Security boundaries of an optical-power limiter for protecting quantum-key-distribution systems

Qingquan Peng,[1,2,3] Binwu Gao,[1] Konstantin Zaitsev,[4,5,6,7,8] Dongyang Wang,[1] Jiangfang Ding,[1] Yingwen Liu,[1] Qin Liao,[9] Ying Guo,[10,2,*] Anqi Huang,[1,†] and Junjie Wu[1,‡]

[1] *Institute for Quantum Information and State Key Laboratory of High Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China*
[2] *School of Automation, Central South University, Changsha 410083, China*
[3] *China Greatwall Research Institute, China Greatwall Technology Group Co., Ltd., Shenzhen 518057, China*
[4] *Russian Quantum Center, Skolkovo, Moscow 121205, Russia*
[5] *NTI Center for Quantum Communications, National University of Science and Technology MISiS, Moscow 119049, Russia*
[6] *Vigo Quantum Communication Center, University of Vigo, 36310 Vigo, Spain*
[7] *Escuela de Ingeniería de Telecomunicación, Department of Signal Theory and Communications, University of Vigo, 36310 Vigo, Spain*
[8] *atlanTTic Research Center, University of Vigo, 36310 Vigo, Spain*
[9] *College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410082, China*
[10] *School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China*

(Received 1 April 2023; revised 12 October 2023; accepted 21 December 2023; published 16 January 2024)

Unauthorized light injection has always been a vital threat to the practical security of a quantum-key-distribution (QKD) system. An optical-power limiter (OPL) based on the thermo-optical defocusing effect has been proposed and implemented, limiting the injected hacking light. As a hardware countermeasure, the performance of the OPL under various light-injection attacks will be tested to clarify the security boundary before it is widely deployed. To investigate the security boundary of the OPL in quantum cryptography, we comprehensively test and analyze the behavior of the OPL under continuous-wave (cw) light-injection attacks and pulse-illumination attacks with pulse repetition rates of 0.5 Hz, 40 MHz, and 1 GHz. The test results illuminate the security boundary of the OPL, which allows one to properly employ the OPL in use cases. The methodology of testing and analysis proposed here is applicable to other power-limitation components in a QKD system.

DOI: 10.1103/PhysRevApplied.21.014026

I. INTRODUCTION

Quantum-key-distribution (QKD), based on the laws of quantum mechanics, provides a promising path for the sharing of a symmetric key between two parties, which has been proven to be informational-theoretically secure [1–3]. Due to the rapid development of QKD, it has become one of the most mature applications in the field of quantum information and it has gradually become globalized and commercialized [4–6]. However, the imperfections in QKD implementations may disclose loopholes, which are employed by quantum hackers to compromise the practical security of QKD systems [7–34]. Quantum hacking can be classified as passive attacks and active attacks. Passive attacks do not change the features of a QKD system. On the other hand, a quantum hacker can actively modify the operation mechanism of a QKD system, creating a loophole. These quantum active attacks that usually use laser light as a tool are unpredictable, which threatens the practical security of QKD systems at a high-risk level.

Active attacks on the source unit, including laser-seeding attacks [29,35], laser-damage attacks [30,36], and Trojan-horse attacks [37–39], compromise the security of QKD implementation. One commonality of these attacks is that the adversary (henceforth called Eve) injects unauthorized light into the source apparatus of the QKD to eavesdrop on secret-key information. Taking laser-seeding attacks as an example, Eve injects tailored light into Alice's laser diode, which results in an increase in the intensity of the emitted optical pulses. Under such an attack, the secret-key rate of the original protocol may be overestimated, since Alice and Bob are not aware of

[*] guoying@bupt.edu.cn
[†] angelhuang.hn@gmail.com
[‡] junjiewu@nudt.edu.cn

the existence of the attack [29]. This indicates that Eve can actively open security loopholes to successfully obtain information about secret keys in these QKD systems. Consideration of these attacks in the security proof may eliminate their security threats.

Active attacks on measurement devices, including aftergate attacks [17], blinding attacks [25,34,40,41], and laser-damage attack [23], are another particularly powerful category of side-channel attacks. In these attacks, Eve exploits the imperfection of the single-photon detector, using laser light to control the behavior of the detectors. For example, in blinding attacks, Eve sends relatively strong laser light that makes the detector unable to operate in Geiger mode, such that it is no longer sensitive to a single photon [18]. By applying the blinding-attack method, Eve could obtain 100% of the key information, without being noticed by Alice and Bob. Fortunately, innovative QKD protocols, such as measurement-device-independent (MDI) QKD, are immune to these detection-side-channel attacks [42].

While conducting the above-mentioned active quantum attacks, the injected unauthorized light is the essential tool for Eve. This unauthorized bright light modifies the characteristics of the targeted QKD system, breaking some of the vital security assumptions of the QKD protocol [29,39]. Although we cannot stop Eve injecting hacking light into a QKD system, a module as protection to limit the injected light could be applied, either in the source unit or in the measurement unit. This injection-power-limitation module will present dynamical insertion loss, i.e., the higher the input power, the higher is the insertion loss. Moreover, this module will not affect the other properties, except for limiting the injection power. Typically, a module called an optical-power limiter (OPL) uses nonlinear optical effects to keep the output optical power stable and below a threshold [43]. A preliminary study of an OPL and its application to a QKD system is presented in Ref. [44], which highlights its power-limiting effects. Specifically, when the injected power is between 31 and 100 mW, its output power can only be stabilized at around 1 mW.

However, a hardware patch will be investigated to verify its security performance, to iterate on enhancing security. For each iteration, the security boundary of the patch will be investigated to show its capabilities and limitations. Under these guidelines, the security boundary of the OPL that limits eavesdropping light will be comprehensively studied under various quantum attacks that employ higher powers of continuous-wave (cw) light and that may also use pulsed light. Although some basic testing of the OPL has already been conducted in Ref. [44], the security boundary of the OPL will be comprehensively studied under various quantum attacks that employ higher powers of cw light and also pulsed light. For this purpose, we test the OPL in the following hacking scenarios that are not fully investigated in Ref. [44]. In scenario one, Eve illuminates the OPL with high-intensity cw light, the optical power of which is up to 5 W. In scenario two, Eve illuminates the OPL with strong optical pulses. In this case, we adopt different repetition frequencies (0.5 Hz, 40 MHz, and 1 GHz) for the injected pulses to thoroughly test the behavior of the OPL under pulse-illumination attacks and Trojan-horse attacks. The test results present the security boundary of the OPL, which exposes its protection limitations for a QKD system. This work provides some reference and guidance on using the OPL properly. Furthermore, the methodology of testing and analyzing the security boundaries proposed in this work can be applied to other hardware patches. The study aims to gain a better and more comprehensive understanding of the devices in safeguarding QKD systems, thus supporting the improvement of their security performance. These research findings may hold significant value for researchers and practitioners in the fields of information security and quantum communication.

The paper is organized as follows. In Sec. II, we first introduce an experimental model of the OPL and then calibrate the performance of the OPL. The cw-light experiments and pulsed-light experiments on the OPL are presented in Secs. III and IV, respectively. In Sec. V, we discuss the security boundary and make relevant recommendations for use. Finally, the work is concluded in Sec. VI.

II. EXPERIMENTAL SETUP

A. Scheme of experiment

The scheme of our experiment is shown in Fig. 1, where the orange solid lines represent the optical signal and the blue dashed lines represent the electrical signal. The OPL is a plug-and-play component, shown as the yellow-shaded module in Fig. 1, which consists of two collimators, an acrylic prism, and a diaphragm aperture. In this experiment, three types of acrylic prism with lengths of 25.4 mm, 50.8 mm, and 101.6 mm are used, respectively. The diaphragm pore is chosen to be 800 μm in diameter. For the complete testing of power limitations, the experimental setup includes Eve's and Alice's light source. Eve's laser is located on the input side of the OPL to model hacking injected light via a quantum channel. Alice's laser is placed on the other side of the OPL, acting as the laser source used to prepare the weak coherence states in a QKD system.

Eve's laser can produce both pulsed and cw light. The maximum output power of the cw light reaches 10 W and the maximum peak power of the pulsed light used in the experiment is 800 mW. The laser (cw or pulse light) produced by Eve is split into two by a 50:50 beam splitter (50:50 BS). One of the split optical paths is connected to Detector 2, monitoring the output power of Eve's laser in real time. The other half of the split light is injected

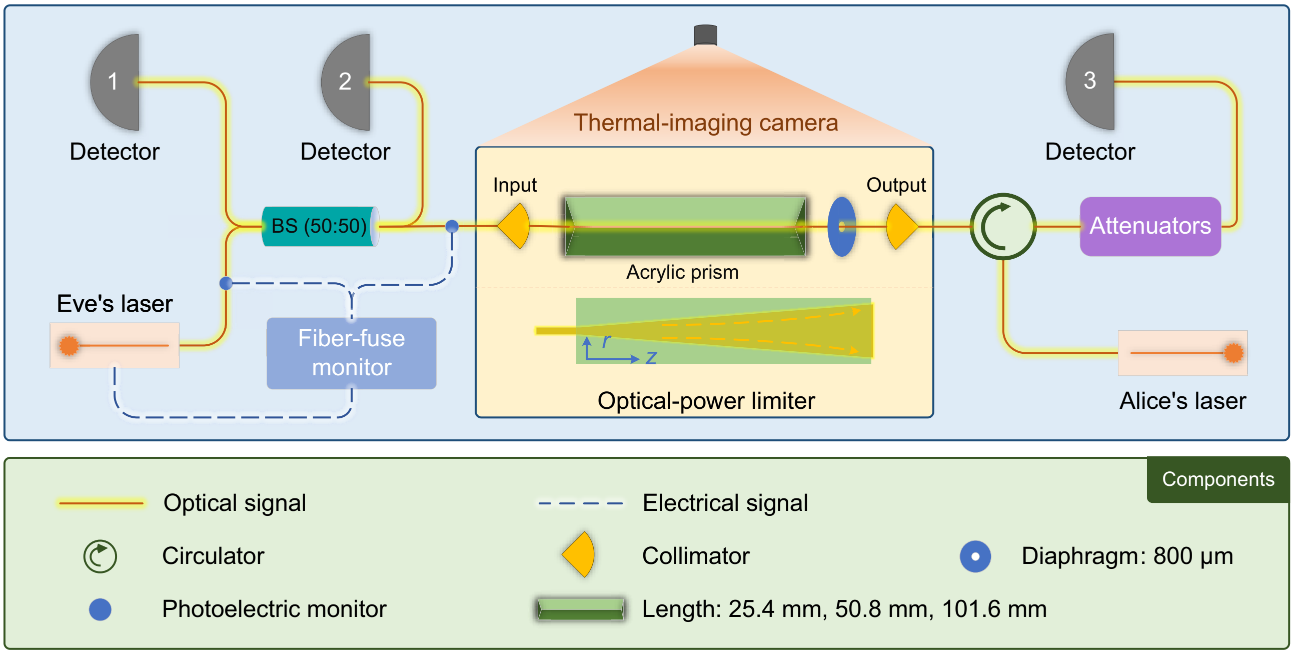

FIG. 1. The experimental scheme of the OPL. The core part of the OPL consists of two collimators, an acrylic prism, and a diaphragm. The internal temperature of the acrylic prism changes when absorbing energy, causing the incident collimated Gaussian beam to diverge due to the thermal-defocusing effects of light. A diaphragm behind the acrylic prism controls the collected light power. The plot below the diagram of the OPL shows the light-signal divergence inside the acrylic prism. Since acrylic prisms are isotropic, both the optical and thermal responses are axially symmetric along the optical axis. Detectors 1 and 2 are optical-power detectors. Detector 3 is an optical-power meter in the cw-light and 0.5-Hz pulsed-light experiments, while in the experiments of injection pulses at repetition frequencies of 40 MHz and 1 GHz, an optical-electrical converter is used as Detector 3 and a variable optical attenuator is added. A fiber-fuse monitor protects Eve's laser. A thermal imager records the temperature change of the OPL during the experiment in real time.

into the OPL. When passing through the OPL, the laser beam first passes through a circulator and is detected by Detector 3. It is worth noting that Detector 3 refers to an optical-power meter used in the following cw-light and 0.5-Hz pulsed-light experiments but an optical-electrical converter applied in the 40-MHz and 1-GHz pulse experiments is introduced later in the paper. When Detector 3 represents an optical-electrical converter, a variable attenuator is needed to ensure that the power of the input light is in the linear-response range of the optical-electrical converter. Alice's laser produces 3-mW cw light, which passes through the OPL and 50:50 BS and is finally detected by Detector 1.

In order to prevent the setup from damage to the experimental equipment other than the OPL when using the high-power laser, a set of protective measures are also employed. A fiber-fuse monitor containing two fiber-fuse sensors and an automatic shutdown system is applied to protect the setup from fiber fuse that may possibly occur under high-power injection. The fiber-fuse sensors are placed along the fiber jacket, shown as blue dots in Fig. 1. Once fiber fuse is detected, the monitoring circuit automatically shuts down Eve's high-power laser, stopping the fiber fuse and preventing extensive damage to the equipment. Fortunately, fiber fuse did not occur during our experiment. In addition, we add a circulator to the output of the OPL to protect Alice's laser from being destroyed by Eve's high-power laser beam.

B. Calibration

Before the experiment, we used a 1550-nm laser to calibrate the power-limitation feature of the OPL in both the forward (input-output) and the backward (output-input) directions. During calibration, we minimize attenuation from the forward direction for each sample. The input power from the forward direction is monitored by Detector 2 and converted to the power at the input port of the OPL by considering the specific splitting ratio of the 50:50 BS. Consequently, the output power is measured by Detector 3 and converted to the power at the output port of the OPL by compensating the insertion loss of the circulator. The calibrated characteristics of the power limiter from the forward direction are shown in Fig. 2. Similarly, Fig. 3 presents the power-limitation property from the backward direction. The red solid line, yellow dashed line, and blue dotted line in Figs. 2 and 3 represent the lengths of the acrylic prisms, which are 25.4 mm, 50.8 mm, and 101.6 mm. In both

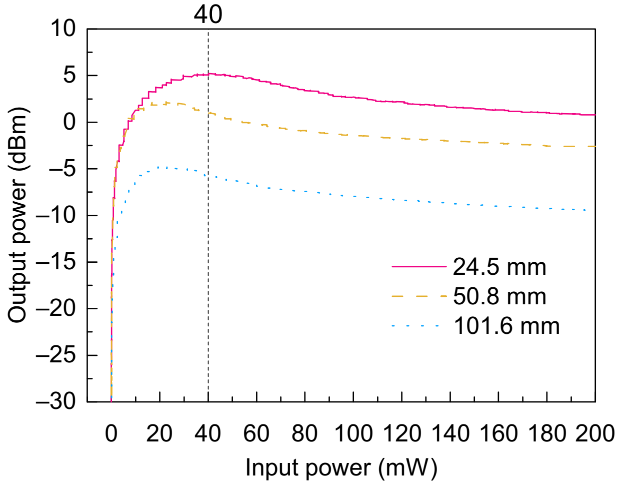

FIG. 2. The calibrated power-limitation characteristics from the forward direction of the optical-power limiter. The “Input power” represents the laser power that Eve injects into the OPL from its input port. The “Output power” represents the optical power measured at the output port of the OPL.

Figs. 2 and 3, the functionality of limiting the transmitted optical power is clearly illustrated from both directions and the output power remains stable when the input power ranges from 40 to 200 mW, which is the maximal power applied in the calibration. It is also found that the longer the acrylic prism fitted in the OPL is, the greater is the attenuation that is provided. For example, in Fig. 2, when the input power is 200 mW, the output power of the OPL with a 25.4-mm acrylic prism is 1.2 mW, while that of the OPL with a 50.8-mm and a 101.6-mm acrylic prism is reduced to 551 and 114 mW, respectively.

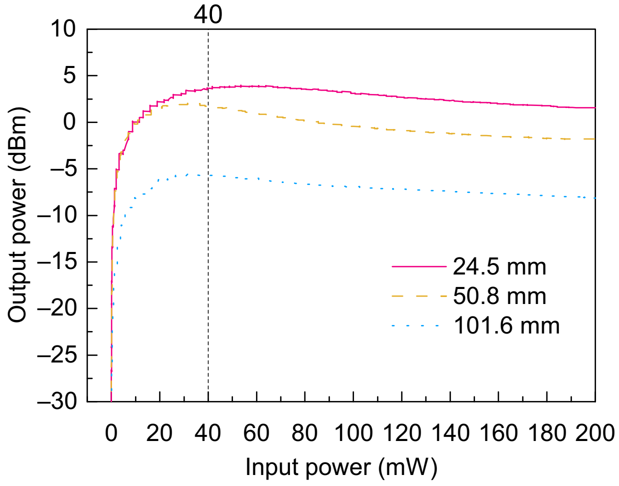

FIG. 3. The calibrated power-limitation characteristics from the backward direction of the optical-power limiter. The “Input power” represents the laser power that Alice injects into the OPL from its output port. The “Output power” represents the optical power measured at the input port of the OPL.

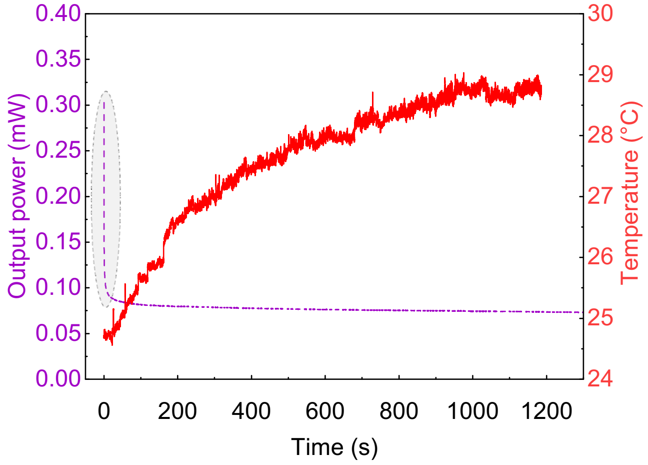

FIG. 4. The variation of the output power and temperature during OPL operation for 20 min. The input power is 200 mW and the length of the acrylic prism is 101.6 mm.

The core component of the OPL is the acrylic prism, which effectively controls the output optical power based on the thermo-optical defocusing effect [45–47]. In order to characterize the stability of the power limitation, we send 200-mW cw light into the OPL from the forward direction for 20 min and obtain the change in the output power and temperature over time, as shown in Fig. 4. The results show that the temperature of the acrylic prisms increases rapidly in the first 10 min of being irradiated and tends to reach a balance between heat accumulation and heat dissipation in the second 10 min. It is worth noting that once we turn the laser on, there is a set of very strong optical powers detected by Detector 3, as shown in the gray-shaded area in Fig. 4. In other words, the cw light emitted by Eve is not limited until the OPL activates the thermo-optical defocusing effect. It is shown that the thermo-optical defocusing effect time of the acrylic prism is about 200 ms [44]. This time gap may provide some opportunity for Eve to use pulsed light to mitigate the effect of the power limitation, which we show in the experiments presented later in this paper.

As typical cases, here we only illustrate the calibration results for three OPLs equipped with different acrylic prism lengths. In practice, each OPL used in the following experiment is calibrated according to the above procedure in order to be precise.

III. TESTING UNDER EVE’S cw LASER

In principle, Eve is able to hack a QKD system via applying an optical power that is limited by the handling capability of the quantum channel [30]. For a fiber-based QKD system, Eve’s hacking power will be below the laser-induced-damage threshold (LIDT) of the standard single-mode fiber that acts as the quantum channel. Thus, as the component that Eve's hacking light first reaches, the power-limitation performance of the OPL will be experimentally tested under the full range of the allowed hacking power, instead of only 200 mW as the maximum tested in Ref. [44]. Previous research has shown that 20-m single-mode optical fiber can withstand a 10-W cw laser [30], which is also the specified maximum power of our laser source used in this study. To fully investigate the characterization of the OPL from both the forward and backward directions under Eve's cw-light illumination, we design the testing cycle using the following steps:

(1) *Eve's high-power laser testing*. Eve injects the cw light at 1550 nm, the power of which is 223 mW $+ k *$ 200 mW (where k is the cycle number, starting from 0) into the OPL from the forward direction, in the first round. Each test lasts for 2 min, at which point the measurement results are relatively stable. Detector 2 monitors the power injected into the OPL by Eve in real time and Detector 3 measures the light power passing through the OPL. Meanwhile, Alice's laser is turned off. We call this process the "Eve-strong test."

(2) *Eve's weak-power laser testing*. Eve adjusts her laser power to be 223 mW (constant during all cycles) and injects into the OPL again. The injection time lasts for 10–20 s, until the detection value is relatively stable. Detector 3 measures the light power transmitted through the OPL from the forward direction. We call this process the "Eve-weak test."

(3) *Alice's laser testing*. Alice generates 3-mW cw light and sends it to the OPL from the backward direction. The injection time lasts for 10–20 s, until the detection value is relatively stable. Detector 1 measures the optical power transmitted through the OPL. Meanwhile, Eve's laser is turned off. We call this process the "Alice-weak test."

(4) *Increases Eve's high-power laser.* Eve increases her laser power in step (1) by 200 mW.

(5) *Test repeat.* We repeat steps (1)–(4) until Eve's injection power reaches 5 W.

According to the above-mentioned test procedure, we conduct three groups of cyclic experiments. In each group of experiments, OPLs with three different acrylic prism lengths, 25.4 mm, 50.8 mm, and 101.6 mm, are tested. Since acrylic appears to suffer irreversible damage after each round of experiments, each acrylic prism sample can only be tested once. The experimental results are shown in Fig. 5, in which the yellow, blue, and purple lines represent the measured optical power after transmission through the OPL in the Eve-strong test, the Eve-weak test, and the Alice-weak test, respectively. During the experiment, the OPL device is placed inside a box surrounded by a black metal plate but the top of the box is open. A thermal-imaging camera looks down on the entire OPL device from the top of the box and records the entire experiment. As shown in Fig. 5, the green stars represent the highest values obtained by the thermal-imaging camera in each round of testing.

Although the OPL under test indeed limits the transmitted power, a more significant variation in power is still shown in certain areas. After being illuminated by Eve's light up to 1 W, the transmittance of the OPL from the backward direction increases slightly, as shown by the purple lines in the red-shaded areas in Fig. 5. We take the OPL with the acrylic prism length of 25.4 mm in group (a) of Fig. 5 as an example. When Alice's injection power remains stable at 3 mW, the transmitted power from the backward direction increases from 0.13 to 0.61 mW after illumination by Eve's 1-W light. A similar phenomenon also appears in the OPLs tested in group (c) of Fig. 5. That is, the output power increases from 0.057 to 0.063 mW in the sample with a 50.8-mm-long acrylic prism, after being illuminated by 0.388 W light from the forward direction. For the OPL with a 101.6-mm-long acrylic prism, the output power increases from 0.002 to 0.012 mW under Eve's injection power of 0.6 W.

The transmittance from the forward direction under Eve-weak testing becomes higher than its original value after the OPL is illuminated by Eve's light with a power between 2 and 3 W, as shown by the blue lines in the shaded areas in Fig. 5. Taking group (a) of Fig. 5 as an example, in the sample with an acrylic prism length of 25.4 mm, the optical power injected into the OPL by Eve's laser is stable at 223 mW but the output power is increased from 1.5464 to 2.672 mW. The same phenomenon also appears in groups (b) and (c) of Fig. 5. For samples with an acrylic prism length of 50.8 mm in groups (b) and (c) of Fig. 5, the output power increases from 0.56 to 0.677 mW and from 0.784 to 0.853 mW, respectively. For the sample with an acrylic prism length of 25.4 mm in group (c) of Fig. 5, the output power increases from 1.9 to 2.315 mW.

As the injected optical power increases from 2 W to 4 W, the transmitted power under Eve-strong testing also increases, as shown by the orange line in the yellow- and gray-shaded area in Fig. 5. In group (a) of Fig. 5, the output power of the sample with an acrylic prism length of 50.8 mm increases from 0.76 to 0.98 mW. Similarly, for the samples with an acrylic prism length of 25.4 mm in groups (b) and (c) of Fig. 5, the output power increases from 1.06 to 1.366 mW and from 1.83 to 2.315 mW, respectively. In general, the longer the acrylic prism is, the lower is the probability of increased output power. For example, when the acrylic prism is 101.6 mm long, in groups (a) and (b) of Fig. 5, the transmitted power of the OPL does not exceed the initial value. Once the injected power is beyond 4 W, the transmittance drops due to the physical damage.

Each acrylic prism sample suffers irreversible physical damage after completing the cycling test. The damaged OPL has a very high optical attenuation, transmitting only a few tens of nanowatts optical power under 5-W injection

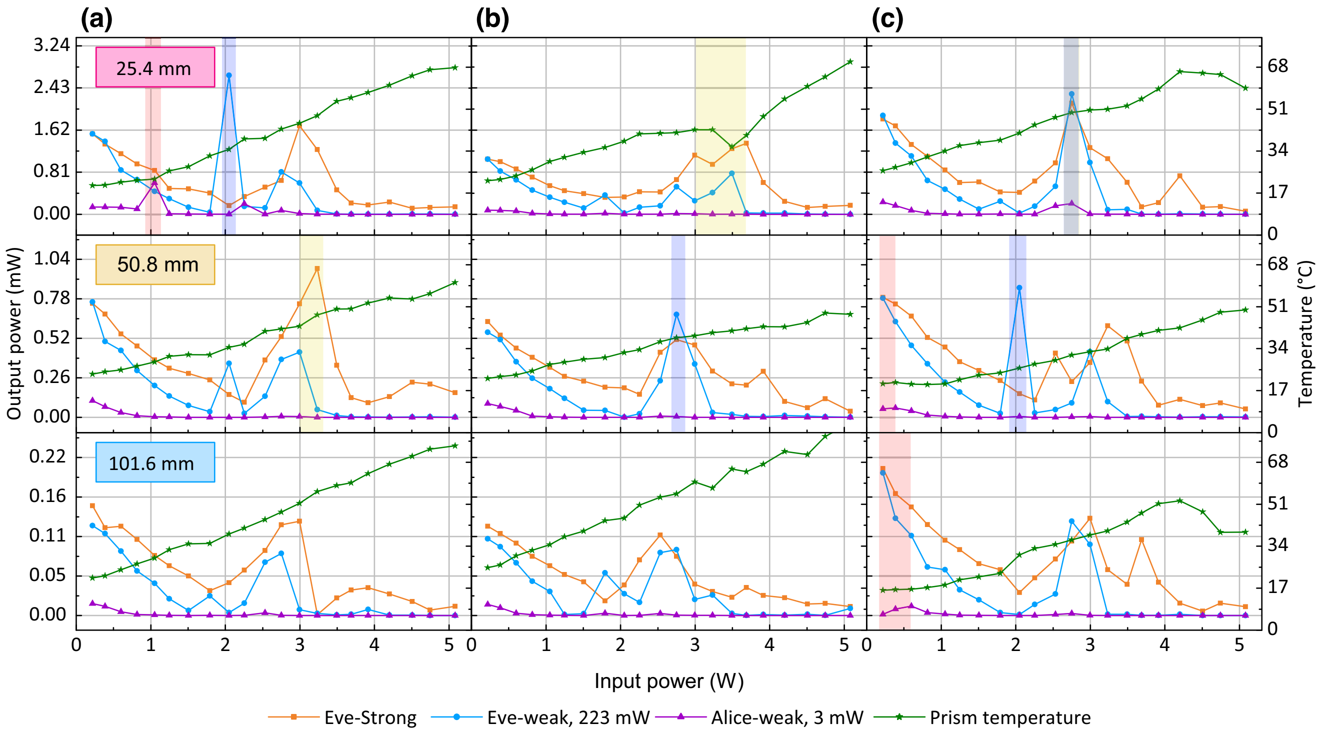

FIG. 5. The experimental results for the OPL illuminated by Eve's cw light. The "Input power" refers specifically to the optical power that Eve injects into the OPL laser during the Eve-strong test. The "Output power" corresponds to the output optical power of the OPL tested during the experiment. (a)–(c) The results of three groups of replicative experiments, the groups containing an OPL with the following acrylic prism lengths: (a) 25.4 mm, (b) 50.8 mm, and (c) 101.6 mm.

light. In order to further understand this damage, we use a thermal imager to continuously observe the temperature of the OPL, which is shown in Fig. 6(a). The power value marked under each thermal image represents the optical power injected into the OPL under Eve-strong testing. It can be seen that the heat accumulates in the acrylic prism as the injected optical power continues to increase. Thus, the overall temperature of the OPL begins to rise due to the thermal radiation from the acrylic prism. The highest-temperature point is right behind the contact surface, between the acrylic prism and the collimator at the input port. Figure 6(b) shows photographs of the acrylic prism after being tested. A small bubble has appeared on the surface at the point first reached by Eve's laser beam, which shows that the high-power laser causes permanent damage to the OPL.

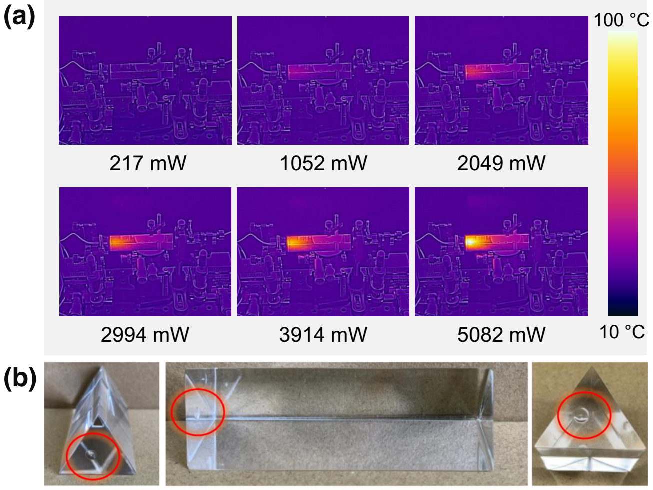

FIG. 6. An analysis of the OPL response to high-power laser exposure. (a) A thermal image of an acrylic prism illuminated by a high-power cw laser. (b) Photographs of the acrylic prism after the experiment. The length of the acrylic prism is 101.6 mm. The red marks indicate the positions of raised bubbles after the acrylic prism has been damaged.

IV. TESTING UNDER EVE'S PULSED LASER

In quantum attacks, Eve is not only allowed to apply cw light to the QKD system [23,25,30,41] but she also can use optical laser pulses to exploit loopholes [34,37,38,40]. This shows that the threat caused by pulsed light is as much as that caused by cw light, which is because some attacks rely on the instantaneous energy instead of the average energy. Therefore, in this section, we extend the investigation of the power-limitation property of the OPL to optical pulses with repetition rates of 0.5 Hz, 40 MHz, and 1 GHz.

A. Eve's optical pulses with 0.5-Hz repetition rate

During the calibration, Fig. 4 shows that the transmitted power reaches a peak immediately after the optical power is sent to the OPL and then drops to a steady level. This indicates that the power-limitation capability may vary over time and may relate to the illumination duration. In order to study the power-limitation response under the hertz-level illumination duration, we first apply optical pulses working at a 0.5-Hz repetition rate and a wavelength of 1550 nm to the OPL as an initial trial. The duty cycle of the light pulse is 50% and the peak power is 200 and 400 mW, respectively. The output power is recorded every 0.1 s by Detector 3, for 10 min, to produce a set of scatter plots.

Figure 7 shows the power-limitation response of the OPL illuminated by Eve's optical pulses working at the repetition frequency of 0.5-Hz. The red, orange, and blue scatter plots in the figure represent the output power of the OPL with 25.4-mm-long, 50.8-mm-long, and 101.6-mm-long acrylic prisms, respectively. The phenomenon that a longer acrylic prism leads to a lower output power is also shown in this testing. In addition, the purple horizontal line in each prism diagram indicates the optical power detected by Detector 3 if Eve apples cw light with a power the same as the peak power of the optical pulses. It can be seen from Fig. 7 that most measured values of the output power under a 0.5-Hz pulsed laser exceed the purple line. Crucially, this suggests that Eve could exploit the finite response time of the OPL to inject more light into a QKD system. It is also notable that the longer acrylic prism limits the peak power to being lower and closer to the purple line.

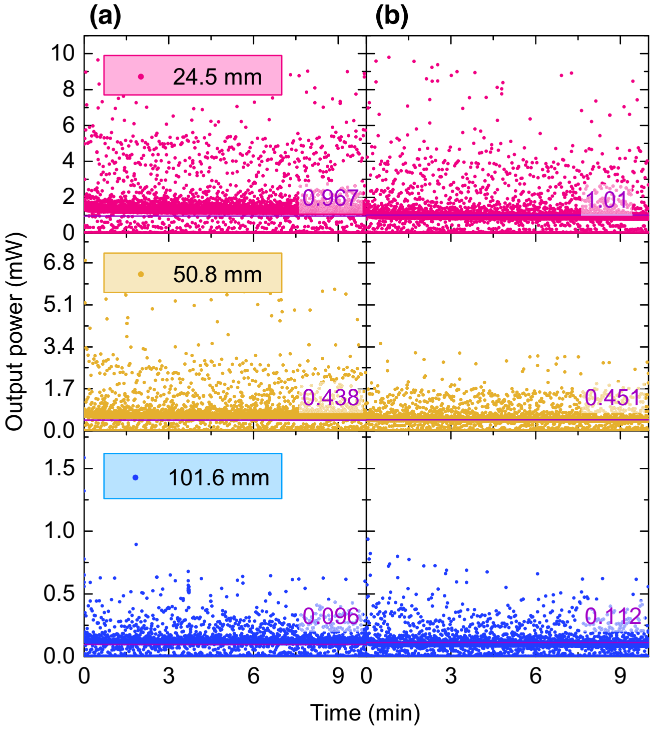

FIG. 7. Scatter plots of the output power of the OPL under Eve's injection pulses with a repetition frequency of 0.5-Hz. (a) The peak power of Eve's optical pulses is 200 mW. (b) The peak power of Eve's optical pulses is 400 mW.

B. Eve's optical pulses with 40-MHz repetition rate

In a quantum attack, Eve may apply optical pulses with a repetition rate the same as that of a QKD system, such as the pulse-illumination attack presented in Ref. [34,40]. Restricting the power of Eve's illumination light to the avalanche photodiode (APD) is one of the most direct and effective means of preventing such attacks. Therefore, it is necessary to investigate whether the OPL can protect an APD from these attacks. In this subsection, we test the transmittance of the OPL under light injection by Eve, who employs optical pulses with the same repetition rate as that used in Ref. [40].

Eve's laser emits optical pulses at 1550 nm and the repetition frequency is 40-MHz, the same as that of the APD tested in Ref. [40]. In our experiments, the width of each optical pulse is set as 4 ns. Detector 2 monitors the average power of Eve's optical pulses. The attenuator dynamically adjusts its attenuation according to Eve's injection power, ensuring that Detector 3 responds linearly. We test the performance of the OPL when the average power of Eve's optical pulses is 10, 20, 30, 60, and 80 mW. Additionally, we test the OPL with different acrylic prism lengths under each average power of Eve's light.

The test results are shown in Fig. 8. It is clear that while Eve increases the average power of the optical pulses, the output power of the OPL first increases and then decreases, no matter what length of acrylic prism is used. During the process of Eve increasing the pulse power, the maximal output power of the OPL can be observed when the average power of the optical pulses is 30 or 60 mW. For the OPL with a 25.4-mm-long acrylic prism, the maximal output power of the OPL measured in the experiment is 56.59 mW. For the OPL with a 50.8-mm-long (101.6-mm-long) acrylic prism, the maximal output power of the OPL measured in the experiment is 38.83 mW (13.69 mW). Thus, the longer acrylic prism limits the output power to the lower value.

In addition, the pulsed light can be easily passed through the OPL compared to the cw light. The purple line in each subfigure in Fig. 8 represents the output power of the OPL under cw-light injection, the power of which is as the same as the average power of the optical pulses. The peak power of the transmitted pulses in each test demonstrates a much higher value than the purple line. For example, in Fig. 8(c), for the OPL equipped with a 25.4-mm-long acrylic prism, Eve's optical pulses result in a transmitted power of 38.83 mW, while only 3.78-mW optical power is transmitted through the OPL under the cw-laser test. In the

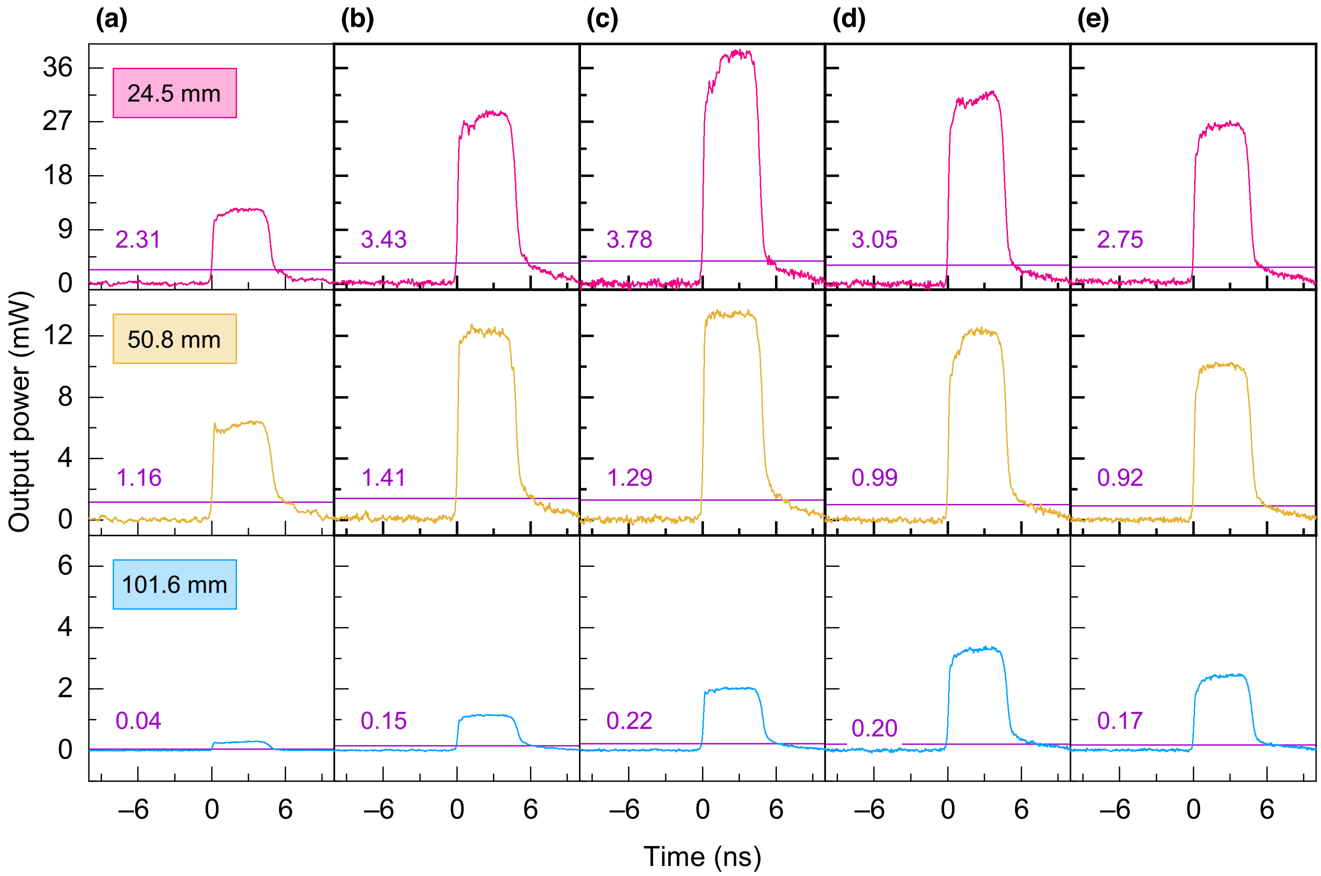

FIG. 8. The output wave forms of the OPL under 40-MHz pulsed light. The average optical power injected by Eve into the OPL is (a) 10, (b) 20, (c) 30, (d) 60, and (e) 80 mW.

test, the peak power of the optical pulses is 5.5656–16.969 times higher than that of the cw light under the same input optical power, on average.

C. Eve’s optical pulses with 1-GHz repetition rate

In this subsection, we investigate the power-limitation response of the OPL under high-speed pulse illumination, which might be used in Eve’s attack on a high-speed QKD system [34]. Usually, a high-speed QKD system would have a gigahertz-level repetition frequency. Therefore, in this test, we prepare a 1550-nm optical pulse with a 1-GHz repetition frequency and about 200-ps full width at half maximum as a typical case.

The specific experimental procedure is similar to that of Sec. IV B. Each experimental setup consists of OPLs equipped with acrylic prisms of three different lengths. The average optical power of the pulse injected by Eve into the OPL is gradually increased from 10 to 80 mW, the same as that of the last test in Sec. IV B. The output power is first measured by Detector 3 and is then converted to the power at the output port of the OPL. The test results are presented in Fig. 9. In each subfigure, the purple solid line represents the output power of the OPL if the cw-light injection is performed with an optical power the same as the average power of the optical pulses. From this test, the following phenomena can be disclosed.

Similar to the 40-MHz optical-pulse test, the maximum output power in this test can be obtained when the average optical power injected by Eve into the OPL is 30 or 60 mW. In the OPL equipped with an acrylic prism length of 25.4 mm (101.6 mm), the maximum output power is 55.31 mW (5.41 mW) when the average input power is 30 mW. The maximum output power of the OPL with a 50.8-mm-long acrylic prism is 19.33 mW under input optical pulses with average power of 60 mW.

When the average input power of the optical pulses is as the same as that of the cw light, their output power is higher than that under cw-light injection. For example, when Eve applies optical pulses with an average optical power of 10 mW to the OPL with a 25.4-mm-long acrylic prism, its output peak optical power is 28.87 mW; whereas the output power is only 2.094 mW under 10-mW cw input light. All the cases that we test follow this regularity.

When the average optical power is constant, the peak output power of the OPL increases with the pulse light frequency. For example, compare the same situation in Figs. 8(c) and 9(c), when the length of the acrylic prism

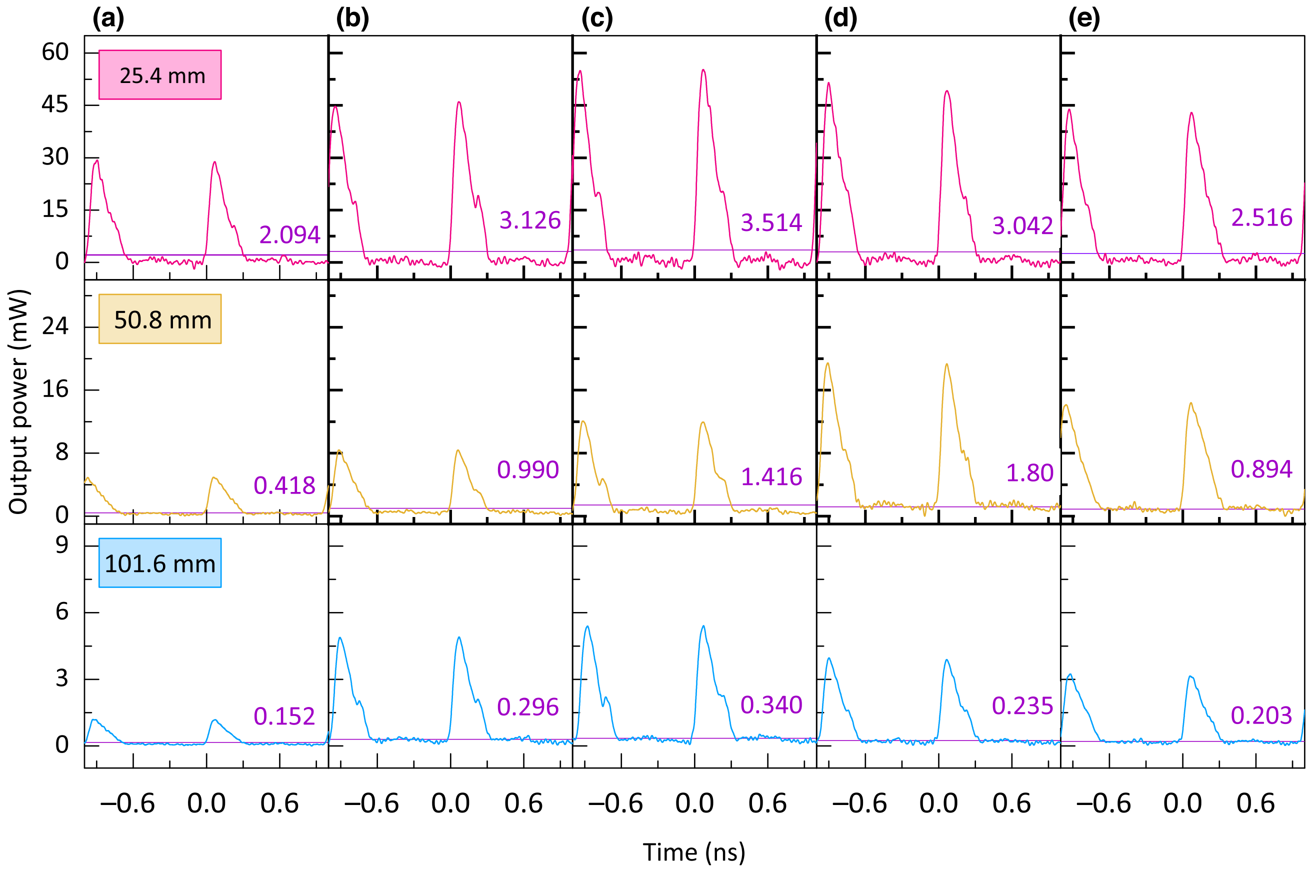

FIG. 9. The output wave forms of the OPL under 1-GHz pulsed light. The average optical power injected by Eve into the OPL is (a) 10, (b) 20, (c) 30, (d) 60, and (e) 80 mW.

assembled in the OPL is 25.4 mm. We find that the maximum output power corresponding to the pulsed light at 40 MHz is 39.6046 mW, while the maximum output power corresponding to the 1-GHz light is 56.869 mW. In this test, the peak power of the optical pulses is 5.627–16.969 times higher than that of the cw light, assuming that the average input power is the same.

In the pulsed-laser experiments demonstrated above, it is shown that the peak output power of the pulsed light transmitted through the OPL is higher than that for cw light, assuming that the input power is equal on average. The reason for this phenomenon is that the OPL limits the output optical power based on the thermo-optical defocusing effect. In the case of the pulsed light, the OPL only receives a short period of light radiation followed by a long idle time as a repetition cycle, in which it is difficult to accumulate the heat to produce the thermo-optical defocusing effect. Thus, more light is transmitted through the OPL under pulsed injection.

V. DISCUSSION ABOUT SECURITY BOUNDARY

In the tests shown above, the OPL indeed demonstrates its power-limitation effects. Additionally, the OPL shows variation in response under cw high-power light and pulsed light. From the results and principles of the OPL, it can be seen that the length and temperature changes of the acrylic prism will affect its power-limiting performance. Therefore, the security boundary of the OPL is studied under the potential risk of quantum attacks, considering the parameters of the acrylic prism as a necessity.

A. Security boundary for cw-light injection

In the cw-light experiments, generally, the OPL keeps the transmitted power at the same order of magnitude as the optical power no matter how much input light there is. Nevertheless, the OPL still may expose some security threats that will be noticed. This is because the significant increase in the optical power allows us to observe the fluctuation and degradation of the performance of the OPL under high-intensity light, which has also not been explored in the previous work [44], presenting the security boundary of the OPL.

First, for the source unit of a QKD system, the amount of Alice’s transmitted light increases slightly, as shown in our test, which may allow the mean photon number to be higher than the set value. Usually, in Alice’s case, the insertion loss of the OPL is counted as a part of the process that attenuates the weak coherence laser to the single-photon level. In a normal case, the mean photon numbers of Alice's output follow the set values required by a QKD protocol. However, if Eve injects about 1 W cw light, there is a chance of increasing the mean photon number of Alice's pulses, as shown by the red area in Fig. 5. This unnoticed increase of the optical power sent by Alice results in an incorrect estimation of the key rate, which compromises the security of a QKD system that is running either a prepare-and-measure QKD protocol or a measurement-device-independent QKD protocol [29].

Specifically, as shown in Fig. 10, the dashed and solid lines represent the incorrect and correct secret-key rates for a decoy-state BB84 QKD system with an increase in Alice's mean photon number, denoted as R_I and R_C, respectively. The green, orange, and blue curves indicate that the mean photon numbers of Alice's pulses are increased to $g = 1.17$ times, $g = 4.41$ times, and $g = 7.16$ times the set values, according to the test results shown in Fig. 5. The black solid line represents the secret-key rate without an attack. It is apparent that the secret-key rate estimated by Alice and Bob, R_I—given by the dashed lines—is significantly higher than the accurate key rate, R_C. More precisely, once the attenuation value of the OPL decreases due to the attack, Alice and Bob are not aware of the increase in the transmitted mean photon number and they wrongly estimate the secret key according to the security proof introduced in Refs. [48,49]. Remarkably, the dashed lines are higher than the black solid line, which illustrates that the incorrectly estimated secret-key rate cannot be guaranteed as a secure one.

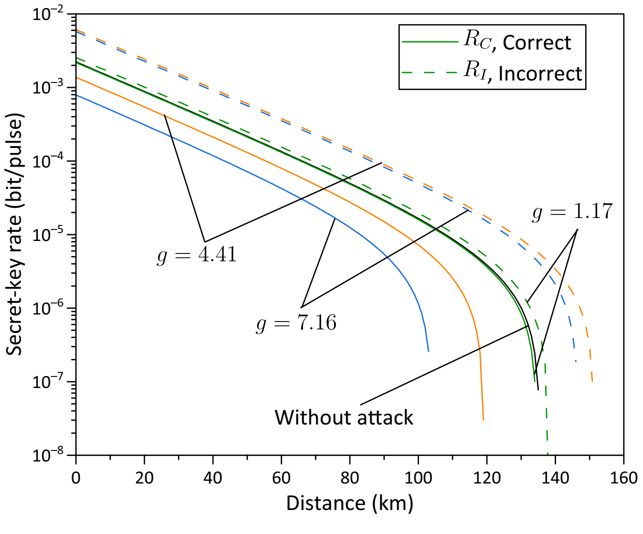

FIG. 10. The secret-key rate as a function of the distance for the standard decoy-state BB84 protocol at different mean photon numbers of Alice's pulses. The average number of photons in the signal state and the decoy state are set to 0.1 and 0.05, respectively. The other parameters in the simulation are as follows: the dark-count probability per gate is 2×10^{-6}; the error-correction efficiency is 1; the probability that a photon will hit the erroneous detector is 0.01; Bob's internal optics loss is 0.4; and the detection efficiency of Bob's single-photon detector is 10%.

We also consider the case of MDI QKD with weak coherent pulses [42]. Similar to the previous example, we assume that the attenuation value decreases after the OPL is attacked, so that both Alice's and Bob's output intensity increases. The resulting secret-key rates are shown in Fig. 11. From the overall observations, we find that the results are analogous to those illustrated in Fig. 10. In particular, in the presence of an attack, the R_I values estimated by Alice and Bob are much higher than the correct one, R_C, provided that Alice and Bob are not aware of the increase in intensity under the attack. It is worth noting that when the light intensity is increased by factors of 4.41 and 7.16, no secret key can be generated.

Besides, we can also see from Fig. 5 that at very beginning of the experiment with a slow temperature rise, there is small decrease in the optical power of Alice's pulses. Thus, the lowest temperature corresponds to the highest optical power of Alice's pulses. When considering the security of a QKD system with the OPL, temperature manipulation may cause a loophole. For a QKD system placed in a room with a temperature 20 °C, Eve may try to manipulate the conditions in a laboratory to decrease the temperature to, say, 15 °C. In this case, the decreased temperature may allow the optical power of Alice's pulses to be higher, which again leads to an unnoticed increase of the mean photon number sent by Alice. It is notable that

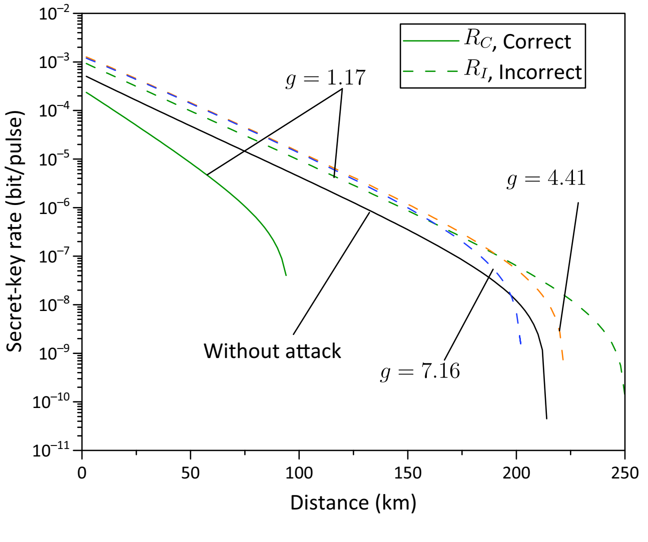

FIG. 11. The secret-key rate as a function of the distance for the symmetric MDI-QKD protocol at different mean photon numbers of Alice's pulses. The parameters used in the simulations are taken from Ref. [50]. The light intensity set by Alice and Bob in the vacuum state, the decoy state, and the signal state is 0, 0.03, and 0.5, respectively.

whether or not the OPL has the above-mentioned vulnerabilities still requires further experimental verification.

To counter this potential threat, one possible solution is to characterize the minimal insertion loss of the OPL and always take this minimal value into account to calculate the required attenuation for Alice's set mean photon number. Of course, this countermeasure provides a conservative estimation of the secret-key rate, which is lower than the optimal one. Another more conservative but secure countermeasure is to assume that in Alice's case, the OPL is untrusted. Thus, the attenuation needed to prepare quantum states with a set mean photon number will be accounted for without the OPL. In this case, no matter how the insertion loss of the OPL changes, the security assumption about the mean photon number at the set single-photon level will not be affected. On the other hand, the trade-off is the lower secret-key rate and the shorter communication distance for the QKD system.

Second, for the detection unit of a QKD system, the OPL can still transmit milliwatt-level light under several hundreds of milliwatts or 2–4 W of injected light, which is usually sufficient to conduct a detector-blinding attack on an APD. Previous studies have been shown that nanowatt to milliwatt cw light is able to blind the APD, no matter whether it is a passively quenched module [20,51], an actively quenched module [52], or a gated module [25,41,53]. Thus, the OPL cannot prevent a QKD system from these blinding attacks. An exception, an APD with a low-impedance bias voltage supply, which requires about 10 mW to blind the APD detector [14]. This is because the blinding light is no longer reduces the bias voltage of this low-impedance biased APD. Instead, the blinding light is used to accumulate heat in the APD to increase its temperature, increasing the temperature-dependent breakdown voltage [14]. Apparently, this thermal effect requires more injection light.

However, the OPL is likely to prevent the source unit in a QKD system from laser-seeding attacks [29,54] and laser-damage attacks [26,30,36]. Since the injection power is at the milliwatt level after the OPL, a laser-seeding attack would not work if Alice were to apply more than 40-dB attenuation to ensure that less than 100-nW light was injected into the laser diode. With regard to laser-damage attacks, it is shown in our test that high-power optical light can be attenuated to several milliwatts, which in Alice's case is usually not enough to damage the optical components. The high-power optical light even may destroy the OPL to block the transmission, protecting other components behind it from being hacked. Moreover, we find that cw high-power light causes a significant increase in the temperature of the OPL. Thus, the OPL combined with a thermal sensor may be able to sense the increased temperature and trigger further protective action once the temperature is beyond a certain threshold.

B. Security boundary for pulsed-light injection

In the pulsed-light experiments, the optical pulses are transmitted through the OPL. After that, the peak power of the optical pulses varies from 0.370 to 56.599 mW, which is 5.627–16.943 times that of the cw light. The peak power of the transmitted pulses indeed reduces and becomes much lower than the original pulses. However, the remaining optical pulses may still help Eve to conduct pulse-illumination attacks on single-photon avalanche diodes (SPADs) and Trojan-horse attacks on the modulators. In this subsection, we focus on the security boundary in the scenarios of pulse-based attacks equipped with the OPL, providing a more comprehensive picture about the security performance of the OPL.

For optical pulses with a repetition rate of 40 MHz, the peak power of the transmitted optical pulses can reach from several milliwatts to a maximum of 38.83 mW. Our testing shows that optical pulses transmitted through the OPL may be sufficient to perform a pulse-illumination attack on the APDs as shown in Ref. [40]. Specifically, under the average input power of 10 mW, the sequence of optical pulses transmitted through the OPL with an 50.8-mm-long acrylic prism is shown in Fig. 12, presenting the peak power above 6 mW. This amount of peak power with about 5-ns pulse width provides a similar pulse shape to that of the blinding pulses used in the pulse-illumination attack, which indicates that there is enough energy in each optical pulse to blind the APD tested in Ref. [40]. Moreover, the repetition rate tested in this study, 40 MHz, is also the same as that of the pulse-illumination attack demonstrated in Ref. [40]. Thus, an OPL equipped with such an acrylic prism may not be able to prevent a QKD system from a pulse-illumination attack. It is notable that a OPL with a longer acrylic prism, of length 101.6 mm,

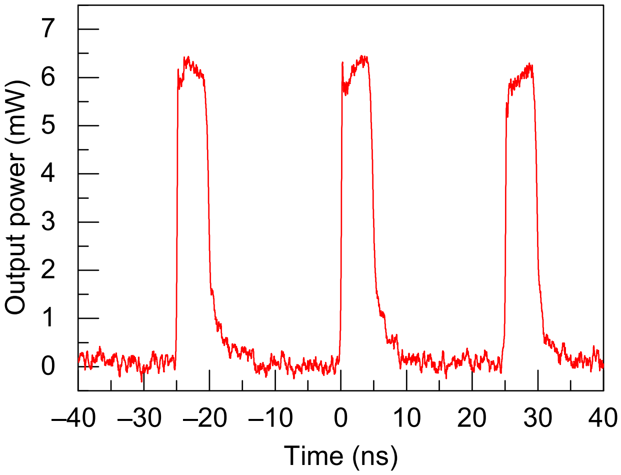

FIG. 12. The performance of the OPL with 40-MHz pulsed light. The length of the acrylic prism is 50.8 mm and the average optical power of the pulsed light is 10 mW.

attenuates the peak power further to mitigate the threat of pulse-illumination attack.

Optical pulses with repetition rates of 1 GHz and 40 MHz can support Eve's Trojan-horse attack on a QKD system, especially for a high-speed attack. As shown by the test in Sec. IV C, optical pulses with 1-GHz repetition frequency injected into the OPL result in a transmitted peak power higher than that of the cw light and the 40-MHz optical pulses. For a QKD system also working at a 1-GHz repetition rate, these high-speed optical pulses transmitted through the OPL may reach the quantum state modulators—such as phase modulators, intensity modulators, or polarization modulators—and then be reflected. The reflected optical pulses carry the modulation information of the prepared quantum states. Thus, Eve can obtain the information about the secret key via the injected and reflected Trojan-horse pulses.

A previous study has shown the passive architecture against Trojan-horse attack (schematically shown in Fig. 13) when the average photon number of the reflected light is sufficiently low [37]. As can be seen from Fig. 13, the Trojan-horse light enters via the quantum channel on the right-hand side, passes through the OPL, an optical isolator I, and an optical attenuator A, and is then reflected at the internal optical components (R is the total reflection of the internal optical elements to the left of the dot-dashed line), finally returning along the original path. Therefore, the optical isolation against the Trojan-horse attack can be calculated as $\gamma = O^2 I^n A^2 R$, where n represents the number of optical isolators present in the system. Under normal circumstances, the attenuation value of the OPL itself increases with the enhancement of the injected light. However, our experimental results show that Eve's hacking makes the attenuation value of the OPL decrease. Therefore, the optical isolation of the OPL itself becomes untrusted from the security point of view. To be secure, the optical isolation against the Trojan-horse attack can be further written as

$$\gamma = I^n \times A^2 \times R. \tag{1}$$

Converting the absolute optical isolation value to a decibel (dB) value can more intuitively reflect the attenuation of the component. If the absolute value of the optical isolation of the component is x, we use the notation $\dot{x} = 10 \log_{10} x$ [37]. Equation (1) can be conveniently rewritten in decibels, as

$$\dot{\gamma} = n\dot{I} + 2\dot{A} + \dot{R}. \tag{2}$$

It should be noted that the passive architecture in Fig. 13 guarantees security against a Trojan-horse attack only when the average photon leakage number is $\mu_{\text{out}} = 10^{-6}$ ($\dot{\mu}_{\text{out}} = -60$ dB). That is, we can guarantee security against a Trojan-horse attack by reasonably controlling the isolation of the components, i.e.,

$$\dot{\mu}_{\text{out}} = \dot{\gamma} + \dot{\chi}. \tag{3}$$

Figure 8 shows that when the system frequency is 40 MHz, the maximum number of photons passing through is $\chi = 1.39113 \times 10^9$ ($\dot{\chi} \simeq 90$ dB). From Eq. (3), we then obtain $\dot{\gamma} = \dot{\mu}_{\text{out}} - \dot{\chi} = (-60 - 90)$ dB $= -150$ dB. Figure 9 shows that when the system frequency is 1 GHz, the maximum number of photons passing through is $\chi = 0.7461649 \times 10^8$ ($\dot{\chi} \simeq 80$ dB). Similarly, we then obtain $\dot{\gamma} = -140$ dB. This result is the total optical isolation required for security in Alice's module. Based on this, we provide several combinations of attenuation values for other components after using an OPL, so that the system can prevent Trojan-horse attacks. For convenience, we report the absolute values of the components, as shown in Table I, which contains some possible combinations.

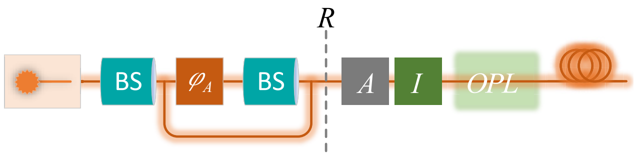

FIG. 13. The reflected-signal model for QKD with a Trojan-horse attack. φ_A is the encoding device, R is the total reflection from all components to the left of the dot-dashed line, A is an attenuator, and I is an optical isolator.

TABLE I. The practical combinations of system components to passive architecture against Trojan-horse attacks. All dotted quantities are in decibels and are given as absolute values.

Clock rate	$\lvert\dot{\gamma}\rvert$	$\lvert\dot{R}\rvert$	$\lvert\dot{A}\rvert$	$\lvert\dot{I}\rvert$
1 GHz	140	40	40	60(1)
1 GHz	140	20	0	60(2)
40 MHz	150	40	10	50(2)
40 MHz	150	30	0	60(2)

C. Security boundary of a blinding attack under the protection of the OPL

Some attacks, such as Bob's blinding attack on a SPAD, may use both cw light and pulsed light. Therefore, the security boundary under the protection of the OPL depends on its responses to both cw and pulsed-injection light. To further investigate this mixed case, we demonstrate an experiment of a blinding attack on a SPAD with the protection of the OPL as a countermeasure, as depicted in Fig. 14(a). The specific steps of the experiment are as follows. First, cw light is injected into the OPL, the results of which are shown in Fig. 14(b). When the optical power injected into the OPL reaches 34.58 nW, the

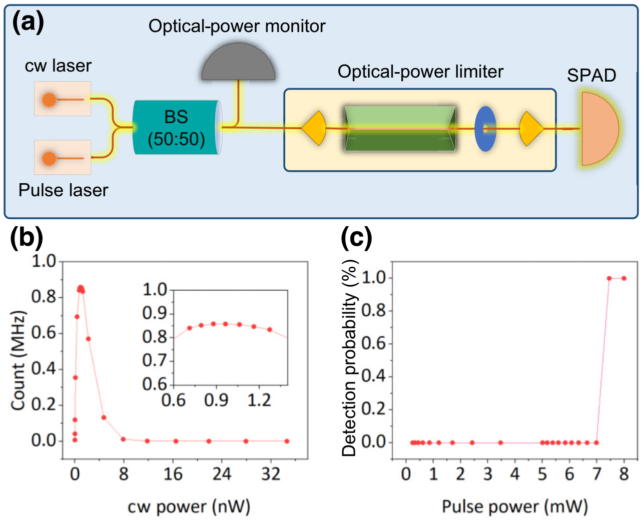

FIG. 14. (a) The testing scheme of a blinding attack on a SPAD under the protection of the OPL. Eve applies a cw laser to blind the SPAD, which is superposed by a pulsed laser to trigger the click of the SPAD. The power of these injected lasers is monitored by an optical-power monitor. The length of the acrylic prism is 50.8 mm and the diaphragm width is 800 μm. The model number of the SPAD is ID210, which operates in external triggering mode with a detection efficiency of 10%, a gate width of 4.99 ns, a gated frequency of 125 MHz, and a dead time of 100 ns. (b) The experimental graph shows the relationship between the power of the cw blinding light and the count rate of the SPAD. (c) The detection probability of the SPAD as a function of the trigger-pulse energy. The trigger pulses are generated with a repetition frequency of 125 MHz and a full width at half maximum of 2.2 ns.

SPAD exhibits blinding effects. Then, pulsed light superposed upon the blinding light is injected into the OPL. As shown in Fig. 12(c), when the peak optical power injected into the OPL reaches 8 mW, the detection efficiency of the SPAD becomes 100% to fully control the detection result. Furthermore, we also conduct the same blinding-attack experiment on an OPL equipped with a 101.6-mm-long acrylic prism. In this scenario, the optical power of the cw light needs to exceed 105.2 nW to blind the SPAD. Additionally, if Eve intends to control the responses of the SPAD, the peak optical power of the pulsed light should be greater than 28.2 mW.

These experiments demonstrate that even with the protection of the OPL, Eve still has an opportunity to blind the SPAD and control its response. This phenomenon occurs because the required optical power injected into the OPL to blind the SPAD is much lower than the power-limitation threshold (40 mW) of the OPL. At this point, the OPL behaves as a component with an almost fixed insertion loss, as confirmed in Fig. 2. For the trigger pulse, although the OPL starts to limit the injection power, the transmitted power of the optical pulses is still enough to trigger the click of the blinded SPAD. These experiments show that the OPL has limited effectiveness in resisting weak light attacks, especially when the hacking optical power is less than the power-limitation threshold of the OPL. Moreover, the attacks employing cw and pulsed light reveal a complex scenario for the OPL as a countermeasure to be implemented. This indicates that, in this mixed case, an effective countermeasure is challenging and the security-boundary investigation and security evaluation is also complicated.

VI. CONCLUSIONS

In this paper, we have conducted comprehensive security tests on a passive OPL based on the acrylic prism thermo-optical defocusing effect. The experiments have utilized light sources covering a range from 0 to 5 W of cw light, as well as pulsed light with frequencies of 0.5 Hz, 40 MHz, and 1 GHz. Through experimental verification, we have disclosed the security boundary of the OPL. First, it is undeniable that when the OPL is operating normally, the power of the injected hacking light can be limited below a certain value. However, when there is a physical change in the OPL introduced by cw injected light, there is a window of reduced restrictions, which provides opportunities for Eve to launch an attack. Moreover, under the same average input optical power, the peak power of pulsed light can pass through the OPL to a greater extent than that of cw light. The higher the repetition rate of the injected pulses, the larger is the peak power. Further security analysis based on the test results is given to present the security boundary. This study provides a more comprehensive understanding of the OPL, which allows one to use this device, with a known security boundary, properly. The testing methodology is also applicable to other types of OPL to verify their security performance.

ACKNOWLEDGMENTS

We thank V. Makarov, C. Wang, and G. Zhang for helpful discussions. This work was funded by the National Natural Science Foundation of China (No. 62371459 and No. 62061136011), the National Key Research and Development Program of China (Grant No. 2019QY0702), the Research Fund Program of the State Key Laboratory of High Performance Computing (Grant No. 202001-02), the Innovation Program for Quantum Science and Technology (Grant No. 2021ZD0300704), the Key Research and Development Program of Hunan Province (Grant No. 2022GK2016), and the Key Project of Scientific Research of Hunan Provincial Education Department (Grant No. 22A0669). K.Z. was supported by the Ministry of Science and Education of Russia (program of the NTI Center for Quantum Communications) and the Russian Science Foundation (Grant No. 21-42-00040), the Galician Regional Government (consolidation of Research Units: AtlantTIC), "Ministerio de Ciencia e Innovacion" (Ministry of Science and Innovations, Spanish) (MICIN), with funding from the European Union NextGenerationEU (PRTR-C17.I1), and the Galician Regional Government with own funding through the "Planes Complementarios de I+D+I con las Comunidades Aut'onomas."

Q.P., B.G., and A.H. conducted the experiment. Q.P. and K.Z. analyzed the data. Q.P. and A.H. wrote the paper, with input from all authors. A.H., J.W., and Y.G. supervised the project.

[1] H.-K. Lo, M. Curty, and K. Tamaki, Secure quantum key distribution, Nat. Photonics **8**, 595 (2014).

[2] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, Rev. Mod. Phys. **74**, 145 (2002).

[3] E. Diamanti, H. Lo, B. Qi, and Z. Yuan, Practical challenges in quantum key distribution, npj Quantum Inf. **2**, 1 (2016).

[4] H. Takesue, S. W. Nam, Q. Zhang, R. H. Hadfield, T. Honjo, K. Tamaki, and Y. Yamamoto, Quantum key distribution over a 40-dB channel loss using superconducting single-photon detectors, Nat. Photonics **1**, 343 (2007).

[5] P. Sibson, J. E. Kennard, S. Stanisic, C. Erven, J. L. O'Brien, and M. G. Thompson, Integrated silicon photonics for high-speed quantum key distribution, Optica **4**, 172 (2017).

[6] T. A. Eriksson, T. Hirano, B. J. Puttnam, G. Rademacher, R. S. Luís, M. Fujiwara, R. Namiki, Y. Awaji, M. Takeoka, and N. Wada, Wavelength division multiplexing of continuous variable quantum key distribution and 18.3 tbit/s data channels, Commun. Phys. **2**, 1 (2019).

[7] N. Lütkenhaus, Security against individual attacks for realistic quantum key distribution, Phys. Rev. A **61**, 052304 (2000).

[8] V. Makarov, A. Anisimov, and J. Skaar, Effects of detector efficiency mismatch on security of quantum cryptosystems, Phys. Rev. A **74**, 022313 (2006). **78**, 019905 (2008), erratum ibid.

[9] C. H. F. Fung, B. Qi, K. Tamaki, and H. K. Lo, Phase-remapping attack in practical quantum-key-distribution systems, Phys. Rev. A **75**, 032314 (2007).

[10] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, Time-shift attack in practical quantum cryptosystems, Quantum Inf. Comput. **7**, 73 (2007).

[11] A. Lamas-Linares and C. Kurtsiefer, Breaking a quantum key distribution system through a timing side channel, Opt. Express **15**, 9388 (2007).

[12] Y. Zhao, C. H. F. Fung, B. Qi, C. Chen, and H.-K. Lo, Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems, Phys. Rev. A **78**, 042333 (2008).

[13] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Hacking commercial quantum cryptography systems by tailored bright illumination, Nat. Photonics **4**, 686 (2010).

[14] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Thermal blinding of gated detectors in quantum cryptography, Opt. Express **18**, 27938 (2010).

[15] F. Xu, B. Qi, and H.-K. Lo, Experimental demonstration of phase-remapping attack in a practical quantum key distribution system, New J. Phys. **12**, 113026 (2010).

[16] H.-W. Li, S. Wang, J.-Z. Huang, W. Chen, Z.-Q. Yin, F.-Y. Li, Z. Zhou, D. Liu, Y. Zhang, G.-C. Guo, W.-S. Bao, and Z.-F. Han, Attacking a practical quantum-key-distribution system with wavelength-dependent beam-splitter and multiwavelength sources, Phys. Rev. A **84**, 062308 (2011).

[17] C. Wiechers, L. Lydersen, C. Wittmann, D. Elser, J. Skaar, C. Marquardt, V. Makarov, and G. Leuchs, After-gate attack on a quantum cryptosystem, New J. Phys. **13**, 013043 (2011).

[18] L. Lydersen, M. K. Akhlaghi, A. H. Majedi, J. Skaar, and V. Makarov, Controlling a superconducting nanowire single-photon detector using tailored bright illumination, New J. Phys. **13**, 113042 (2011).

[19] L. Lydersen, N. Jain, C. Wittmann, Ø. Marøy, J. Skaar, C. Marquardt, V. Makarov, and G. Leuchs, Superlinear threshold detectors in quantum cryptography, Phys. Rev. A **84**, 032320 (2011).

[20] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Full-field implementation of a perfect eavesdropper on a quantum cryptography system, Nat. Commun. **2**, 349 (2011).

[21] S.-H. Sun, M.-S. Jiang, and L.-M. Liang, Passive Faraday-mirror attack in a practical two-way quantum-key-distribution system, Phys. Rev. A **83**, 062331 (2011).

[22] N. Jain, C. Wittmann, L. Lydersen, C. Wiechers, D. Elser, C. Marquardt, V. Makarov, and G. Leuchs, Device calibration impacts security of quantum key distribution, Phys. Rev. Lett. **107**, 110501 (2011).

[23] A. N. Bugge, S. Sauge, AinaMardhiyahM Ghazali, J. Skaar, L. Lydersen, and V. Makarov, Laser damage helps the eavesdropper in quantum cryptography, Phys. Rev. Lett. **112**, 070503 (2014).

[24] S. Sajeed, P. Chaiwongkhot, J.-P. Bourgoin, T. Jennewein, N. Lütkenhaus, and V. Makarov, Security loophole in free-space quantum key distribution due to spatial-mode detector-efficiency mismatch, Phys. Rev. A **91**, 062301 (2015).

[25] A. Huang, S. Sajeed, P. Chaiwongkhot, M. Soucarros, M. Legré, and V. Makarov, Testing random-detector-efficiency countermeasure in a commercial system reveals a breakable unrealistic assumption, IEEE J. Quantum Electron. **52**, 8000211 (2016).

[26] V. Makarov, J.-P. Bourgoin, P. Chaiwongkhot, M. Gagné, T. Jennewein, S. Kaiser, R. Kashyap, M. Legré, C. Minshull, and S. Sajeed, Creation of backdoors in quantum communications via laser damage, Phys. Rev. A **94**, 030302(R) (2016).

[27] A. Huang, S.-H. Sun, Z. Liu, and V. Makarov, Quantum key distribution with distinguishable decoy states, Phys. Rev. A **98**, 012330 (2018).

[28] Y.-J. Qian, D.-Y. He, S. Wang, W. Chen, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han, Hacking the quantum key distribution system by exploiting the avalanche-transition region of single-photon detectors, Phys. Rev. Appl. **10**, 064062 (2018).

[29] A. Huang, Á. Navarrete, S.-H. Sun, P. Chaiwongkhot, M. Curty, and V. Makarov, Laser-seeding attack in quantum key distribution, Phys. Rev. Appl. **12**, 064043 (2019).

[30] A. Huang, R. Li, V. Egorov, S. Tchouragoulov, K. Kumar, and V. Makarov, Laser damage attack against optical attenuators in quantum key distribution, Phys. Rev. Appl. **13**, 034017 (2020).

[31] S. Sun and A. Huang, A review of security evaluation of practical quantum key distribution system, Entropy **24**, 260 (2022).

[32] P. Chaiwongkhot, J. Zhong, A. Huang, H. Qin, S.-c. Shi, and V. Makarov, Faking photon number on a transition-edge sensor, EPJ Quantum Technol. **9**, 23 (2022).

[33] A. Huang, A. Mizutani, H.-K. Lo, V. Makarov, and K. Tamaki, Characterisation of state preparation uncertainty in quantum key distribution, ArXiv:2205.11870.

[34] B. Gao, Z. Wu, W. Shi, Y. Liu, D. Wang, C. Yu, A. Huang, and J. Wu, Strong pulse illumination hacks self-differencing avalanche photodiode detectors in a high-speed quantum key distribution system, Phys. Rev. A **106**, 033713 (2022).

[35] Y. Zheng, P. Huang, A. Huang, J. Peng, and G. Zeng, Security analysis of practical continuous-variable quantum key distribution systems under laser seeding attack, Opt. Express **27**, 27369 (2019).

[36] A. Ponosova, D. Ruzhitskaya, P. Chaiwongkhot, V. Egorov, V. Makarov, and A. Huang, Protecting fiber-optic quantum key distribution sources against light-injection attacks, PRX Quantum **3**, 040307 (2022).

[37] M. Lucamarini, I. Choi, M. B. Ward, J. F. Dynes, Z. L. Yuan, and A. J. Shields, Practical security bounds against the Trojan-horse attack in quantum key distribution, Phys. Rev. X **5**, 031030 (2015).

[38] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, Trojan-horse attacks on quantum-key-distribution systems, Phys. Rev. A **73**, 022320 (2006).

[39] N. Jain, E. Anisimova, I. Khan, V. Makarov, C. Marquardt, and G. Leuchs, Trojan-horse attacks threaten the security of practical quantum cryptography, New J. Phys. **16**, 123030 (2014).

[40] Z. Wu, A. Huang, H. Chen, S. Sun, J. Ding, X. Qiang, X. Fu, P. Xu, and J. Wu, Hacking single-photon avalanche detectors in quantum key distribution via pulse illumination, Opt. Express **28**, 25574 (2020).

[41] L. Lydersen and J. Skaar, Security of quantum key distribution with bit and basis dependent detector flaws, Quantum Inf. Comput. **10**, 60 (2010).

[42] H.-K. Lo, M. Curty, and B. Qi, Measurement-device-independent quantum key distribution, Phys. Rev. Lett. **108**, 130503 (2012).

[43] A. E. Siegman, Nonlinear optical effects: An optical power limiter, Appl. Optics **1**, 127 (1962).

[44] G. Zhang, I. W. Primaatmaja, J. Y. Haw, X. Gong, C. Wang, and Charles Ci Wen Lim, Securing practical quantum communication systems with optical power limiters, PRX Quantum **2**, 030304 (2021).

[45] D. C. Smith, High-power laser propagation: Thermal blooming, Proc. IEEE **65**, 1679 (1977).

[46] R. C. C. Leite, S. Porto, and T. C. Damen, The thermal lens effect as a power-limiting device, Appl. Phys. Lett. **10**, 100 (1967).

[47] M. E. DeRosa and S. L. Logunov, Fiber-optic power limiter based on photothermal defocusing in an optical polymer, Appl. Opt. **42**, 2683 (2003).

[48] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, Security of quantum key distribution with imperfect devices, Quantum Inf. Comput. **4**, 325 (2004).

[49] H.-K. Lo, X. Ma, and K. Chen, Decoy state quantum key distribution, Phys. Rev. Lett. **94**, 230504 (2005).

[50] X. Ma and M. Razavi, Alternative schemes for measurement-device-independent quantum key distribution, Phys. Rev. A **86**, 062319 (2012).

[51] V. Makarov, Controlling passively quenched single photon detectors by bright light, New J. Phys. **11**, 065003 (2009).

[52] G. Gras, N. Sultana, A. Huang, T. Jennewein, F. Bussières, V. Makarov, and H. Zbinden, Optical control of single-photon negative-feedback avalanche diode detector, J. Appl. Phys. **127**, 094502 (2020).

[53] V. Chistiakov, A. Huang, V. Egorov, and V. Makarov, Controlling single-photon detector ID210 with bright light, OE **27**, 32253 (2019).

[54] S.-H. Sun, F. Xu, M.-S. Jiang, X.-C. Ma, H.-K. Lo, and L.-M. Liang, Effect of source tampering in the security of quantum cryptography, Phys. Rev. A **92**, 022304 (2015).