Eve strikes back
in the era of measurement-device-independent quantum key distribution

Anqi Huang,[1,2] Álvaro Navarrete,[3] Ruoping Li,[4] Vladimir Egorov,[5]
Shi-Hai Sun,[6] Poompong Chaiwongkhot,[1,4] Marcos Curty,[3] and Vadim Makarov[4]

[1]*Institute for Quantum Computing, University of Waterloo, Waterloo, ON, N2L 3G1 Canada*
[2]*Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1 Canada*
[3]*EI Telecomunicación, Department of Signal Theory and Communications, University of Vigo, Vigo E-36310, Spain*
[4]*Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, N2L 3G1 Canada*
[5]*Department of Photonics and Optical Information Technology, ITMO University, 199034 Kadetskaya line 3b, St. Petersburg, Russia*
[6]*College of Liberal Arts and Science, National University of Defense Technology, Changsha 410073, P.R.China*

Introduction.—In theory, quantum key distribution (QKD) shares a secret random bit string between two separated parties (Alice and Bob), which provides information-theoretic security based on the laws of physics. In practice, however, it does not, owing to equipment vulnerabilities in implementation.

Researchers have been working on bridging the gap between perfect theory and imperfect practice. As one of the most promising approaches, measurement-device-independent QKD (MDI QKD) [1] was proposed to remove all side channels of a single-photon detector that is regarded as the "Achilles' heel" of QKD [2]. The security of MDI QKD is equivalent to the Einstein-Podolsky-Rosen (EPR) based QKD protocol [3], as it is a time-reversed version of EPR-based QKD [1]. Remarkably, MDI QKD is also highly practical, and it can be realized using current technology [4–7]. Please note that an essential assumption in MDI QKD is that the source is trusted [1]. That is, Alice and Bob are believed to be located at secure laboratories perfectly shielded from the eavesdropper, and they fully know the prepared states. However, our study shows that the practical source is vulnerable to a laser seeding attack and a laser damage attack. Therefore, attacks exploiting practical loopholes in the source might compromise the security of MDI QKD.

The main contributions of this work are twofold. First, we show that the laser seeding attack on a laser diode and the laser damage attack on an optical attenuator can increase the intensities of Alice's pulses. Second, we theoretically analyze the effects of the increased intensity on MDI QKD. The analysis shows that the Alice and Bob wrongly overestimate the secure key rate in this scenario, and they could even consider secure a totally insecure key.

Eve is able to increase the intensities of Alice's pulses.—To investigate Eve's controllability of a semiconductor laser diode, we conduct a laser seeding attack. The testing scheme is shown in Fig. 1. At Alice's side, the laser diode, as a testing target, generates optical pulses. Eve employs a tunable laser (Agilent 8164B), whose wavelength and output power are adjustable, to send continuous-wave (c.w.) bright light to Alice via a single-mode fibre. Thus, Eve injects photons with a proper wavelength into Alice's laser. The energy of each injected photon matches the energy difference between the excited state and the ground state, triggering stimulated emission. In the test, a polarization controller is used to make the polarization of Eve's laser the same as that of Alice's laser, which maximizes the injection efficiency. A circulator isolates the injected light and Alice's emitted light. The energy of each of Alice's pulses is measured using an optical-to-electrical converter

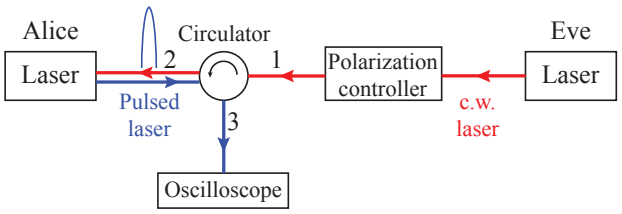

FIG. 1. Setup for testing laser seeding. The red path represents Eve's injection, and the blue path shows the normal photon emission.

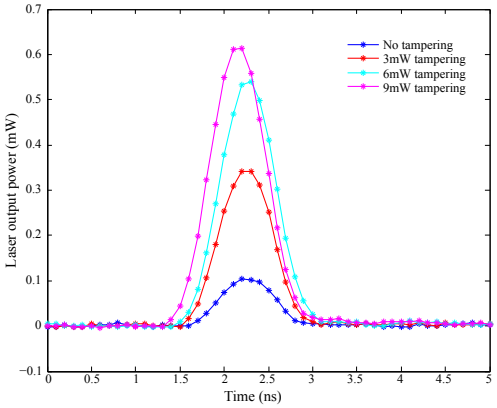

FIG. 2. Waveforms of ID300's laser pulses with and without Eve's tampering.

connected to an oscilloscope (Agilent DSOX93303Q).

Two ID300 short-pulse laser modules from ID Quantique with repetition frequency of 1 MHz have been tested. Our experiment shows that Eve can manipulate the output power of Alice's laser module. The matched wavelengths for sample 1 and sample 2 are 1556.90 nm and 1557.18 nm, respectively. We then gradually increase the power of the c.w. laser. The energy of Alice's laser pulse increases with the injected the c.w. power. Figure 2 shows waveforms of Alice's optical pulses under different amounts of injected laser power. Compared to the original laser pulse, the amplitude of the pulse under attack becomes much higher. Also, the injected light broadens the width of Alice's optical pulse. This is because simulated emission triggered by the injected light takes less time than the spontaneous emission in a normal case. Remarkably, under 9 mW light injection at the laser, its pulse energy increases 5.2 times for sample 1 and 6.0 times for sample 2. As we show in the theory section below, a factor of 2 increase in intensity is sufficient to compromise the security of QKD. This is achieved when seeding this laser module

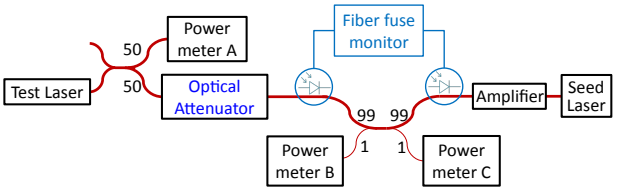

FIG. 3. Setup for testing laser damage of optical attenuators.

with 2 mW power, or realistic 10 W at Alice's channel interface provided she has 37 dB channel-to-laser isolation.

Another method to increase the intensity of Alice's pulse is to perform laser damage attack on an optical attenuator, which is typically her last component before the communication line. We test Eve applying several watt c.w. power via the single-mode fiber communication line. The experimental scheme that simulates hacking a running QKD system is shown in Fig. 3. A test laser simulating Alice's laser emits 20 mW at 1550 nm into the input of an attenuator under test. A high-power erbium-doped fiber amplifier applies up to 9 W c.w. 1550 nm light to the optical attenuator in the opposite direction. The attenuation value before and after the laser damage test is measured by power meter C. The amplifier applies high power starting from 316 mW (25 dBm) for at least 10 s. Afterwards, we turn off the high-power laser and record the attenuation value. If no attenuation change has occurred, the laser power is increased by 0.5–1 dBm, and the steps above are repeated. Once a change in attenuation is registered, the testing stops. If the maximum power, 9 W (39.5 dBm), is applied, but still no change in attenuation, the testing stops.

We have tested four types of optical attenuators from different QKD systems. A manual variable attenuator with a screw tip appears to be essentially unaffected up to 9 W. A fixed attenuator consistently shows a short-term temporary decrease in attenuation, but no permanent decrease. Remarkably, the other two attenuator types exhibit permanent decrease in attenuation. These are a variable optical attenuator (VOA) using micro-electro-mechanical-system (MEMS) and a programmable VOA employing a glass disk covered with variable density metal coating (VDMC).

Out of 13 tested samples of MEMS VOAs, we have been successful in permanently decreasing the attenuation for 8 samples with an average decline of 5.34 dB at their default setting after laser damage. Figure 4 shows a typical voltage-attenuation curve of a successfully compromised sample. The grey area denotes where permanent attenuation drop is observed after 4.47 W (36.5 dBm) exposure.

We have tested one VDMC VOA performing in total 25 measurements over different areas of the glass disk. Out of these, 18 testing points demonstrate permanent decrease in attenuation, with mean decrease of 9.59 dB. Typical attenuation curves before and after successful attacks are shown in Fig. 5. A dip of attenuation appears around the affected point as the result of VDMC being locally destroyed. Optical attenuation gradually increased around the area and returned to normal after 0.5 dB shift in any direction.

Intensity-increased attacks compromise the security of MDI QKD.—In MDI QKD and, more generally, in decoy-state based QKD protocols, the transmitter randomly varies the mean photon number of the pulses sent to the quantum channel, in order to estimate *a posteriori* a

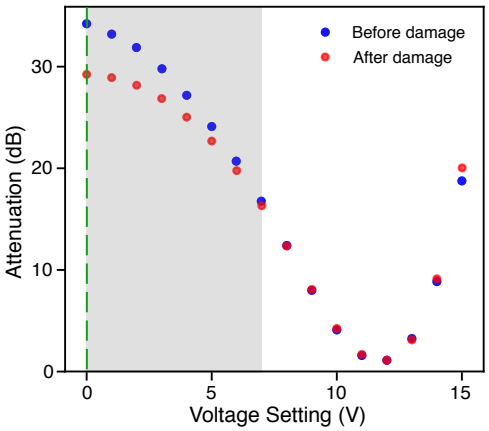

FIG. 4. A typical sample of MEMS VOA before and after laser damage testing. The green dashed line indicates the setting during high-power exposure.

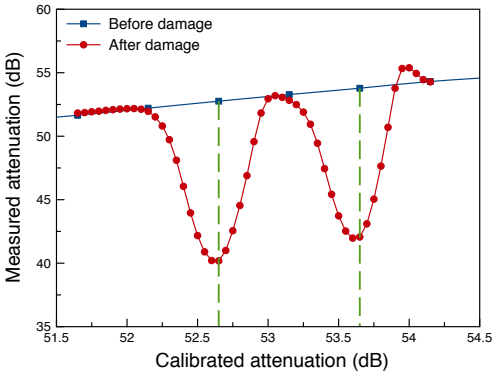

FIG. 5. A typical sample of VDMC VOA before and after laser damage testing. The green dashed lines indicate the settings during high-power exposure.

set of parameters necessary to evaluate the security of the communication. For instance, and to keep the discussion simple, let us consider first the case of the standard decoy-state BB84 protocol. The case of MDI QKD is analogous and will be briefly discussed afterwards. In the decoy-state BB84 protocol, the users can use the relations

$$G^{\mu} = \sum_{n} p_n^{\mu} Y_n,$$
$$E^{\mu} G^{\mu} = \sum_{n} p_n^{\mu} e_n Y_n, \qquad (1)$$

where G^{μ} and Y_n (E^{μ} and e_n) are the overall gain (overall quantum bit error rate) of the signals and the yield (error rate) of an n-photon signal, to obtain bounds on Y_0, Y_1 and e_1. This is possible because they observe, after the quantum transmission step of the protocol, G^{μ} and $E^{\mu}G^{\mu}$ for the different signal and decoy intensities μ. Furthermore, by assumption they know the probabilities p_n^{μ} of sending an n-photon Fock state, which for phase-randomized weak coherent pulses follow a Poissonian distribution $p_n^{\mu} = e^{-\mu}\mu^n/n!$ that depends only on the mean photon number μ. Similar ideas apply to MDI QKD. However, as shown for instance in Fig. 2, a malicious adversary Eve can manipulate the final mean photon number of the pulses sent in each transmission round. This implies that the users, who are unaware of this fact, would base their

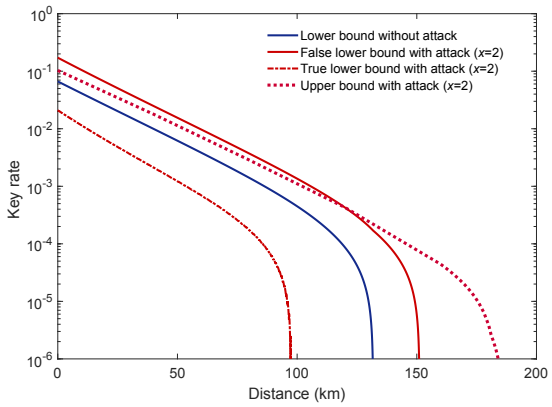

FIG. 6. Upper and lower bounds on the secret key rate for the decoy-state BB84 protocol. In the legend, *true* (*false*) means that the lower bound was calculated using the manipulated (using the original unmodified) intensity values in the analytical expressions.

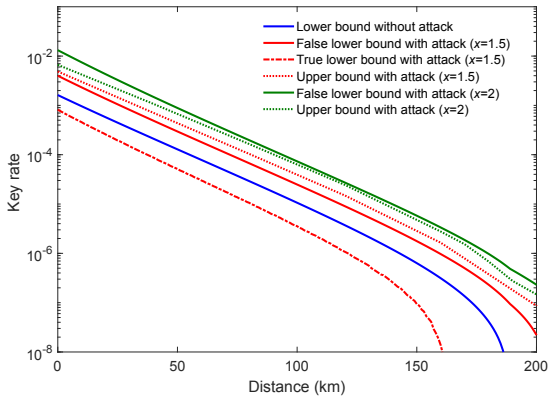

FIG. 7. Upper and lower bounds for the secret key rate in the MDI QKD scenario for different increase factor of intensities. In the legend, *true* (*false*) means that the lower bound was calculated using the manipulated (using the original unmodified) intensity values in the analytical expressions.

estimations on false probabilities p_n^{μ} (instead of the actual probabilities $p_n^{\mu'}$) given the observable quantities $G^{\mu'}$ and $E^{\mu'}G^{\mu'}$ that depend on the modified mean photon number μ'. Hence, under this attack, the parameters estimated by the users might not be proper bounds for Y_0, Y_1 and e_1, and the security of the shared key could be compromised.

To show how these intensity-increased attacks jeopardize the security of decoy-state based QKD protocols, we show here, as an example, some simulations for the decoy-state BB84 protocol and for MDI QKD. In these simulations we assume, for simplicity, that the transmitters send an infinite number of pulses, so we disregard statistical fluctuations. Also, motivated by the experimental results presented above, we consider that Eve's attack increases all the intensities μ by the same factor x. We use known analytical expressions [8, 9] to calculate lower bounds on the key rate. Specifically, three different intensities are used (the signal intensity and two decoys). Furthermore, we use the method proposed in Ref. 10 to calculate an upper bound on the key rate for the BB84 scenario, and we extend these later results to the framework of MDI QKD.

The results are shown in Fig. 6 for the BB84 protocol and in Fig. 7 for MDI QKD. We calculate upper and lower bounds assuming that the intensities of Alice's pulses approximately double ($x = 2$) under Eve's attack. For MDI QKD, we also show the case if the intensities are increased by a factor $x = 1.5$. The original intensities and their associated probabilities were optimized for a typical channel model in the absence of Eve, and the same optimized parameters were used afterwards to simulate the variation on the key rate due to Eve's attack. We set the channel loss to be $\alpha = 0.2$ dB/km, the dark-count probability $p_d = 2.64 \times 10^{-5}$, the misalignment of the quantum channel $e_d = 1.5\%$, and the efficiency of the detectors $\eta_D = 0.3$. The simulations show the lower bound on the secret key rate that the users would calculate in the previously described scenario (*false*) as well as the real one (*true*), which takes into account the variation of the intensity values. We can observe that, in these examples, when the intensities of Alice's pulses are doubled, the *false* lower bound is largely overestimated due to the wrong use of the original unmodified intensities μ in the photon statistics p_n^{μ} in comparison to the *true* lower bound calculated by using the modified intensities μ' in the statistics $p_n^{\mu'}$. Furthermore, in this scenario the *false* lower bound even surpasses the upper bound, implying the shared key would be insecure no matter what security proof is used.

Conclusion—This study reveals new vulnerabilities and loopholes in the source, which indicates that the practical security of the source in a QKD system should be deeply investigated. We have experimentally demonstrated that Eve can increase the intensities of Alice's pulses exploiting the loopholes in the laser diodes and the optical attenuators. Furthermore, we have theoretically shown that the intensity-increased attacks compromise the security of the decoy-state BB84 and MDI QKD protocols. These hacking methods can be applied to test most QKD systems, offering a promising avenue for evaluating and verifying their practical security.

This submission is based on two manuscripts under preparation by the authors.

[1] H.-K. Lo, M. Curty, and B. Qi, Phys. Rev. Lett. **108**, 130503 (2012).
[2] H.-K. Lo, M. Curty, and K. Tamaki, Nat. Photonics **8**, 595 (2014).
[3] H. Inamori, Algorithmica **34**, 340 (2002).
[4] A. Rubenok, J. A. Slater, P. Chan, I. Lucio-Martinez, and W. Tittel, Phys. Rev. Lett. **111**, 130501 (2013).
[5] Z. Tang, Z. Liao, F. Xu, B. Qi, L. Qian, and H.-K. Lo, Phys. Rev. Lett. **112**, 190503 (2014).
[6] S. Pirandola, C. Ottaviani, G. Spedalieri, C. Weedbrook, S. L. Braunstein, S. Lloyd, T. Gehring, C. S. Jacobsen, and U. L. Andersen, Nat. Photonics **9**, 397 (2015).
[7] Y.-L. Tang, H.-L. Yin, Q. Zhao, H. Liu, X.-X. Sun, M.-Q. Huang, W.-J. Zhang, S.-J. Chen, L. Zhang, L.-X. You, Z. Wang, Y. Liu, C.-Y. Lu, X. Jiang, X. Ma, Q. Zhang, T.-Y. Chen, and J.-W. Pan, Phys. Rev. X **6**, 011024 (2016).
[8] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, Phys. Rev. A **72**, 012326 (2005).
[9] F. Xu, H. Xu, and H.-K. Lo, Phys. Rev. A **89**, 052333 (2014).
[10] M. Curty, T. Moroder, X. Ma, H.-K. Lo, and N. Lütkenhaus, Phys. Rev. A **79**, 032335 (2009).