OPTICA

Hacking measurement-device-independent quantum key distribution

Feng-Yu Lu,[1,2,†] Peng Ye,[1,2,†] Ze-Hao Wang,[1,2] Shuang Wang,[1,2,3,*] Zhen-Qiang Yin,[1,2,3,5] Rong Wang,[1,2,4] Xiao-Juan Huang,[1,2] Wei Chen,[1,2,3] De-Yong He,[1,2,3] Guan-Jie Fan-Yuan,[1,2] Guang-Can Guo,[1,2,3] and Zheng-Fu Han[1,2,3]

[1]CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei, Anhui 230026, China
[2]CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei, Anhui 230026, China
[3]Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China
[4]Department of Physics, University of Hong Kong, Hong Kong SAR, China
[5]yinzq@ustc.edu.cn
[†]These authors contributed equally to this work.
*wshuang@ustc.edu.cn

Received 10 January 2023; revised 21 March 2023; accepted 22 March 2023; published 20 April 2023

Decoy-state measurement-device-independent quantum key distribution (MDI-QKD) is regarded as one of the most promising protocols since it is physically immune to all fatal detection-side attacks. However, the loopholes in the source side still open side channels to eavesdroppers. Previous work has indicated that a system is insecure if its decoy-state modulation is inaccurate, but a practical method for Eve to exploit the inaccurate modulation is still absent. In this work, we propose a practical hacking method and provide a detailed theoretical analysis. Using our method, we experimentally hacked a MDI-QKD system and successfully obtained all final keys. By demonstrating the hacking experiment, we hope to bring attention to source-side security and provide inspiration for the development of protocols with better source-side security. © 2023 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement

https://doi.org/10.1364/OPTICA.485389

1. INTRODUCTION

Quantum key distribution (QKD) [1] allows two distant parties, named Alice and Bob, to share encryption keys with information-theoretic security based on physical laws [2–5], even if there is an eavesdropper, Eve. As one of the most successful technologies in the field of quantum information, QKD has aroused wide attention and achieved increasingly higher key rates and longer communication distances. However, hidden side channels are lurking in practical QKD systems since the theoretical security assumes an unbreakable source and detection side, but the behavior of real devices typically deviates from that considered in the security proofs. Eve may take advantage of these side channels to hack the practical systems [6–26].

The practical single-photon detectors (SPDs) were once the most vulnerable point in practical QKDs [6–13]. Fortunately, an elegant solution named measurement-device-independent quantum key distribution (MDI-QKD) [27] was proposed to solve the detection-side loopholes once and for all. The MDI-QKD is physically immune to all detection-side attacks, so its security has nothing to do with the measurement unit (MU) Charlie, even if Charlie is an eavesdropper.

However, the MDI-QKD still requires ideal sources, where there is still a gap in the practical situation. For instance, the absence of ideal single-photon sources brings a well-known quantum hacking named the photon-number-splitting (PNS) attack [14–16]. Practical QKDs usually employ phase-randomized weak coherent sources whose photon number obeys the Poisson distribution. Eve could fully control the quantum bits by splitting some photons from the multi-photon pulses, performing quantum nondemolition detections, and storing her results in quantum memory. These insecure quantum bits are named tagged bits [28–31]. As a countermeasure, a milestone method named the decoy-state method [29–31] was proposed. By randomly and precisely modulating several different intensities, Alice and Bob can accurately estimate their information leakage from the tagged bits and further reduce them by privacy amplification. The decoy-state method can reach a comparable key rate compared with using ideal single-photon sources and has been employed in almost all QKDs [32–41], including MDI-QKD [42–46].

In recent years, outstanding works have caused decoy-state MDI-QKD to be regarded as one of the best solutions for practical applications [47–64]. However, these great achievements do not mean that the QKD users can rest easy. As mentioned above, the decoy-state method requires random and precise intensity modulation. Some works found that, in practical sources, the intensity modulation would lose control due to environmental changes [65,66], limited electronic systems, and imperfect modulators [19,67–69]. Some of the errors come from the intrinsic imperfections of practical devices. For instance, the correlated deviation [18,19,67–71] brings correlated intensity deviation in decoy-state

modulation; the random intensity fluctuation [69,72,73] brings an unknown photon number distribution so that the users cannot accurately solve the decoy-state equation anymore; the inhomogeneous electrical properties [65,74–76] in the widely employed $LiNbO_3$ modulators make its DC-bias drift with time. Some of the modulation errors could be actively introduced by Eve such as the laser-seeding attack [17,20] and laser-damage attack [21,23].

Although the above works have pointed out that the imperfect modulations inevitably exist on the source side and bring potential threats, the source-side security still has not received enough attention. An important reason is that the details about how to practically exploit the modulation errors are absent. However, the source-side security should not be ignored because nobody can guarantee that there would not be stronger techniques in the future. To arouse people's attention, in this work, we propose a practically usable hacking strategy that borrows from the idea of the PNS attack but only requires general devices rather than quantum memory and quantum nondemolition detections. We also proposed a theoretical model to analyze what error is in Eve's favor. The hacking strategy was demonstrated through two experiments in this work. In the first experiment, the decoy states are inaccurately modulated, and Eve passively exploits the modulation error. The purpose of the first experiment is to verify the feasibility of our hacking method and the correctness of our theory. In the second one, Eve actively disturbs Alice's modulation by developing an induced-photorefractive technique [77]. In both of the experiments, Eve successfully obtained almost all final keys. By verifying that Eve can indeed threaten practical MDI-QKD systems with commonly used devices, we hope to arouse attention to the source-side security and bring inspiration for designing new techniques and protocols.

2. HACKING STRATEGY

A. Tagging All Key Bits

As shown in the well-known Gottesman-Lo-Lutkenhaus-Preskill (GLLP) formula [28], the most common form of the final secret key rate is

$$R = p_1^{\mu} p_1^{\mu'} Y_{11}(1 - h(e_{11}^X)) - f Q_{\mu\mu'}^Z h(E_{\mu\mu'}^Z), \tag{1}$$

where $Q_{\mu\mu'}^Z$ is the raw key bit rate, $E_{\mu\mu'}^Z$ is the quantum bit error rate, $p_n^{\lambda} = e^{-\lambda}\lambda^n/n!$ denotes the n-photon probability when intensity (mean photon number) is λ, Y_{11} is the single-photon-pulse-pair yield of ZZ basis (both Alice and Bob select Z basis), e_{11}^X denotes single-photon-pulse-pair error rate of XX basis (both Alice and Bob select X basis), $h(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary Shannon entropy function, and f is the error correction efficiency. Equation (1) can also be revised as

$$\begin{aligned} R &= Q_{\mu\mu'}^Z \left[(1-\Delta)(1-h(e_{11}^X)) - f h(E_{\mu\mu'}^Z)\right] \\ &= Q_{\mu\mu'}^Z \left\{1 - \left[\Delta + (1-\Delta)h(e_{11}^X)\right] - f h(E_{\mu\mu'}^Z)\right\} \\ &= Q_{\mu\mu'}^Z \left[1 - \delta - f h(E_{\mu\mu'}^Z)\right], \end{aligned} \tag{2}$$

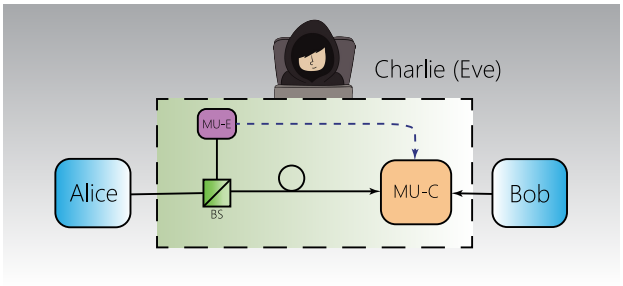


Fig. 1. Schematic diagram of our tagged method. MU-E, the measurement unit for projecting Alice's states on the Z basis; MU-C, the measurement unit for BSM; BS, beam splitter.

where $\Delta = 1 - p_1^{\mu} p_1^{\mu'} Y_{11}/Q_{\mu\mu'}^Z$ denotes the proportion of the key bits from multi-photon pulses in the raw key bits; $\delta = \Delta + (1-\Delta)h(e_{11}^X)$ is the information leakage. In the QKD procedure, Alice and Bob estimate the lower bound of Y_{11} (equivalent to estimating the upper bound of Δ) and the upper bound of e_{11}^X by the decoy-state method. The estimated information leakage is expressed as

$$\bar{\delta} = \bar{\Delta} + (1-\bar{\Delta})h(\bar{e}_{11}^X), \tag{3}$$

where the overline and underline denote the upper bound and lower bound of a variable, respectively. We define κ_a and κ_e as Alice and Eve's raw key bit, respectively, τ as the overlap between κ_a and κ_e, and $\hat{\delta} = 1 - h(1-\tau)$ as the practical information leakage. The key point of the protocol security is that the estimation satisfies

$$\bar{\delta} \geq \hat{\delta}. \tag{4}$$

Based on the above expressions, we can conclude that the key point of a successful hacking is misleading Alice and Bob to underestimate $\bar{\delta}$ to break the Eq. (4). Especially, if $\bar{\delta} < \hat{\delta}$ and $\tau = 1$, all final key bits are obtained by the unreliable measurement unit (named Eve in the following).

Our hacking strategy consists of tagging all key bits and exploiting the modulation error. The first part is illustrated in Fig. 1, Eve divides Alice's pulses into two parts, measures one of the parts on the Z basis by her MU-E for tagging the raw key bits, and sends the other part to her MU-C for Bell-state measurement (BSM). If $|0\rangle$ or $|1\rangle$ is measured by MU-E, Eve records the corresponding result and announces her BSM result normally. If nothing is measured by MU-E, Eve always announces failure. This method does not bring large errors in any bases (see Supplement 1) and can tag all raw key bits in the Z basis round. Noting that MDI-QKDs only generate key bits by the ZZ basis, Eve has a chance to obtain all final key bits if she can mislead Alice and Bob to break Eq. (4).

Using the tagged method alone can tag all raw key bits but cannot get any information about the final keys due to the protection of the decoy-state method. So the second part is exploiting the modulation errors to make the decoy-state method lose its power. The modulation errors can come from the intrinsic imperfections of practical devices such as the correlated deviation [18,19,67–71], random intensity fluctuation [69,72,73], and inhomogeneous electrical properties [65,74–76]; the errors can also be actively introduced by some recently proposed techniques such as the laser-seeding attack [17,20] and laser-damage attack [21,23]. Besides, we also gave a theoretical model to describe the relationship

between the estimated information leakage $\bar{\delta}$ and the modulation errors to show what error is in Eve's favor (see Supplement 1).

B. Actively Introducing Modulation Error

To indicate the source-side security is not negligible, we further develop an induced-photorefractive technique to actively introduce modulation errors. The induced-photorefractive technique combining with the theoretical model in Supplement 1 allows Eve to introduce modulation errors according to her optimal strategy. Besides, the previous techniques [17,20,21,23,24] such as the laser-damage attack [21] proposed by Huang *et al.* can also help Eve to actively introduce modulation errors.

The key idea of the induced-photorefractive technique is to exploit the photorefractive effect described as follows: under the illumination of the attack laser, photogenerated charges are generated and create an inhomogeneous space charge field; this field modulates the refractive index of the lithium niobate waveguide through the Pockels effect [78–82] and proceeds to affect the working states of the lithium-niobate-based devices. We note that the photorefractive effect is widespread in almost all lithium niobate devices [78,83,84] including the lithium niobate-based Mach–Zehnder interferometer (MZI), which is the most widely employed intensity modulator (IM) for decoy-state modulation in QKD systems [39,49–64,85–90]. Therefore, a DC-bias deviation can be actively introduced by injecting an irradiation laser from the quantum channel into Alice's IM due to the photorefractive effect.

When we inject an irradiation laser inversely into a lithium niobate-based MZI, the input laser is equally split into the two paths. However, due to the wavelength dependence of the splitter in the MZI, its splitting ratio at the irradiation wavelength deviates from 0.5, which causes different irradiation intensities between these two paths. Besides, considering that the fabrication errors and the applied electric fields during the working of these two paths are both different, the photorefractive effect on these two arms is unbalanced. As shown in Fig. 2, the unbalanced photorefractive effect introduces the DC-bias deviation [78] in the voltage modulation curve of the IM, which changes all modulation points of the decoy states so that the modulation error is introduced.

The modulation error can be quantitatively controlled by taking advantage of the property that the photorefractive effect is related to the wavelength, power, and injection duration of the irradiated laser [78,83,91]. Eve can inject a periodic irradiation pulse train into Alice's IM from the quantum channel, and the working point of the IM would be freely and accurately controlled by adjusting the duty ratio and injection time of the periodic irradiation pulse train.

We simulated the hacking performance of our induced-photorefractive technique and the hacking performance of employing the induced-photorefractive technique and laser damage simultaneously (the details are shown in Supplement 1). The simulation is based on the condition that Eve has successfully tagged all raw key bits by our tagged method ($\tau = 1$) so that Eve can obtain all secret keys as long as the final key rate is not 0 (equivalent to $\bar{\delta} + fh(E^{Z}_{\mu\mu'}) < 1$). The simulation results are shown in Fig. 3. We can find that increasing the DC-bias deviation by the induced-photorefractive technique and amplifying the output intensity by the laser-damage attack can efficiently mislead users to underestimate their $\bar{\delta}$. Especially, the induced-photorefractive technique can mislead users without the laser-damage attack, and the laser damage can make the induced-photorefractive technique more dangerous.

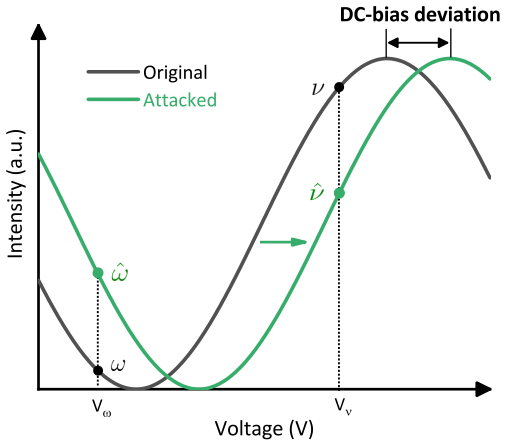


Fig. 2. Response curve of a lithium niobate-based MZI before and after injecting the irradiation laser. The black and green lines denote the response curve before and after the attack, respectively. Before the irradiation laser injection, when Alice (Bob) loads voltage V_ν and V_ω on her (his) IM, the expected intensities ν and ω are prepared, respectively. However, the irradiation laser injection introduces an additional bias so the practically prepared intensities become $\hat{\nu}$ and $\hat{\omega}$, respectively.

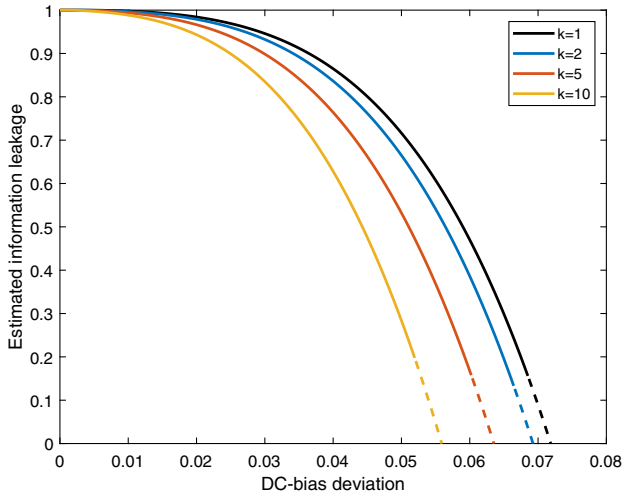


Fig. 3. Simulation of the hacking performance of actively introducing modulation error. The x axis denotes $\theta/2\pi$, where θ is the DC-bias deviation rooted in the photorefractive effect. The y axis denotes Alice's estimated $\bar{\delta}$ on the condition that Eve has tagged all raw key bits ($\tau = 1$). The black line denotes that Eve introduces modulation error only with our induced-photorefractive technique. The blue, red, and yellow lines denote the intensity amplifications k due to the laser-damage attack, which are 2, 5, and 10, respectively. The dashed line parts denote Alice's $\bar{\Delta} < 0$, where the users may terminate their protocol.

3. PRACTICAL HACKING EXPERIMENT

We demonstrated our hacking strategy by experimentally hacking a time-bin coding MDI-QKD system. The source side of the MDI-QKD system consists of a laser module to generate a phase randomized pulse train with 500 ps pulse width and 40 ns pulse interval, an encoding module to randomly prepare the four BB84 states [27], a decoy-state module to randomly modulate several

Table 1. Optimized Decoy-State Parameters

μ	ν	ω	P_μ	P_ν	P_ω
0.307	0.246	0.069	41.6%	10.0%	33.6%

different intensities, and an attenuator (ATT) to attenuate the output intensity to a single-photon level. Alice and Bob's pulses each pass through a 5 km fiber and are received by the measurement unit Charlie who performs the BSM and announces a success event if $|\psi^-\rangle = (|0\rangle|1\rangle - |1\rangle|0\rangle)/\sqrt{2}$ is measured. According to Charlie's announcement, Alice and Bob record their raw key bits κ_a and κ_b at their local FPGA-based driver board (PXIe-6547, National Instruments) and computer.

Before inserting our hacking devices, we first tested our MDI-QKD system with a four-intensity decoy-state method [46,54]. The four intensities μ, ν, ω, o and the corresponding probabilities P_λ ($\lambda \in \{\mu, \nu, \omega, o\}$) are optimized to maximize the final key rate. The optimized values are listed in Table 1. We ran the system to collect 10^{11} pulse pairs for parameter estimation. The observed raw key rate $Q^Z_{\mu\mu}$ and error rate $E^Z_{\mu\mu}$ are 2.89×10^{-4} and 0.0179, respectively. The estimated information leakage $\bar{\delta}$ is 80.7%. The final key rate $R = 2.14 \times 10^{-6}$.

Then we inserted our hacking device to demonstrate hacking a practical MDI-QKD system. As illustrated in Fig. 4, the hacking device is inserted in Alice's quantum channel and controlled by a hostile measurement unit Eve. Alice's pulses are fed to the hacking device in the forward direction. The pulses are first split by an 1:3 BS (25% transmission and 75% reflection). The transmission part is sent to the MU-C for BSM, and the reflected part goes through a 99:1 BS (99% transmission and 1% reflection) to be measured by MU-E that consists of an 1:1 BS and two SPDs named SPD-E0 and SPD-E1. The MU-E outputs 0, 1, or Ø when Alice's signal is projected to $|0\rangle$, $|1\rangle$, or nothing detected. Especially when both of the SPDs click, the MU-E randomly outputs 0 or 1. *The results of the MU-E are recorded as Eve's raw key bit κ_e by a local FPGA-based driver board (PXIe-6547, National Instruments) and a computer*. To obtain all key bits, the MU-C normally announces the BSM result when the MU-E outputs 0 or 1 and always announces failure when the MU-E outputs Ø. In the backward direction, a 405 nm ultraviolet laser diode is employed to generate a pulse train to introduce the photorefractive effect. The ultraviolet pulses reversely pass through the 99:1 BS, the 1:3 BS, and Alice's ATT and finally enter into Alice's IM to introduce the photorefractive effect.

The demonstration of our hacking is carried out in two steps. First, we close the ultraviolet laser and intentionally load a wrong DC-bias on Alice's IM. The purpose of this step is to verify the accuracy of our theoretical analysis (see Supplement 1) and demonstrate the harm of the intensity modulation error. We also reduce Alice's ATT from 51.8 dB to 41.8 dB to mimic the laser-damage attack [21], which has indicated that high-power lasers could damage the ATT and reduce its attenuation. We emphasize that the laser damage is not necessary for our hacking strategy, but it can make our demonstration easier. In our demonstration, Alice and Bob always process their data with a data size of $N = 10^{11}$. As shown in Fig. 5, when we intentionally added a DC-bias with a 0.44 rad deviation (compared with the correct DC-bias), Alice and Bob estimated their information leakage as $\bar{\delta} = 79.9\%$ and generated key bits with the final key rate of 1.01×10^{-6} bit per pulse. We compared κ_a and κ_e, and found that the overlap τ reaches 99.886% so that the practical information leakage $\hat{\delta}$ reaches 98.47%. Similarly, when we added a DC-bias with a 0.49 rad deviation, the estimated $\bar{\delta} = 78.2\%$, and the final key rate $R = 2.46 \times 10^{-6}$ bit per pulse even though the overlap τ reaches 99.887% and the practical information leakage $\hat{\delta}$ reaches 98.57%. It is obvious that the condition $\bar{\delta} > 1 - h(1 - \tau)$ is not satisfied anymore and Eve can obtain almost all final keys since the overlap is too large. The reason of $\tau \neq 1$ is the bit errors from dark counts and afterpulses in MU-C.

Then we add the correct DC-bias on Alice's IM and open the ultraviolet laser to actively introduce DC-bias deviation. The ultraviolet laser of 405 nm wavelength is employed by Eve since the photorefractive effect of MZI is more sensitive to the 405 nm laser [91] to make it easier to introduce DC-bias deviation with the 405 nm wavelength. Besides, the 405 nm laser can pass some widely employed ATTs (such as the buildout ATT) with less loss. Before the injection, the ultraviolet laser should reversely pass through a 1:99 BS and a 1:3 BS in Eve's hacking setup and an ATT in Alice's export. We measured that the 405 nm laser only

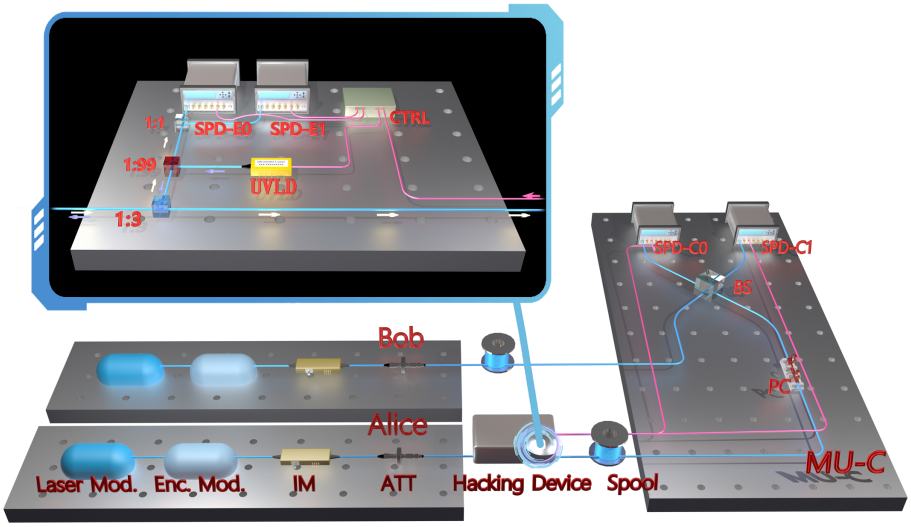


Fig. 4. Experimental setup of our hacking demonstration. The hacking device is inserted in the quantum channel, and the details are shown in the subfigure. Laser Mod., laser module; Enc. Mod., encoding Module; IM, intensity modulator; ATT, attenuator; BS, beam splitter; SPD, single-photon detector; PC, polarization controller; UVLD, ultraviolet laser diode; CTRL, controller. The blue lines and pink lines denote fibers and cables, respectively.

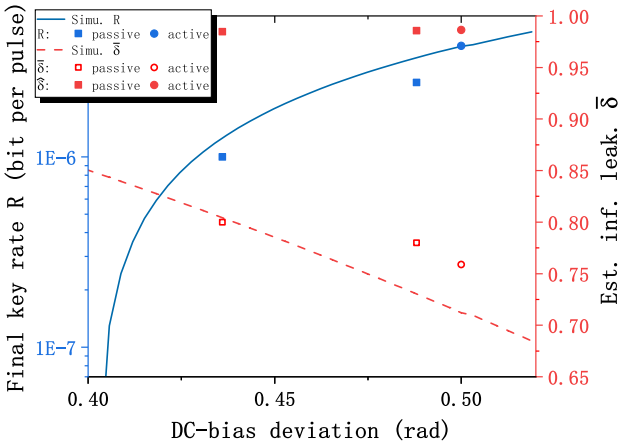


Fig. 5. Simulation and experimental results of the estimated information leakage $\bar{\delta}$, practical information leakage $\hat{\delta}$, and final key rate R (insecure) as a function of the DC-bias deviation θ. The blue solid line and red dash line denote the simulated final key rate and the estimated information leakage as a function of the DC-bias deviation, respectively. The squares denote the experimental results in the cases when Eve passively exploits Alice's modulation flaws, and the circles denote the results in the cases when Eve actively introduces modulation errors. The blue solid shapes denote the experimental final key rates. The red solid shapes denote the experimental observed $\hat{\delta}$, and the red hollow shapes denote the experimental $\bar{\delta}$.

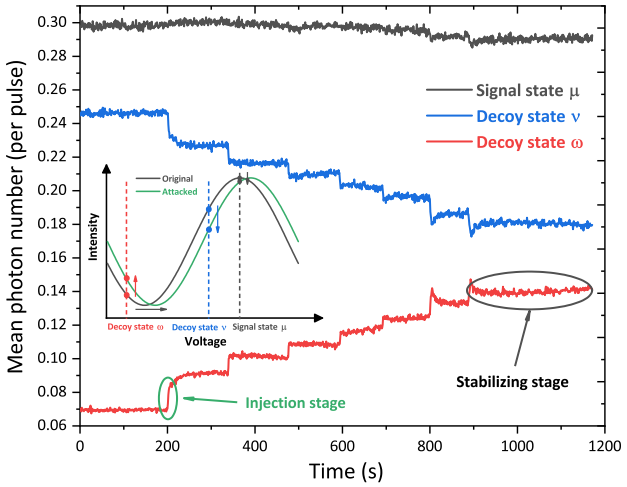


Fig. 6. Practically prepared intensities during ultraviolet laser injection. The black, blue, and red lines denote the variation of the μ, ν, and ω in the hacking process, respectively. The areas where the intensities change fast belong to the injection stages. The areas where the intensities remain stable belong to the stabilizing stages. The variation of signal state μ is inconspicuous because the signal state is insensitive to modulation errors [68,69]. The subfigure shows the IM's response curves. The black and green lines denote the response curves before and after the injection, respectively. We can find that the hacking introduces a bias deviation, which leads to an unexpected intensity change.

has a 44 dB loss when reversely passing these three devices while the 1550 nm laser has a 67.9 dB loss. Noting that commercial 405 nm lasers can reach a power of several hundred milliwatts or even higher, it is easy to induce the photorefractive effect from the outside. In our experiment, Eve modulated a pulse train with 17.6 milliwatts peak power, and the peak power reaching the IM is 0.7 microwatts, which is sufficient to complete our attack.

The photorefractive effect can be accurately controlled by adjusting the duty ratio, injection duration, and frequency of the irradiation pulse train. In our experiment, we controlled the duty ratio and injection duration to precisely introduce bias deviation as we want. The index change precision depends on the duty ratio, and the total index change depends on both of these two parameters. The induced-photorefractive technique is divided into two stages, namely, the "injection stage" and the "stabilizing stage." In the injection stage, the irradiation pulse train is injected into Alice's IM to change the index and, thus, introduce the bias deviation we want. In the stabilizing stage, we fine-tune the duty ratio to counteract the decay behavior of the photorefractive effect [92] for stabilizing the introduced bias deviation. The process of the laser injection demonstration is shown in Fig. 6, and the hacking process is separated into several sets of injection and stabilizing stages. The areas where the intensities change fast belong to the injection stages. In these stages, the duty ratio is set as 10% to induce a rapid intensity change. The areas where the intensities remain stable belong to the stabilizing stages. In these stages, we reduce and fine-tune the duty ratio to maintain the bias deviation as we (Eve) want. We emphasize that it is not necessary to divide the process into so many stages. The main purpose of the several stabilizing stages is to show that we have the ability to introduce arbitrary bias deviation as we want. By fine-tuning the duty ratio and the injection duration of the ultraviolet pulse train, we successfully introduced a 0.50 rad DC-bias deviation. As shown in Fig. 5, in this case, the estimated $\bar{\delta} = 75.9\%$, and the final key rate is 3.8×10^{-6} while we observed that $\tau = 99.88\%$ and $\hat{\delta} = 98.66\%$ so that Eve can obtain almost all final key bits.

4. DISCUSSION

According to our hacking strategy, we would like to make some recommendations on the design of future QKD systems with higher security. (1) An intensity monitor is a potential solution for avoiding the internal errors and the external attacks. By monitoring the output intensities, the users may find that the DC-bias has deviated from the corrected value. Especially, the users may employ some algorithms to automatically calibrate the DC-bias error. However, the intensity monitor cannot fix the loophole once for all because Eve can hack the monitor [26] and she can also develop a special ultraviolet laser injection strategy to cheat the algorithm. (2) Comparing the detection statistics with the simulation value is also a potential solution because users may notice a difference and abort the protocol. However, our simulation results in Fig. 3 have indicated that small DC-bias deviation can also bring risks. The users may not distinguish the deviations coming from inaccurate simulation models, environmental disturbances, and Eve's induced-photorefractive technique. (3) A bias-free IM [68] can efficiently protect QKD systems from this loophole; however, it is still in the research stage, and the commercial MZI-IM is still the mainstream for decoy-state modulation. (4) In addition, protective devices such as bandpass filters, isolators, and circulators can protect practical systems from this loophole [22,93]. Employing these devices can effectively protect against the induced-photorefractive technique. However, the protective devices are not absolutely secure because some previous works indicated that these devices can also be damaged by Eve [23,24]. (6) Finally, we suggest that the devices that are susceptible to Eve's interference should stay as far away from Alice's exit port as possible [94]. Placing insensitive

optical passive devices near the exit port can efficiently mitigate the impact of the induced-photorefractive technique.

In summary, we proposed a hacking strategy accompanied by a theoretical analysis to show the importance of the source-side security. Our work proves that the modulation errors can practically threaten the system security since our experiment has indeed obtained all final key bits with the commonly used devices. In Supplement 1, we further present a more general hacking strategy that also threatens other decoy-state protocols.

The detection side is usually regarded as the most fragile part of QKD systems, and the source-side threats are relatively underestimated. In this work, the source-side security is proved to be non-negligible, and we trust that more countermeasures against the source-side loopholes would be unearthed in the future. We hope this work can make people pay more attention to the source-side security and bring more inspiration for developing new protocols and techniques to improve the source-side security.

Funding. National Key Research and Development Program of China (2018YFA0306400); National Natural Science Foundation of China (62271463, 62171424, 62105318); China Postdoctoral Science Foundation (2022M723064, 2021M693098).

Acknowledgment. This work was partially carried out at the USTC Center for Micro and Nanoscale Research and Fabrication.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. C. H. Bennett and G. Brassard, "Quantum cryptography: public key distribution and coin tossing," in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing* (IEEE, 1984), pp. 175–179.
2. H.-K. Lo and H. F. Chau, "Unconditional security of quantum key distribution over arbitrarily long distances," Science **283**, 2050–2056 (1999).
3. P. W. Shor and J. Preskill, "Simple proof of security of the BB84 quantum key distribution protocol," Phys. Rev. Lett. **85**, 441–444 (2000).
4. V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, "The security of practical quantum key distribution," Rev. Mod. Phys. **81**, 1301–1350 (2009).
5. R. Renner, "Security of quantum key distribution," Int. Symp. Inf. Theory **6**, 1–127 (2008).
6. B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, "Time-shift attack in practical quantum cryptosystems," arXiv, arXiv:quant-ph/0512080 (2005).
7. V. Makarov, "Controlling passively quenched single photon detectors by bright light," New J. Phys. **11**, 065003 (2009).
8. V. Makarov, A. Anisimov, and J. Skaar, "Effects of detector efficiency mismatch on security of quantum cryptosystems," Phys. Rev. A **74**, 022313 (2006).
9. 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).
10. 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**, 686–689 (2010).
11. 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).
12. S. Sajeed, A. Huang, S. Sun, F. Xu, V. Makarov, and M. Curty, "Insecurity of detector-device-independent quantum key distribution," Phys. Rev. Lett. **117**, 250505 (2016).
13. 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).
14. N. Lütkenhaus and M. Jahma, "Quantum key distribution with realistic states: photon-number statistics in the photon-number splitting attack," New J. Phys. **4**, 44 (2002).
15. W.-T. Liu, S.-H. Sun, L.-M. Liang, and J.-M. Yuan, "Proof-of-principle experiment of a modified photon-number-splitting attack against quantum key distribution," Phys. Rev. A **83**, 042326 (2011).
16. L. O. Mailloux, D. D. Hodson, M. R. Grimaila, R. D. Engle, C. V. Mclaughlin, and G. B. Baumgartner, "Using modeling and simulation to study photon number splitting attacks," IEEE Access **4**, 2188–2197 (2016).
17. 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).
18. M. Pereira, G. Kato, A. Mizutani, M. Curty, and K. Tamaki, "Quantum key distribution with correlated sources," Sci. Adv. **6**, eaaz4487 (2020).
19. V. Zapatero, Á. Navarrete, K. Tamaki, and M. Curty, "Security of quantum key distribution with intensity correlations," arXiv, arXiv:2105.11165 (2021).
20. X.-L. Pang, A.-L. Yang, C.-N. Zhang, J.-P. Dou, H. Li, J. Gao, and X.-M. Jin, "Hacking quantum key distribution via injection locking," Phys. Rev. Appl. **13**, 034008 (2020).
21. 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).
22. G. Zhang, I. W. Primaatmaja, J. Y. Haw, X. Gong, C. Wang, and C. C. W. Lim, "Securing practical quantum communication systems with optical power limiters," PRX Quantum **2**, 030304 (2021).
23. 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).
24. H. Tan, W.-Y. Zhang, L. Zhang, W. Li, S.-K. Liao, and F. Xu, "External magnetic effect for the security of practical quantum key distribution," Quantum Sci. Technol. **7**, 045008 (2022).
25. P. R. Smith, D. G. Marangon, M. Lucamarini, Z. L. Yuan, and A. J. Shields, "Out-of-band electromagnetic injection attack on a quantum random number generator," Phys. Rev. Appl. **15**, 044044 (2021).
26. S. Sajeed, I. Radchenko, S. Kaiser, J.-P. Bourgoin, A. Pappa, L. Monat, M. Legré, and V. Makarov, "Attacks exploiting deviation of mean photon number in quantum key distribution and coin tossing," Phys. Rev. A **91**, 032326 (2015).
27. H.-K. Lo, M. Curty, and B. Qi, "Measurement-device-independent quantum key distribution," Phys. Rev. Lett. **108**, 130503 (2012).
28. D. Gottesman, H.-K. Lo, N. Lutkenhaus, and J. Preskill, "Security of quantum key distribution with imperfect devices," in *International Symposium on Information Theory (ISIT)* (IEEE, 2004), pp. 136.
29. W.-Y. Hwang, "Quantum key distribution with high loss: toward global secure communication," Phys. Rev. Lett. **91**, 057901 (2003).
30. X.-B. Wang, "Beating the photon-number-splitting attack in practical quantum cryptography," Phys. Rev. Lett. **94**, 230503 (2005).
31. H.-K. Lo, X. Ma, and K. Chen, "Decoy state quantum key distribution," Phys. Rev. Lett. **94**, 230504 (2005).
32. X.-B. Wang, "Decoy-state protocol for quantum cryptography with four different intensities of coherent light," Phys. Rev. A **72**, 012322 (2005).
33. X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, "Practical decoy state for quantum key distribution," Phys. Rev. A **72**, 012326 (2005).
34. Q. Wang, W. Chen, G. Xavier, M. Swillo, T. Zhang, S. Sauge, M. Tengner, Z.-F. Han, G.-C. Guo, and A. Karlsson, "Experimental decoy-state quantum key distribution with a sub-Poissionian heralded single-photon source," Phys. Rev. Lett. **100**, 090501 (2008).
35. M. Lucamarini, Z. L. Yuan, J. F. Dynes, and A. J. Shields, "Overcoming the rate–distance limit of quantum key distribution without quantum repeaters," Nature **557**, 400–403 (2018).
36. X. Ma, P. Zeng, and H. Zhou, "Phase-matching quantum key distribution," Phys. Rev. X **8**, 031043 (2018).
37. X.-B. Wang, Z.-W. Yu, and X.-L. Hu, "Twin-field quantum key distribution with large misalignment error," Phys. Rev. A **98**, 062323 (2018).
38. R. Wang, Z.-Q. Yin, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han, "Round-robin-differential-phase-shift quantum key distribution with monitoring signal disturbance," Opt. Lett. **43**, 4228–4231 (2018).
39. A. Boaron, G. Boso, D. Rusca, and C. Vulliez, "Secure quantum key distribution over 421 km of optical fiber," Phys. Rev. Lett. **121**, 190502 (2018).

40. F.-Y. Lu, Z.-Q. Yin, G.-J. Fan-Yuan, R. Wang, H. Liu, S. Wang, W. Chen, D.-Y. He, W. Huang, B.-J. Xu, G.-C. Guo, and Z.-F. Han, "Efficient decoy states for the reference-frame-independent measurement-device-independent quantum key distribution," Phys. Rev. A **101**, 052318 (2020).
41. C. C. W. Lim, M. Curty, N. Walenta, F. Xu, and H. Zbinden, "Concise security bounds for practical decoy-state quantum key distribution," Phys. Rev. A **89**, 022307 (2014).
42. X.-B. Wang, "Three-intensity decoy-state method for device-independent quantum key distribution with basis-dependent errors," Phys. Rev. A **87**, 012320 (2013).
43. X. Ma and M. Razavi, "Alternative schemes for measurement-device-independent quantum key distribution," Phys. Rev. A **86**, 062319 (2012).
44. F. Xu, M. Curty, B. Qi, and H.-K. Lo, "Practical aspects of measurement-device-independent quantum key distribution," New J. Phys. **15**, 113007 (2013).
45. Z.-W. Yu, Y.-H. Zhou, and X.-B. Wang, "Statistical fluctuation analysis for measurement-device-independent quantum key distribution with three-intensity decoy-state method," Phys. Rev. A **91**, 032318 (2015).
46. Y.-H. Zhou, Z.-W. Yu, and X.-B. Wang, "Making the decoy-state measurement-device-independent quantum key distribution practically useful," Phys. Rev. A **93**, 042324 (2016).
47. T. F. da Silva, D. Vitoreti, G. B. Xavier, G. C. do Amaral, G. P. Temporão, and J. P. von der Weid, "Proof-of-principle demonstration of measurement-device-independent quantum key distribution using polarization qubits," Phys. Rev. A **88**, 052303 (2013).
48. Y. Liu, T.-Y. Chen, L.-J. Wang, H. Liang, G.-L. Shentu, J. Wang, K. Cui, H.-L. Yin, N.-L. Liu, and L. Li, "Experimental measurement-device-independent quantum key distribution," Phys. Rev. Lett. **111**, 130502 (2013).
49. 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, D.-X. Yang, Z. Wang, H. Liang, Z. Zhang, N. Zhou, X. Ma, T.-Y. Chen, Q. Zhang, and J.-W. Pan, "Measurement-device-independent quantum key distribution over 200 km," Phys. Rev. Lett. **113**, 190501 (2014).
50. S. Pirandola, C. Ottaviani, G. Spedalieri, C. Weedbrook, S. L. Braunstein, S. Lloyd, T. Gehring, C. S. Jacobsen, and U. L. Andersen, "High-rate measurement-device-independent quantum cryptography," Nat. Photonics **9**, 397–402 (2015).
51. C. Wang, X.-T. Song, Z.-Q. Yin, S. Wang, W. Chen, C.-M. Zhang, G.-C. Guo, and Z.-F. Han, "Phase-reference-free experiment of measurement-device-independent quantum key distribution," Phys. Rev. Lett. **115**, 160502 (2015).
52. Y.-L. Tang, H.-L. Yin, Q. Zhao, H. Liu, X.-X. Sun, M.-Q. Huang, W.-J. Zhang, S.-J. Chen, L. Zhang, and L.-X. You, "Measurement-device-independent quantum key distribution over untrustful metropolitan network," Phys. Rev. X **6**, 011024 (2016).
53. L. C. Comandar, M. Lucamarini, B. Fröhlich, J. F. Dynes, A. W. Sharpe, S. W.-B. Tam, Z. L. Yuan, R. V. Penty, and A. J. Shields, "Quantum key distribution without detector vulnerabilities using optically seeded lasers," Nat. Photonics **10**, 312–315 (2016).
54. H.-L. Yin, T.-Y. Chen, Z.-W. Yu, H. Liu, L.-X. You, Y.-H. Zhou, S.-J. Chen, Y. Mao, M.-Q. Huang, and W.-J. Zhang, "Measurement-device-independent quantum key distribution over a 404 km optical fiber," Phys. Rev. Lett. **117**, 190501 (2016).
55. G. L. Roberts, M. Lucamarini, Z. L. Yuan, J. F. Dynes, L. C. Comandar, A. W. Sharpe, A. J. Shields, M. Curty, V. Puthoor, and E. Andersson, "Experimental measurement-device-independent quantum digital signatures," Nat. Commun. **8**, 1098 (2017).
56. C. Wang, Z.-Q. Yin, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han, "Measurement-device-independent quantum key distribution robust against environmental disturbances," Optica **4**, 1016–1023 (2017).
57. H. Liu, J. Wang, H. Ma, and S. Sun, "Polarization-multiplexing-based measurement-device-independent quantum key distribution without phase reference calibration," Optica **5**, 902–909 (2018).
58. H. Liu, W. Wang, K. Wei, X.-T. Fang, L. Li, N.-L. Liu, H. Liang, S.-J. Zhang, W. Zhang, and H. Li, "Experimental demonstration of high-rate measurement-device-independent quantum key distribution over asymmetric channels," Phys. Rev. Lett. **122**, 160501 (2019).
59. H. Semenenko, P. Sibson, A. Hart, M. G. Thompson, J. G. Rarity, and C. Erven, "Chip-based measurement-device-independent quantum key distribution," Optica **7**, 238–242 (2020).
60. K. Wei, W. Li, H. Tan, Y. Li, H. Min, W.-J. Zhang, H. Li, L. You, Z. Wang, and X. Jiang, "High-speed measurement-device-independent quantum key distribution with integrated silicon photonics," Phys. Rev. X **10**, 031030 (2020).
61. R. I. Woodward, Y. S. Lo, M. Pittaluga, M. Minder, T. K. Paraïso, M. Lucamarini, Z. L. Yuan, and A. J. Shields, "Gigahertz measurement-device-independent quantum key distribution using directly modulated lasers," npj Quantum Inf. **7**, 58 (2021).
62. X.-Y. Zhou, H.-J. Ding, M.-S. Sun, S.-H. Zhang, J.-Y. Liu, C.-H. Zhang, J. Li, and Q. Wang, "Reference-frame-independent measurement-device-independent quantum key distribution over 200 km of optical fiber," Phys. Rev. Appl. **15**, 064016 (2021).
63. F.-Y. Lu, Z.-H. Wang, Z.-Q. Yin, S. Wang, R. Wang, G.-J. Fan-Yuan, X.-J. Huang, D.-Y. He, W. Chen, and Z. Zhou, "Unbalanced-basis-misalignment-tolerant measurement-device-independent quantum key distribution," Optica **9**, 886–893 (2022).
64. G.-J. Fan-Yuan, F.-Y. Lu, S. Wang, Z.-Q. Yin, D.-Y. He, W. Chen, Z. Zhou, Z.-H. Wang, J. Teng, and G.-C. Guo, "Robust and adaptable quantum key distribution network without trusted nodes," Optica **9**, 812–823 (2022).
65. J. P. Salvestrini, L. Guilbert, M. Fontana, M. Abarkan, and S. Gille, "Analysis and control of the DC drift in $LiNBo_3$-based Mach–Zehnder modulators," J. Lightwave Technol. **29**, 1522–1534 (2011).
66. H. Nagata and K. Kiuchi, "Temperature dependence of DC drift of Ti:$LiNbO_3$ optical modulators with sputter deposited SiO_2 buffer layer," J. Appl. Phys. **73**, 4162–4164 (1993).
67. K. I. Yoshino, M. Fujiwara, K. Nakata, T. Sumiya, T. Sasaki, M. Takeoka, M. Sasaki, A. Tajima, M. Koashi, and A. Tomita, "Quantum key distribution with an efficient countermeasure against correlated intensity fluctuations in optical pulses," npj Quantum Inf. **4**, 8 (2018).
68. G. L. Roberts, M. Pittaluga, M. Minder, M. Lucamarini, J. F. Dynes, Z. L. Yuan, and A. J. Shields, "Patterning-effect mitigating intensity modulator for secure decoy-state quantum key distribution," Opt. Lett. **43**, 5110–5113 (2018).
69. F.-Y. Lu, X. Lin, S. Wang, G.-J. Fan-Yuan, P. Ye, R. Wang, Z.-Q. Yin, D.-Y. He, W. Chen, and G.-C. Guo, "Intensity modulator for secure, stable, and high-performance decoy-state quantum key distribution," npj Quantum Inf. **7**, 75 (2021).
70. X. Kang, F.-Y. Lu, S. Wang, J.-L. Chen, Z.-H. Wang, Z.-Q. Yin, D.-Y. He, W. Chen, G.-J. Fan-Yuan, and G.-C. Guo, "Patterning-effect calibration algorithm for secure decoy-state quantum key distribution," J. Lightwave Technol. **41**, 75–82 (2023).
71. A. Mizutani, G. Kato, K. Azuma, M. Curty, R. Ikuta, T. Yamamoto, N. Imoto, H.-K. Lo, and K. Tamaki, "Quantum key distribution with setting-choice-independently correlated light sources," npj Quantum Information **5**, 8 (2019).
72. X.-B. Wang, L. Yang, C.-Z. Peng, and J.-W. Pan, "Decoy-state quantum key distribution with both source errors and statistical fluctuations," New J. Phys. **11**, 075006 (2009).
73. F. Grasselli and M. Curty, "Practical decoy-state method for twin-field quantum key distribution," New J. Phys. **21**, 073001 (2019).
74. S. Yamada and M. Minakata, "DC drift phenomena in $LiNbO_3$ optical waveguide devices," Jpn. J. Appl. Phys. **20**, 733 (1981).
75. R. A. Becker, "Circuit effect in $LiNBo_3$ channel-waveguide modulators," Opt. Lett. **10**, 417–419 (1985).
76. S. Korotky and J. Veselka, "An RC network analysis of long term Ti:$LiNBo_3$ bias stability," J. Lightwave Technol. **14**, 2687–2697 (1996).
77. P. Ye, W. Chen, G.-W. Zhang, F.-Y. Lu, F.-X. Wang, G.-Z. Huang, S. Wang, D.-Y. He, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han, "Induced-photorefractive attack against quantum key distribution," arXiv, arXiv:2303.10885 (2023).
78. S. M. Kostritskii, "Photorefractive effect in $LiNBo_3$-based integrated-optical circuits at wavelengths of third telecom window," Appl. Phys. B **95**, 421–428 (2009).
79. P. Günter and J.-P. Huignard, *Photorefractive Materials and their Applications* (Springer, 2007), Vol. **114**.
80. M. Carrascosa, J. Villarroel, J. Carnicero, A. Garca-Cabañes, and J. M. Cabrera, "Understanding light intensity thresholds for catastrophic optical damage in $LiNBo_3$," Opt. Express **16**, 115–120 (2008).
81. M. M. Howerton and W. K. Burns, "Photorefractive effects in proton exchanged $LiTaO_3$ optical waveguides," J. Lightwave Technol. **10**, 142–149 (1992).

82. T. Fujiwara, R. Srivastava, X. Cao, and R. V. Ramaswamy, "Comparison of photorefractive index change in proton-exchanged and Ti-diffused $LiNBo_3$ waveguides," Opt. Lett. **18**, 346–348 (1993).
83. T. Fujiwara, S. Sato, and H. Mori, "Wavelength dependence of photorefractive effect in Ti-indiffused $LiNBo_3$ waveguides," Appl. Phys. Lett. **54**, 975–977 (1989).
84. F. Mondain, F. Brunel, X. Hua, E. Gouzien, A. Zavatta, T. Lunghi, F. Doutre, M. P. De Micheli, S. Tanzilli, and V. d'Auria, "Photorefractive effect in $LiNBo_3$-based integrated-optical circuits for continuous variable experiments," Opt. Express **28**, 23176–23188 (2020).
85. S. Wang, Z.-Q. Yin, W. Chen, D.-Y. He, X.-T. Song, H.-W. Li, L.-J. Zhang, Z. Zhou, G.-C. Guo, and Z.-F. Han, "Experimental demonstration of a quantum key distribution without signal disturbance monitoring," Nat. Photonics **9**, 832–836 (2015).
86. S. Wang, D.-Y. He, Z.-Q. Yin, F.-Y. Lu, C.-H. Cui, W. Chen, Z. Zhou, G.-C. Guo, and Z.-F. Han, "Beating the fundamental rate-distance limit in a proof-of-principle quantum key distribution system," Phys. Rev. X **9**, 021046 (2019).
87. X. Zhong, J. Hu, M. Curty, L. Qian, and H.-K. Lo, "Proof-of-principle experimental demonstration of twin-field type quantum key distribution," Phys. Rev. Lett. **123**, 100506 (2019).
88. J.-P. Chen, C. Zhang, Y. Liu, *et al.*, "Sending-or-not-sending with independent lasers: secure twin-field quantum key distribution over 509 km," Phys. Rev. Lett. **124**, 070501 (2020).
89. M. Pittaluga, M. Minder, M. Lucamarini, M. Sanzaro, R. I. Woodward, M.-J. Li, Z. Yuan, and A. J. Shields, "600-km repeater-like quantum communications with dual-band stabilization," Nat. Photonics **15**, 530–535 (2021).
90. S. Wang, Z.-Q. Yin, D.-Y. He, W. Chen, R.-Q. Wang, P. Ye, Y. Zhou, G.-J. Fan-Yuan, F.-X. Wang, and Y.-G. Zhu, "Twin-field quantum key distribution over 830-km fibre," Nat. Photonics **16**, 154–161 (2022).
91. T. Fujiwara, X. Cao, R. Srivastava, and R. V. Ramaswamy, "Photorefractive effect in annealed proton-exchanged $LiNBO_3$ waveguides," Appl. Phys. Lett. **61**, 743–745 (1992).
92. T. Fujiwara, H. Mori, and Y. Fujii, "Evolution and decay behavior of the photorefractive effect in Ti:$LiNbO_3$ optical waveguides," Ferroelectrics **95**, 133–137 (1989).
93. 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).
94. N. Jain and G. S. Kanter, "Upconversion-based receivers for quantum hacking-resistant quantum key distribution," Quantum Inform. Process. **15**, 2863–2879 (2016).