Acheva et al. *EPJ Quantum Technology* (2023) 10:22
https://doi.org/10.1140/epjqt/s40507-023-00178-x

EPJ Quantum Technology
a SpringerOpen Journal

EPJ.org

RESEARCH Open Access

Automated verification of countermeasure against detector-control attack in quantum key distribution

Polina Acheva[1,2,3*], Konstantin Zaitsev[1,2,3,4], Vladimir Zavodilenko[4,5], Anton Losev[4], Anqi Huang[6] and Vadim Makarov[1,4]

*Correspondence:
achevap17@yandex.ru
[1]Russian Quantum Center, Skolkovo, Moscow 121205, Russia
[2]Vigo Quantum Communication Center, University of Vigo, Vigo E-36310, Spain
Full list of author information is available at the end of the article

Abstract

Attacks that control single-photon detectors in quantum key distribution using tailored bright illumination are capable of eavesdropping the secret key. Here we report an automated testbench that checks the detector's vulnerabilities against these attacks. We illustrate its performance by testing a free-running detector that includes a rudimentary countermeasure measuring an average photocurrent. While our testbench automatically finds the detector to be controllable in a continuous-blinding regime, the countermeasure registers photocurrent significantly exceeding that in a quantum regime, thus revealing the attack. We then perform manually a pulsed blinding attack, which controls the detector intermittently. This attack is missed by the countermeasure in a wide range of blinding pulse durations and powers, still allowing to eavesdrop the key. We make recommendations for improvement of both the testbench and countermeasure.

1 Introduction

In recent years, quantum key distribution (QKD) [1] has attracted significant attention by both science and industry. This interest is based on its security guaranteed by the laws of quantum mechanics. However, differences exist between the theory of QKD and its practical realisation. They can be exploited by an eavesdropper Eve to steal secret information. For example, a laser pulse attenuated to contain less than one photon on average still sometimes contains two or more photons, which contradicts theoretical assumptions in QKD. A photon-number-splitting attack that exploits these multi-photon pulses was historically the first QKD loophole shown in the year 2000 [2]. Since then, over twenty different other loopholes have been discovered [3–27].

Most loopholes can be closed by additional countermeasures implemented in QKD components or postprocessing. The photon-number-splitting attack [2] can be closed by a decoy-state protocol [28]; Trojan-horse attack [4] and laser-seeding attack [20] can be closed by adding isolators to the optical scheme [29–31]; detector efficiency mismatch problems [7, 8, 11, 19] can be solved by both a proper device calibration and an update

Springer

to the theory [32]. However, some loopholes have not yet been fully solved, notably the detector control loophole [9, 10, 12–15, 22].

The single-photon detector (SPD) currently seems to be the most unsafe QKD system element. There are several attacks focusing on it. Some of them create and exploit mismatch in photon detection efficiency between two or more detectors in the receiver Bob. These include efficiency mismatch in the time domain [7, 8], wavelength [11, 19], spatial mode [21], and during a deadtime [33]. Other attacks control the detector deterministically while blinding it with bright light [9, 10, 13] or injecting bright pulses at the closing edge of a detector gate [14, 34] or in-between the gates [12]. The detector control attack was first proposed in 2009 [35] and found to be applicable to commercial QKD systems the following year [9]. A protection against the attacks on detectors is difficult because Bob has to receive all light from a transmission line with as low loss as possible. (In contrast, a sender Alice can be effectively isolated against attacks that inject light [29–31, 36].)

Several countermeasures against the bright-light attacks on detectors have been proposed [9, 37–50]. The most radical one is a measurement-device-independent (MDI) QKD scheme that eliminates the detectors, and thus all their vulnerabilities, from the secure equipment [38]. However, it is less convenient and more costly for commercial implementation than the standard QKD schemes. Other approaches vary in their maturity and effectiveness. An optical power meter at Bob's entrance with a classical threshold [9] is not fast enough and may overlook a pulsed blinding attack. A random-detector-efficiency patch [40] was shown to contain unrealistic assumptions on hardware after a careful investigation [22]. A measurement of coincidence click rates [45] and application of a random optical attenuation [48] are at a proof-of-principle stage and need further tests. Optical power limiters [49, 50] are not mature and sensitive enough to become a countermeasure. A more mature technology is the measurement of photodiode current to sense the blinding [37, 41], which is implemented in some commercial SPDs [41, 51]. However, its effectiveness as a countermeasure in QKD depends on implementation details and needs to be tested.

For a wide adoption of QKD as a data protection technology, it needs to have certification [52]. The certification standards for QKD include tests for the quality of countermeasures against the known vulnerabilities [53]. Formalising and automating the testing procedure would both simplify its application in a certification lab and reduce human factors. We are therefore developing an automated testbench and algorithm that tests SPDs against the bright-light attacks.

The contribution of this paper is two-fold. First, we report an automated setup for testing SPDs against the bright-light attacks. Second, we apply this setup to an SPD with the current-measurement countermeasure. The latter is proven effective against the attack that uses continuous-wave (cw) blinding [41]. This is confirmed with our automated testbench. However, this countermeasure might miss an attack that blinds the detector intermittently by light pulses [41, 54]. We probe experimentally in a manual regime the limits of the existing countermeasure implementation. We then make recommendations for both countermeasure and testbench improvement that would hopefully make them complete and ready for certification.

The development of a complete countermeasure is nontrivial. In more than ten years elapsed since the discovery of these attacks [9], no countermeasure for non-MDI QKD systems has been independently tested and certified as secure. Although it is obvious in

the hindsight that the rudimentary countermeasure we test here is insufficient, this is not clear to an engineer designing it without the help of independent testers. We have chosen to focus our development on this type of countermeasure, because it is the simplest and cheapest to implement (being just some extra electronics in the SPD) and it has a potential to close this class of loopholes. However our test methodology may be adopted to other types of countermeasures.

The paper is organised as follows. In Sect. 2 we describe the testbench setup, its software, and the detector under test. In Sect. 3 we report experimental results and simulate the attack. We discuss and conclude in Sect. 4.

2 Experimental setup

The SPD control attack using bright light can be realised in several ways. We distinguish three main types of it: continuous blinding [9, 35], pulsed blinding [13, 55], and after-gate attack [12]. Under the continuous blinding attacks, a cw laser light is applied to the SPD, which is then controlled continuously. Under pulsed blinding, the SPD is blinded and controlled for a period of time longer than the SPD's deadtime. The after-gate attack exploits controllability of gated SPDs in-between the gates, sending short bright pulses outside the gates. We think that testing for all three types of attacks can be automated. Here we demonstrate the automated testing for continuous blinding. We perform the pulsed blinding manually, to better understand the requirements for its automation. The after-gate attack is not applicable to a free-running SPD chosen for our experiment.

2.1 Automated testbench

Our testbench setup is shown in Fig. 1 [9]. It uses two lasers, a pulsed one and a cw one. Light from each of them passes through an isolator for stability reasons and then a programmable attenuator. Attenuated light from both lasers is then combined on a 90: 10 beamsplitter, whose outputs are connected to an optical power meter and the detector under test. A computer controls all the devices, runs a testing algorithm, and analyses the data.

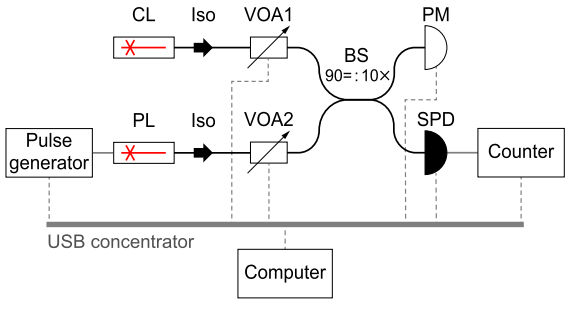

Figure 1 Setup for testing detector control by bright light. CL, continuous-wave laser (1552 nm, 40 mW, Thorlabs SFL1550P); PL, pulsed laser (1552 nm, Gooch & Housego AA1406); Iso, optical isolator; VOA, programmable variable optical attenuator (OZ Optics DA-100); BS, fiber beamsplitter; PM, optical power meter (Thorlabs PM400 with S155C head); SPD, single-photon detector under test. The pulse generator (Highland Technology P400) drives PL directly and can induce relaxation-limited short laser pulses. The counter (Stanford Research Systems SR620) typically accumulates clicks over 1 s for each data point

2.2 Software and methodology

Our software for the automated testbench is written in LabVIEW. It works in two stages. At the first stage, the testbench blinds the SPD by the cw laser. At the second stage, it attempts to control the SPD by the pulsed laser (this method is similar to earlier manual experiments [9]). The program then saves a printable PDF report, an example of which is given in Fig. 5, and all the raw data collected. For this particular report example, the entire test sequence took about 1.5 h.

During the first stage, the program uses CL to apply cw power at the SPD, while PL is turned off. The attenuation of VOA1 is scanned through its full 60 to 0 dB range by a user-settable step (1 dB in our case). The power is measured by the PM and varies from approximately 2.3×10^{-11} W (near the sensitivity limit of the PM) to 1.25×10^{-5} W. At each power level, the detector click rate (measured by the counter) and photocurrent monitor readout value (explained in the next subsection) are recorded. If the click rate drops to zero, the SPD is considered to be blinded.

If the blinding is recorded at one or more power levels, the program proceeds to the second stage. It steps VOA1 again from the maximum attenuation at which the blinding has been recorded through 0 dB. At each power level, it also applies short—240 ps full-width at half-magnitude (FWHM)—pulses from the PL at 10 kHz rate while scanning the attenuation of VOA2 from 60 to 0 dB by a user-settable step (1 dB in our case). The energy E of these control pulses is pre-calibrated and varies from 10^{-18} to 10^{-12} J. The detector click rate and monitor readout value are again recorded at each energy level. The program then analyses whether the detector clicks with above-zero probability in response to these control pulses (if it does, the SPD is declared controllable in the report) and if a change of E by 3 dB or less leads to the change of click probability from 0 to 100%. The latter is a sufficient condition for a perfect attack on Bennett-Brassard 1984 (BB84) QKD protocol [9]. The report also contains monitor readout plots under control, which may be analysed manually by the operator to see if the countermeasure is effective.

2.3 Detector under test

In this work, we investigate a free-running single-photon detector manufactured by QRate (serial number 3-054). This detector does not use gating, which makes it easy to use in a versatile educational kit [56] (whereas QRate's commercial QKD system employs a different detector model with sinusoidal gating that has improved performance [57]). Our free-running detector is based on an InGaAs/InP fiber-pigtailed APD (Wooriro WPACPG-MOSSNCNP serial number PA19H262-0052) thermoelectrically cooled to −35°C. The detector circuit uses passive quenching with enforced deadtime (Fig. 2) [57–59]. For Geiger-mode operation, the voltage across the APD should exceed its breakdown voltage by about 2 V. A high-voltage supply (HV; based on Maxim Integrated MAX1932) applies V_{bias} = +68.6 V at the cathode of the APD via a current mirror and bias resistor R1. When the detector is waiting ready for an avalanche, a stray capacitance between the APD cathode and the circuit ground is charged to the same voltage. This capacitance, on the order of 1 pF, is not shown in the circuit diagram but is essential for the detector operation. Once the avalanche begins, the capacitance supplies its current and discharges via the APD and low-impedance circuits connected at the APD anode. The voltage across the APD quickly drops; once it about equals the breakdown voltage, the current reduces to a value when the avalanche no longer self-sustains and the avalanche then stops [60]. Note that in this

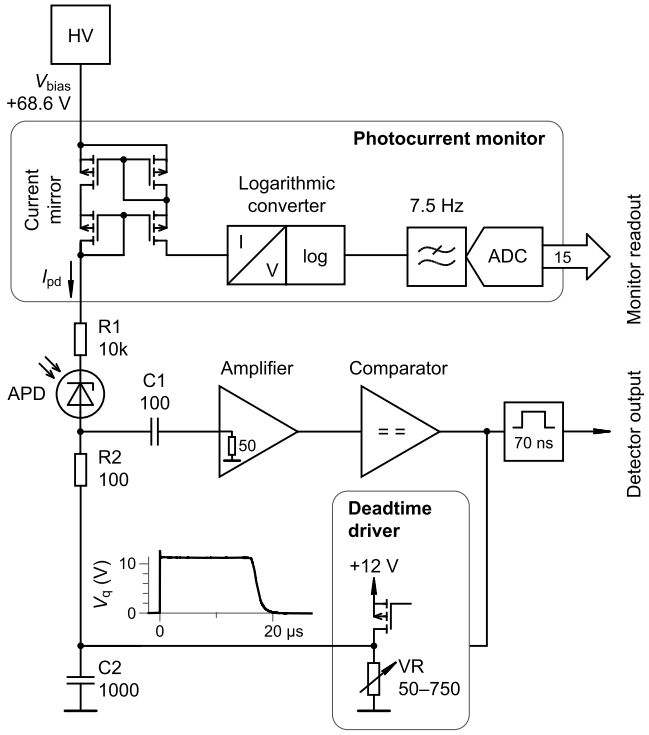

Figure 2 Functional schematic diagram of the single-photon detector. See text for details

passively-quenching circuit, the current supplied from HV via R1 is not sufficient to sustain the avalanche. It merely recharges the stray capacitance relatively slowly to V_{bias} after the avalanche quenches.

The onset of the avalanche current is sensed by an amplifier (Analog Devices HMC589AST89E) and high-speed comparator (Analog Devices ADCMP573), then expanded in duration to 70 ns with a single-shot generator, producing a logic signal at the output of the detector. To reduce afterpulsing, an enforced deadtime $\tau = 20\ \mu$s is applied by raising the voltage at the APD anode by 11.3 V. (This also ensures ending any occasional avalanche that does not cease via the passive quenching [60].) The deadtime driver removes this voltage at the end of the deadtime gradually via a variable resistor VR (implemented with a series of transistor switches), to avoid triggering additional avalanches [57, 59]. Once this process is complete, the stray capacitance charges via R1 to V_{bias} and the detector becomes ready for the next avalanche. The photon detection efficiency of our detector sample is 2.2% at 1550 nm and the dark count rate $D = 412$ Hz.

We remark that a commercial QKD system would use a gated detector with a higher photon detection efficiency, such as the sinusoidally-gated detector [57], which we have also tested on this testbench without a countermeasure [61]. However this free-running detector sample is sufficient for the initial development of our testbench and the countermeasure. Its low photon detection efficiency does not affect the test methodology. In the future, the countermeasure will also be implemented and tested in QRate's sinusoidally-gated detector.

To provide the countermeasure against detector blinding [9], the present free-running SPD employs a photocurrent monitor circuit (Fig. 2) [37, 41]. The current I_{pd} flowing

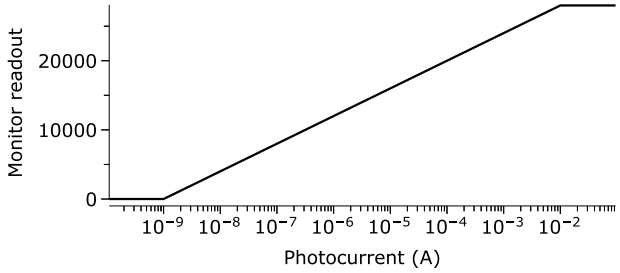

Figure 3 Monitor readout M as a function of a constant APD current I_{pd}. The curve has been theoretically calculated based on the data sheets of the integrated circuits. Note that with V_{bias} = +68.6 V and the resistor values used in this particular detector sample, I_{pd} cannot exceed about 6.8 mA

into R1 is copied by a current mirror, processed by a logarithmic converter (Analog Devices AD8305) configured with a reduced bandwidth, and digitised by an analog-to-digital converter (ADC; Microchip MCP3425) with 15-bit resolution. The analog-to-digital converter outputs a readout value

$$M = 4000 \log_{10} \frac{I_{\mathrm{pd}}}{1\ \mathrm{nA}}, \tag{1}$$

which obeys this equation in a wide range of constant-current values from 10^{-9} to 10^{-2} A (Fig. 3). The logarithmic signal passes a low-pass frequency filter with 7.5 Hz rolloff (intrinsic to the ADCs of delta-sigma type) and is digitised at 15 Hz rate. The latest readout value is made available via a universal serial bus (USB) interface to a computer running either a detector monitoring software supplied by the manufacturer or user-written program like our testbench automation. The latter records a single sampled value whenever it takes a data point for the automatically generated report. These single values of M fluctuate significantly when the detector is producing random counts (the fluctuation is up to ± 700 for dark counts, less at higher count rates). To reduce these random fluctuations, in manual measurements of pulsed blinding we have done additional averaging for each data point, sampling 30 readout values spread evenly over 14.5 s and calculating their mean.

Note that the 7.5 Hz low-pass filter in the monitor circuit is placed *after* the logarithmic converter, which itself outputs a signal with a much higher bandwidth of the order of 1 kHz. When I_{pd} varies in time at a frequency faster than 7.5 Hz but below 1 kHz, this circuit does not average the photocurrent but rather averages its logarithm. The readout M then *underestimates* the mean value of I_{pd}. This helps Eve in defeating this countermeasure, as we show below.

3 Experimental results

3.1 Countermeasure calibration

Before the monitor readout can be used to detect attacks, we need to estimate the values of M observed during normal operation of the QKD system. We simulate the single-photon regime by illuminating the SPD with laser pulses attenuated to 0.8 photon/pulse at a varying rate. The results are shown in Fig. 4. We assume a typical detector click rate

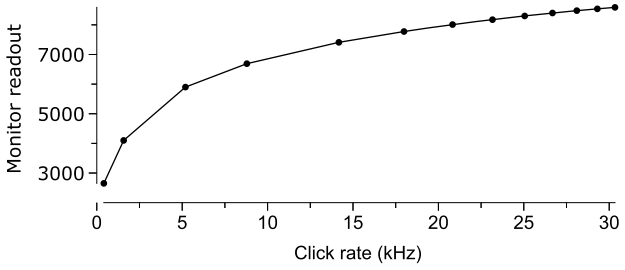

Figure 4 Monitor readout M under detector illumination by 0.8-photon pulses. The leftmost point is for unilluminated detector with dark counts only, while the other points are under illumination at a laser pulse rate from 10 kHz to 1 MHz. An estimated "normal" detector click rate of 20 kHz or less results in single sampled values of $M < 8100$

in a modern QKD system is 20 kHz[1]. The monitor readout is then expected not to exceed 8100 for single sampled values (or 7900 averaged). Any value larger than that indicates an attack.

3.2 Continuous-wave blinding of detector

The continuous blinding attack is executed automatically by the testbench. The report (Fig. 5) includes the following plots: the dependence of the count rate and monitor readout value M on cw laser power (with PL off), the probability that the blinded SPD produces a click versus control pulse's energy E (each curve is at a different cw blinding power), the maximum pulse energy that never produces a click E_{never} and minimum energy that always produces a click E_{always} [9] versus the cw blinding power, and the dependence of M on the SPD click rate when it is being blinded and controlled (each curve is at a different cw blinding power). The software automatically analyses the click rates and makes conclusions that this detector is blindable and the click probability under control is non-zero. Note that the cw power and E are both scanned an order of magnitude or more above and below the values where the full control is observed.

We can manually analyse the monitor readout plots and see that the countermeasure catches this attack. When the SPD is blinded at 2.96 nW, the countermeasure registers $M \approx 14{,}800$, and under total control at 75.6 nW $M \approx 15{,}000$. This significantly violates the safety condition $M \leq 8100$ calibrated in Sect. 3.1.

For completeness, we also need to check M when the blinded SPD is controlled by the PL. The monitor readout drops perceptibly when the detector begins to click, while being otherwise independent on E (see Fig. 6 and the last plot in Fig. 5). The drop of M can be explained by the forced reduction of voltage across the APD during the deadtime, which decreases the APD's internal gain and thus its mean photocurrent in response to cw illumination. However, even the reduced $M \geq 13{,}405$ remains well above the "normal" monitor readout of 8100. This keeps the countermeasure effective against the continuous blinding.

We observe no temporary or permanent deterioration of the SPD during our tests. This type of APD-based detector is known to withstand 10 mW cw optical illumination without damage [10], which is higher than the optical power in our testbench.

[1] In QRate QKD proof-of-principle experiments [62], the system runs at 10 MHz source pulse rate over two lines: a 50 km fiber spool and 30 km urban line. Scaling the click rates reported to a future 1 GHz source rate, we expect a single detector count rate of 50 and 13 kHz. We thus assume 20 kHz to be a typical click rate.

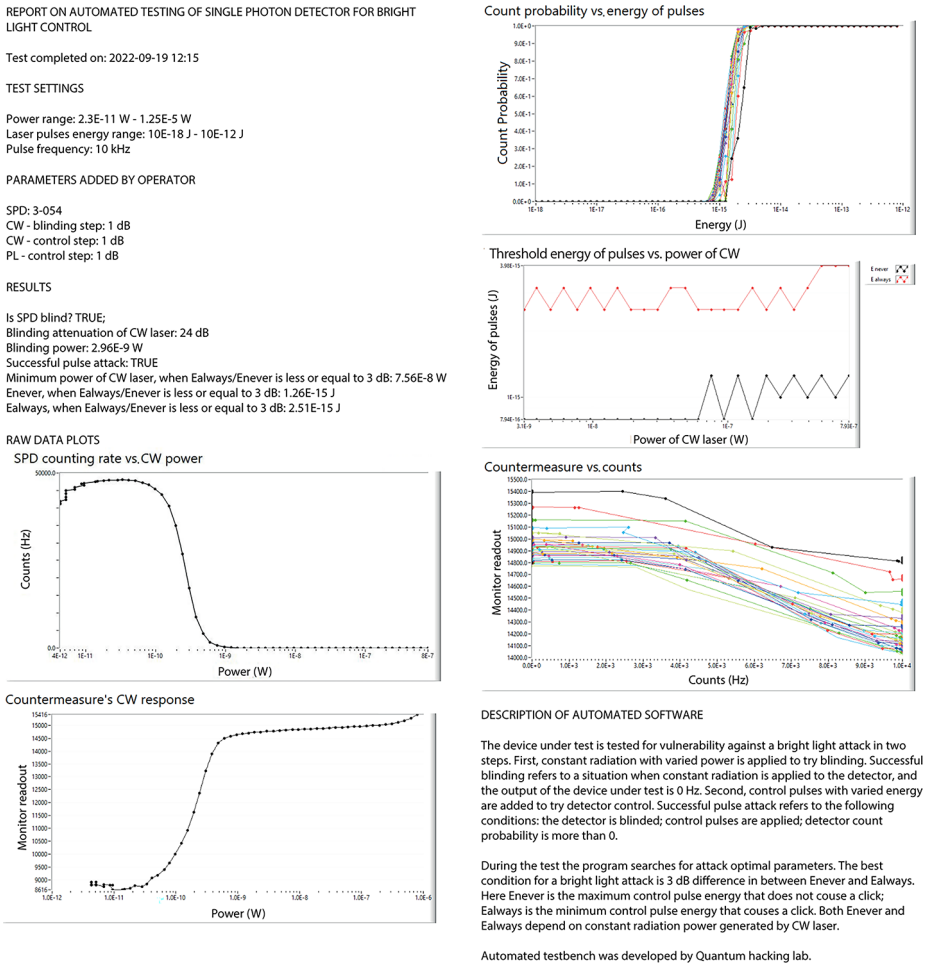

Figure 5 Report generated by software after automated test of cw blinding and control. The multiple curves in the third and fifth plots are taken at cw power values shown in the fourth plot. See text for details

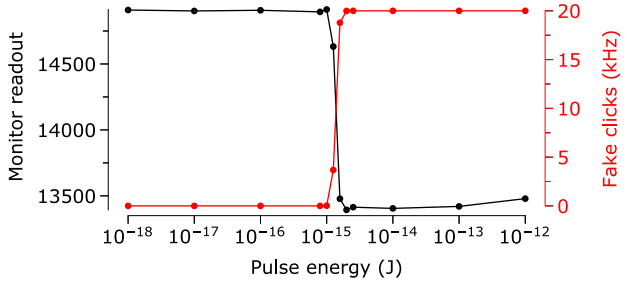

Figure 6 Monitor readout and click rate versus trigger pulse energy E, at 31.5 nW cw blinding power and 20 kHz trigger pulse rate. The countermeasure is not affected by E until it causes a click

3.3 Pulsed blinding of detector

The photocurrent-measuring countermeasure may be ineffective against the pulsed blinding attack [41, 54]. In this attack, the SPD is blinded temporarily and controlled while blinded. It works in the normal photon counting mode between the blinding pulses. We

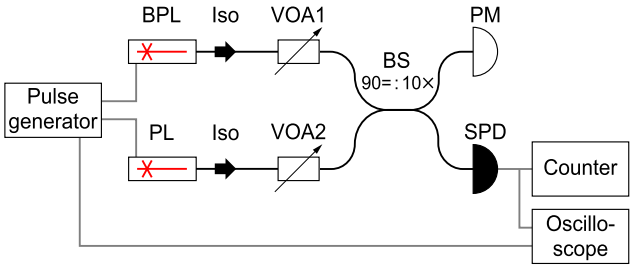

Figure 7 A modified setup for testing detector control under pulsed blinding. BPL, blinding pulse laser [1552 nm, 40 mW, Allwave Lasers SWLD-1550-100-PM(DBF)]. The pulse generator drives BPL directly and induces long laser pulses. The pulsed laser (PL) is also driven by the pulse generator directly and emits a relaxation-limited short laser pulse with 240 ps FWHM delayed in respect to the start of the BPL pulse. The counter typically accumulates clicks over 1 s for each data point

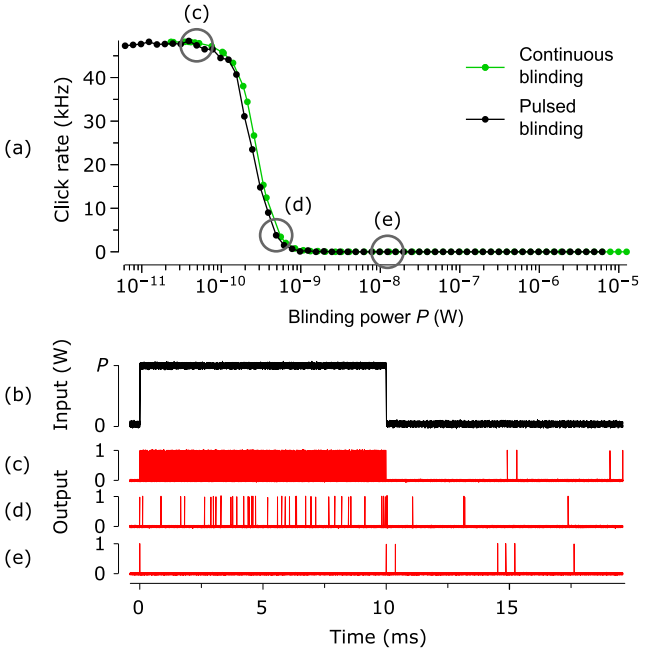

Figure 8 Pulsed blinding of the SPD. (**a**) Click rate under blinding. For pulsed blinding, only the click rate within the blinding pulse is measured; outside the blinding pulse, the detector works in the normal photon counting mode with dark counts. Click rate under cw blinding is shown for comparison. (**b**) The blinding pulse of peak power *P*. (**c**) Oscillogram of detector output (70 ns long logic pulses of 1.5 V amplitude) when $P = 49$ pW is insufficient for blinding and causes a saturated click rate within the pulse. (**d**) At a higher $P = 490$ pW, the click rate within the pulse drops significantly. (**e**) Detector is blinded within the pulse of $P = 12.3$ nW

modify our experimental setup slightly (Fig. 7) and use it to manually test the SPD. We first apply 10 ms long blinding pulse of peak power P (measured by PM) at 20 Hz repetition rate and observe the detector clicks. The results are shown in Fig. 8. We use an oscilloscope (3.5 GHz bandwidth; LeCroy 735Zi) to select the clicks that occur during the blinding pulse and measure their rate. The complete blinding within the pulse occurs at $P = 2.46$ nW, which is almost the same power as in the cw blinding (2.96 nW).

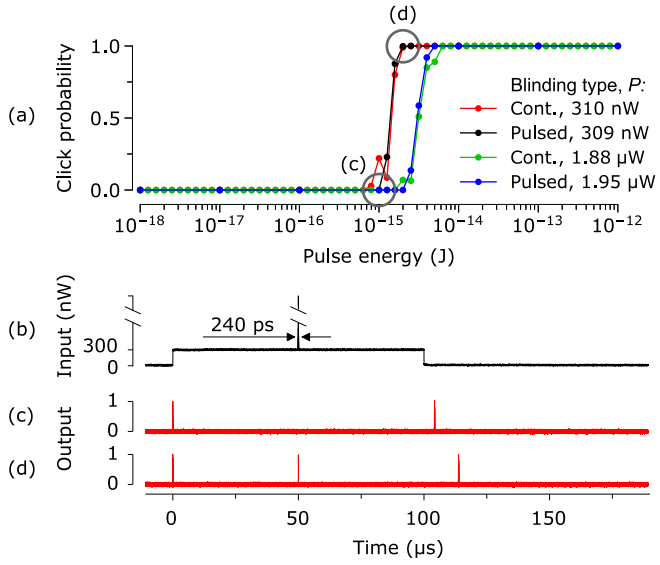

Figure 9 Control of SPD within 100 μs long blinding pulses applied at 20 Hz rate. (**a**) Click probability versus trigger pulse energy E. Click probabilities under cw blinding are shown for comparison. The probabilities are measured over 10^3 pulses. (**b**) The blinding and trigger pulses, the latter being of 240 ps FWHM and applied 50 μs after the start of the blinding pulse. (**c**) Detector output at $P = 309$ nW and $E = 10^{-15}$ J. This trigger pulse never causes a click. (**d**) The trigger pulse energy is increased by 3 dB to 2×10^{-15} J. It always causes a click

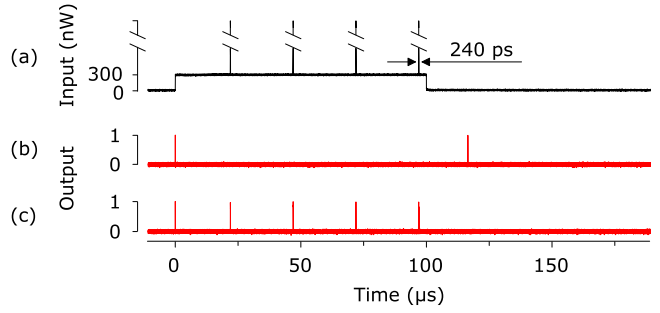

Figure 10 Detector control by multiple trigger pulses applied during one blinding pulse. (**a**) Blinding and trigger pulses, the latter being of 240 ps FWHM. The blinding pulse has $P = 309$ nW and is applied at 20 Hz rate. (**b**) Detector output at $E = 10^{-15}$ J. These trigger pulses never cause clicks. (**c**) The trigger pulses' energy is increased by 3 dB to 2×10^{-15} J. Each of them always causes a click

The detector behaviour within the pulse closely resembles that under the cw blinding. If an additional short trigger pulse of energy E is applied during the blinding pulse, it responds with a click (Fig. 9). Less than 3 dB change of E is required to transition between 0 and 100% click probability. Note that the detector always clicks at the start of the blinding pulse, and often also at the end of it. Because of these additional uncontrollable clicks, it is beneficial for Eve to apply multiple trigger pulses during the blinding pulse. In Fig. 10, control by four trigger pulses is illustrated. The response to the trigger pulse does not depend on its timing within the blinding pulse; the click probability changes less than 2% throughout. We have also verified that up to 199 trigger pulses spaced 20 μs apart (i.e., the exact length of the detector deadtime τ) applied during a longer 4 ms blinding pulse

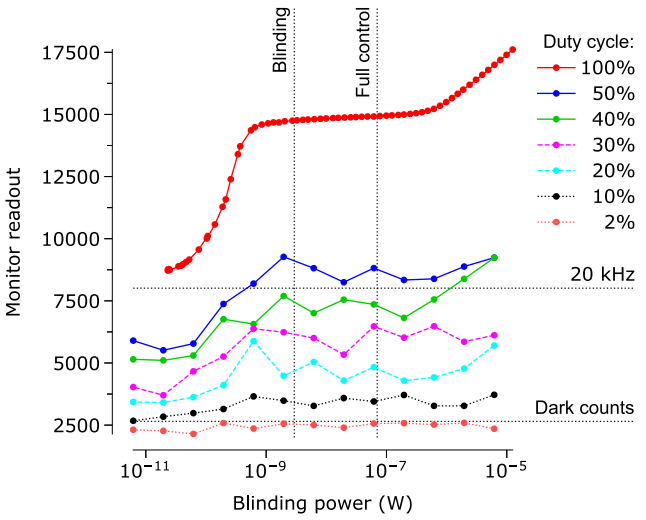

Figure 11 Monitor readout under pulsed blinding at different duty cycle values of the blinding illumination, at 20 Hz repetition rate of the blinding pulses. 100% is cw blinding, 50% is 25 ms long blinding pulse, 40% is 20 ms long blinding pulse, etc. Dotted horizontal lines indicate an expected countermeasure readout in the normal photon counting mode at 20 kHz click rate and at the dark count rate (as calibrated in Sect. 3.1). Dotted vertical lines indicate the minimum blinding power and minimum cw power at which $E_{\text{always}}/E_{\text{never}} \leq 2$

work as well. The monitor readout decreases slightly as the number of triggered clicks increases, similarly to the effect observed in Fig. 6.

Finally, we check how the countermeasure responds to the pulsed blinding of different duty cycle values. We vary the blinding pulse width while keeping its repetition rate constant at 20 Hz, see Fig. 11. At this repetition rate, the blinding pulse can be as long as 20 ms without causing an abnormally high monitor readout of more than 7900. The low monitor readout under the pulsed blinding is partially explained by the unwisely constructed sequence of first taking the logarithm then averaging at the low-pass frequency filter in the photocurrent monitor circuit (Sect. 2.3). This implementation of the countermeasure is thus unable to detect the pulsed blinding of up to 40% duty cycle, which leaves Eve ample room for attack.

3.4 Intercept-resend attack model

The experimental results on pulsed blinding show that Eve has a significant degree of control over the SPD, while not being revealed by the countermeasure. This SPD behaviour violates the assumptions on a measurement apparatus made in most security proofs for QKD, in particular the independence of detection probability on Bob's basis choice [14], rendering these proofs inapplicable. We thus clearly cannot guarantee the security of QKD that employs such SPDs, regardless of whether we know how to construct Eve's attack in detail or not.

Nevertheless, here we attempt to model such attack. We assume the QKD system runs the BB84 protocol [1] with an active basis choice and two detectors at Bob. Eve intercepts Alice's output at the beginning of the lossy quantum channel using a receiver with a high detection efficiency and very low error rate. She then resends blinding pulses and faked states to Bob according to her measurement results [9]. We also assume Bob's dark counts

are the only source of errors in the system without Eve. Let us approximately estimate Alice's and Bob's quantum bit error rate (QBER) under attack and the rate at which Eve can trigger clicks at Bob.

We consider one pulsed blinding period of length T, consisting of CT blinding time and $(1-C)T$ idle time, where $C \in (0,1)$ is the duty cycle. The blinding pulse causes a simultaneous click in both Bob's detectors at its start and, possibly, a click at its end. We assume these, on average, record in the raw key as two clicks, of which one is erroneous. During the blinding time Eve can induce $\gtrsim CT/2\tau$ controlled clicks at Bob (the exact number depends on Eve's detection rate and whether she can send faked states during Bob's deadtime). Bob also registers about $2(1-C)TD$ dark counts during the idle time. Bob's click rate

$$R_{\mathrm{B}} \approx \frac{2 + CT/2\tau + 2(1-C)TD}{T} \tag{2}$$

and

$$\mathrm{QBER} \approx \frac{1 + (1-C)TD}{TR_{\mathrm{B}}}. \tag{3}$$

We stress that the above calculation is approximate and ignores lesser effects like double clicks at Bob, rate reduction owing to his detector deadtime, sources of errors other than Bob's dark counts, etc.

Taking the experimental parameters from this paper (T = 50 ms, C = 0.4, etc.), we get $R_{\mathrm{B}} \approx 10.5$ kHz and QBER $\approx 2.5\%$. These are reasonable parameters for a QKD system and it should generate a key. Since Eve is performing the intercept-resend attack, the generation of secret key under this attack is in fact impossible [63].

The main limitation of this attack is its ability to replicate R_{B} expected by the legitimate users. Its value depends on the system implementation and the line loss, and for many practical settings is less than 10.5 kHz [62]. Also a faster QKD system would use detectors with shorter τ, which helps Eve obtain a higher R_{B} [Eq. (2)]. If Bob's click rate under the attack is still insufficient to replicate the system performance expected by Alice and Bob, Eve can choose to bypass a fraction of Alice's photons into the quantum channel to Bob during the idle time. This strategy would not be an optimal attack. We would then need to consider Eve's key information versus the amount of privacy amplification Alice and Bob apply. We speculate that either attack strategy should also work with the decoy-state protocol [28], as this attack does not drastically affect the yield of different photon-number states.

4 Discussion and conclusion

Our testbench has tested the free-running SPD for cw blinding and control and made conclusions about it fully automatically, essentially replicating the well-known manual testing method [9, 41]. A manual analysis of collected data shows that the countermeasure reveals the cw blinding reliably. We then manually demonstrate pulsed blinding and control of this SPD. The countermeasure fails to reveal the pulsed blinding of up to 40% duty cycle, allowing Eve to control the detector during the blinding pulses. Our modeling shows that the intercept-resend attack on QKD should then still be possible.

To build the testbench good enough for certification purposes, its automatic operation should be extended to pulsed blinding regimes. The testbench should also automatically analyse the countermeasure output under both cw and pulsed blinding, and make a pass/fail conclusion whether the countermeasure reveals all the attacks. Such extension of the testing algorithm is a topic for future work. In order to develop it, we need to have the SPD with a properly implemented countermeasure that reveals the pulsed attacks.

The existing countermeasure implementation fails to reveal the pulsed blinding primarily because of very low ADC sampling rate of the photocurrent (15 Hz). Our results suggest that increasing the bandwidth and processing the monitor signal for peak detection would be a step in the right direction. Direct measurements of the signal at the output of the logarithmic converter with an oscilloscope suggest that an ADC with $\sim$1 MHz sampling rate or an analog comparator (i.e., a voltage threshold detector) would be sufficient to reveal the pulsed blinding. The necessary hardware can easily be added to the next version of QRate's free-running SPD. Implementing and testing this improved countermeasure, as well as adopting it for the sinusoidally-gated detector, will be our next study. Testing superconducting-nanowire single-photon detectors with a built-in countermeasure [55] is also a promising application.

The quantum key distribution protocol needs to be amended to take input from the countermeasure. One obviously secure method is to discard the entire accumulated raw key and start a new QKD session whenever an abnormally high monitor readout value occurs. A less wasteful approach might be to discard potentially compromised raw key data in a limited time range that surrounds the abnormally high monitor readout, while continuing the current QKD session. We remark that the countermeasure might occasionally be triggered by benign transient events like electromagnetic interference, computer glitch, or optical line maintenance [22]. If the problem persists over multiple key distillation sessions, it might be a good idea to alert the human operator of the system of this abnormality, which may be caused by a technical malfunction or the actual attempt of attack. We finally remark that our testbench does not test for effects that may appear at higher optical power, such as thermal blinding [10] and laser damage of APD [18]. While the thermal blinding can be tested in this setup, the laser damage requires significant modifications of the testbench [27, 36].

Acknowledgements
We thank Hao Qin for discussions and for providing motivation for this study.

Funding
This work was funded by the Ministry of Science and Education of Russia (program NTI center for quantum communications and grant 075-11-2021-078), Russian Science Foundation (grant 21-42-00040), 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 Autónomas" in Quantum Communication, the National Natural Science Foundation of China (grants 61901483 and 62061136011), the National Key Research and Development Program of China (grant 2019QY0702), and the Research Fund Program of State Key Laboratory of High Performance Computing (grant 202001-02).

Abbreviations
QKD, quantum key distribution; SPD, single-photon detector; MDI QKD, measurement-device-independent QKD; CL, continuous-wave laser; PL, pulsed laser; Iso, optical isolator; VOA, programmable variable optical attenuator; BS, fiber beamsplitter; PM, optical power meter.

Availability of data and materials
Partial data generated or analysed during this study are included in this published article. Any datasets used and/or analysed during the current study that have not been included in this published article are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors have approved the publication. The research in this work did not involve any human, animal or other participants.

Competing interests
The authors declare no competing interests.

Author contributions
P.A. and K.Z. designed and programmed the automated testbench, performed the experiments, and analysed the data. A.H. and V.M. analysed the data. V.Z. and A.L. developed the detector under test and the countermeasure. P.A., K.Z., and V.M. wrote the paper with input from all authors. V.M. supervised the detector testing project.

Author details
[1]Russian Quantum Center, Skolkovo, Moscow 121205, Russia. [2]Vigo Quantum Communication Center, University of Vigo, Vigo E-36310, Spain. [3]atlanTTic Research Center, University of Vigo, Vigo E-36310, Spain. [4]NTI Center for Quantum Communications, National University of Science and Technology MISiS, Moscow 119049, Russia. [5]HSE University, Moscow 101000, Russia. [6]Institute for Quantum Information & State Key Laboratory of High Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, People's Republic of China.

Received: 10 February 2023 Accepted: 4 June 2023 Published online: 26 June 2023

References

1. Bennett CH, Brassard G. Quantum cryptography: public key distribution and coin tossing. In: Proc. international conference on computers, systems, and signal processing. New York: IEEE Press; 1984. p. 175–9.
2. Brassard G, Lütkenhaus N, Mor T, Sanders BC. Limitations on practical quantum cryptography. Phys Rev Lett. 2000;85:1330–3.
3. Makarov V, Anisimov A, Skaar J. Effects of detector efficiency mismatch on security of quantum cryptosystems. Phys Rev A. 2006;74:022313. Erratum ibid. 2008;78:019905.
4. Gisin N, Fasel S, Kraus B, Zbinden H, Ribordy G. Trojan-horse attacks on quantum-key-distribution systems. Phys Rev A. 2006;73:022320.
5. Qi B, Fung C-HF, Lo H-K, Ma X. Time-shift attack in practical quantum cryptosystems. Quantum Inf Comput. 2007;7:73–82.
6. Lamas-Linares A, Kurtsiefer C. Breaking a quantum key distribution system through a timing side channel. Opt Express. 2007;15:9388–93.
7. Makarov V, Skaar J. Faked states attack using detector efficiency mismatch on SARG04, phase-time, DPSK, and Ekert protocols. Quantum Inf Comput. 2008;8:622–35.
8. Zhao Y, Fung C-HF, Qi B, Chen C, Lo H-K. Quantum hacking: experimental demonstration of time-shift attack against practical quantum-key-distribution systems. Phys Rev A. 2008;78:042333.
9. Lydersen L, Wiechers C, Wittmann C, Elser D, Skaar J, Makarov V. Hacking commercial quantum cryptography systems by tailored bright illumination. Nat Photonics. 2010;4:686–9.
10. Lydersen L, Wiechers C, Wittmann C, Elser D, Skaar J, Makarov V. Thermal blinding of gated detectors in quantum cryptography. Opt Express. 2010;18:27938–54.
11. Li H-W, Wang S, Huang J-Z, Chen W, Yin Z-Q, Li F-Y, Zhou Z, Liu D, Zhang Y, Guo G-C, Bao W-S, Han Z-F. Attacking a practical quantum-key-distribution system with wavelength-dependent beam-splitter and multiwavelength sources. Phys Rev A. 2011;84:062308.
12. Wiechers C, Lydersen L, Wittmann C, Elser D, Skaar J, Marquardt C, Makarov V, Leuchs G. After-gate attack on a quantum cryptosystem. New J Phys. 2011;13:013043.
13. Lydersen L, Akhlaghi MK, Majedi AH, Skaar J, Makarov V. Controlling a superconducting nanowire single-photon detector using tailored bright illumination. New J Phys. 2011;13:113042.
14. Lydersen L, Jain N, Wittmann C, Marøy Ø, Skaar J, Marquardt C, Makarov V, Leuchs G. Superlinear threshold detectors in quantum cryptography. Phys Rev A. 2011;84:032320.
15. Gerhardt I, Liu Q, Lamas-Linares A, Skaar J, Kurtsiefer C, Makarov V. Full-field implementation of a perfect eavesdropper on a quantum cryptography system. Nat Commun. 2011;2:349.
16. Sun S-H, Jiang M-S, Liang L-M. Passive Faraday-mirror attack in a practical two-way quantum-key-distribution system. Phys Rev A. 2011;83:062331.
17. Jain N, Wittmann C, Lydersen L, Wiechers C, Elser D, Marquardt C, Makarov V, Leuchs G. Device calibration impacts security of quantum key distribution. Phys Rev Lett. 2011;107:110501.
18. Bugge AN, Sauge S, Mardhiyah A, Ghazali M, Skaar J, Lydersen L, Makarov V. Laser damage helps the eavesdropper in quantum cryptography. Phys Rev Lett. 2014;112:070503.
19. Huang J-Z, Weedbrook C, Yin Z-Q, Wang S, Li H-W, Chen W, Guo G-C, Han Z-F. Quantum hacking of a continuous-variable quantum-key-distribution system using a wavelength attack. Phys Rev A. 2013;87:062329.
20. Sun S-H, Xu F, Jiang M-S, Ma X-C, Lo H-K, Liang L-M. Effect of source tampering in the security of quantum cryptography. Phys Rev A. 2015;92:022304.
21. Sajeed S, Chaiwongkhot P, Bourgoin J-P, Jennewein T, Lütkenhaus N, Makarov V. Security loophole in free-space quantum key distribution due to spatial-mode detector-efficiency mismatch. Phys Rev A. 2015;91:062301.

22. Huang A, Sajeed S, Chaiwongkhot P, Soucarros M, Legré M, Makarov V. Testing random-detector-efficiency countermeasure in a commercial system reveals a breakable unrealistic assumption. IEEE J Quantum Electron. 2016;52:8000211.
23. Sajeed S, Huang A, Sun S, Xu F, Makarov V, Curty M. Insecurity of detector-device-independent quantum key distribution. Phys Rev Lett. 2016;117:250505.
24. Makarov V, Bourgoin J-P, Chaiwongkhot P, Gagné M, Jennewein T, Kaiser S, Kashyap R, Legré M, Minshull C, Sajeed S. Creation of backdoors in quantum communications via laser damage. Phys Rev A. 2016;94:030302.
25. Huang A, Sun SH, Liu Z, Makarov V. Quantum key distribution with distinguishable decoy states. Phys Rev A. 2018;98:012330.
26. Zheng Y, Huang P, Huang A, Peng J, Zeng G. Practical security of continuous-variable quantum key distribution with reduced optical attenuation. Phys Rev A. 2019;100:012313.
27. Huang A, Li R, Egorov V, Tchouragoulov S, Kumar K, Makarov V. Laser-damage attack against optical attenuators in quantum key distribution. Phys Rev Appl. 2020;13:034017.
28. Lo H-K, Ma X, Chen K. Decoy state quantum key distribution. Phys Rev Lett. 2005;94:230504.
29. Lucamarini M, Choi I, Ward MB, Dynes JF, Yuan ZL, Shields AJ. Practical security bounds against the Trojan-horse attack in quantum key distribution. Phys Rev X. 2015;5:031030.
30. Tamaki K, Curty M, Lucamarini M. Decoy-state quantum key distribution with a leaky source. New J Phys. 2016;18:065008.
31. Wang W, Tamaki K, Curty M. Finite-key security analysis for quantum key distribution with leaky sources. New J Phys. 2018;20:083027.
32. Fung C-HF, Tamaki K, Qi B, Lo H-K, Ma X. Security proof of quantum key distribution with detection efficiency mismatch. Quantum Inf Comput. 2009;9:131–65.
33. Weier H, Krauss H, Rau M, Fürst M, Nauerth S, Weinfurter H. Quantum eavesdropping without interception: an attack exploiting the dead time of single-photon detectors. New J Phys. 2011;13:073024.
34. Qian Y-J, He D-Y, Wang S, Chen W, Yin Z-Q, Guo G-C, Han Z-F. Hacking the quantum key distribution system by exploiting the avalanche-transition region of single-photon detectors. Phys Rev Appl. 2018;10:064062.
35. Makarov V. Controlling passively quenched single photon detectors by bright light. New J Phys. 2009;11:065003.
36. Ponosova A, Ruzhitskaya D, Chaiwongkhot P, Egorov V, Makarov V, Huang A. Protecting fiber-optic quantum key distribution sources against light-injection attacks. PRX Quantum. 2022;3:040307.
37. Yuan ZL, Dynes JF, Shields AJ. Resilience of gated avalanche photodiodes against bright illumination attacks in quantum cryptography. Appl Phys Lett. 2011;98:231104.
38. Lo H-K, Curty M, Qi B. Measurement-device-independent quantum key distribution. Phys Rev Lett. 2012;108:130503.
39. da Silva TF, Xavier GB, Temporao GP, von der Weid JP. Real-time monitoring of single-photon detectors against eavesdropping in quantum key distribution systems. Opt Express. 2012;20:18911–24.
40. Lim CCW, Walenta N, Legré M, Gisin N, Zbinden H. Random variation of detector efficiency: a countermeasure against detector blinding attacks for quantum key distribution. IEEE J Sel Top Quantum Electron. 2015;21:6601305.
41. Gras G, Sultana N, Huang A, Jennewein T, Bussières F, Makarov V, Zbinden H. Optical control of single-photon negative-feedback avalanche diode detector. J Appl Phys. 2020;127:094502.
42. Koehler-Sidki A, Dynes JF, Lucamarini M, Roberts GL, Sharpe AW, Yuan ZL, Shields AJ. Best-practice criteria for practical security of self-differencing avalanche photodiode detectors in quantum key distribution. Phys Rev Appl. 2018;9:044027.
43. Qian Y-J, Li H-W, He D-Y, Yin Z-Q, Zhang C-M, Chen W, Wang S, Han Z-F. Countermeasure against probabilistic blinding attack in practical quantum key distribution systems. Chin Phys B. 2015;24:090305.
44. Koehler-Sidki A, Dynes JF, Martinez A, Lucamarini M, Roberts GL, Sharpe AW, Yuan ZL, Shields AJ. Intrinsic mitigation of the after-gate attack in quantum key distribution through fast-gated delayed detection. Phys Rev Appl. 2019;12:024050.
45. Alhussein M, Inoue K, Honjo T. Monitoring coincident clicks in differential-quadrature-phase shift QKD to reveal detector blinding and control attacks. Jpn J Appl Phys. 2019;58:012006.
46. Marøy Ø, Makarov V, Skaar J. Secure detection in quantum key distribution by real-time calibration of receiver. Quantum Sci Technol. 2017;2:044013.
47. Lee MS, Park BK, Woo MK, Park CH, Kim Y-S, Han S-W, Moon S. Countermeasure against blinding attacks on low-noise detectors with a background-noise-cancellation scheme. Phys Rev A. 2016;94:062321.
48. Qian Y-J, He D-Y, Wang S, Chen W, Yin Z-Q, Guo G-C, Hans Z-F. Robust countermeasure against detector control attack in a practical quantum key distribution system. Optica. 2019;6:1178–84.
49. Zhang G, Primaatmaja IW, Haw JY, Gong X, Wang C, Lim CCW. Securing practical quantum communication systems with optical power limiters. PRX Quantum. 2021;2:030304.
50. Parra J, Navarro-Arenas J, Menghini M, Recaman M, Pierre-Locquet J, Sanchis P. Low-threshold power and tunable integrated optical limiter based on an ultracompact VO2/Si waveguide. APL Photon. 2021;6:121301.
51. Chen Y-A, Zhang Q, Chen T-Y, Cai W-Q, Liao S-K, Zhang J, Chen K, Yin J, Ren J-G, Chen Z, Han S-L, Yu Q, Liang K, Zhou F, Yuan X, Zhao M-S, Wang T-Y, Jiang X, Zhang L, Liu W-Y, Li Y, Shen Q, Cao Y, Lu C-Y, Shu R, Wang J-Y, Li L, Liu N-L, Xu F, Wang X-B, Peng C-Z, Pan J-W. An integrated space-to-ground quantum communication network over 4,600 kilometres. Nature. 2021;589:214–9.
52. Länger T, Lenhart G. Standardization of quantum key distribution and the ETSI standardization initiative ISG-QKD. New J Phys. 2009;11:055051.
53. ISO/IEC DIS 23837-2(en): information technology security techniques — security requirements, test and evaluation methods for quantum key distribution — part 2: evaluation and testing methods, https://www.iso.org/obp/ui/#iso:std:iso-iec:23837:-2:dis:ed-1:v1:en, visited 3 Dec 2022.
54. Gao B, Wu Z, Shi W, Liu Y, Wang D, Yu C, Huang A, Wu J. Ability of strong-pulse illumination to hack self-differencing avalanche photodiode detectors in a high-speed quantum-key-distribution system. Phys Rev A. 2022;106:033713.
55. Tanner MG, Makarov V, Hadfield RH. Optimised quantum hacking of superconducting nanowire single-photon detectors. Opt Express. 2014;22:6734–48.

56. Rodimin VE, Kiktenko EO, Usova VV, Ponomarev MY, Kazieva TV, Miller AV, Sokolov AS, Kanapin AA, Losev AV, Trushechkin AS, Anufriev MN, Pozhar NO, Kurochkin VL, Kurochkin YV, Fedorov AK. Modular quantum key distribution setup for research and development applications. J Russ Laser Res. 2019;40:221.
57. Losev A, Zavodilenko V, Koziy A, Kurochkin Y, Gorbatsevich A. Dependence of functional parameters of sine-gated InGaAs/InP single-photon avalanche diodes on the gating parameters. IEEE Photonics J. 2022;14:6817109.
58. Stipčević M, Christensen BG, Kwiat PG, Gauthier DJ. Advanced active quenching circuit for ultra-fast quantum cryptography. Opt Express. 2017;25:21861–76.
59. Losev AV, Zavodilenko VV, Koziy AA, Filyaev AA, Khomyakova KI, Kurochkin YV, Gorbatsevich AA. Dead time duration and active reset influence on the afterpulse probability of InGaAs/InP single-photon avalanche diodes. IEEE J Quantum Electron. 2022;58:4500111.
60. Kim Y-S, Jeong Y-C, Sauge S, Makarov V. Ultra-low noise single-photon detector based on Si avalanche photodiode. Rev Sci Instrum. 2011;82:093110.
61. Makarov V, Abrikosov A, Chaiwongkhot P, Fedorov A, Huang A, Kiktenko E, Petrov M, Ponosova A, Ruzhitskaya D, Sajeed S, Tayduganov A, Trefilov D, Zaitsev K. Preparing a commercial quantum key distribution system for certification against implementation loopholes. 2023. unpublished.
62. Duplinskiy A, Ustimchik V, Kanapin A, Kurochkin V, Kurochkin Y. Low loss QKD optical scheme for fast polarization encoding. Opt Express. 2017;25:28886–97.
63. Curty M, Lewenstein M, Lütkenhaus N. Entanglement as a precondition for secure quantum key distribution. Phys Rev Lett. 2004;92:217903.

Publisher's Note