Countermeasure Against Quantum Hacking Using Detection Statistics

Gaëtan Gras,[1,2,*] Davide Rusca,[2] Hugo Zbinden,[2] and Félix Bussières[1,2]

[1] *ID Quantique SA, CH-1227 Carouge, Switzerland*

[2] *Group of Applied Physics, University of Geneva, CH-1211 Geneva, Switzerland*

(Received 19 October 2020; accepted 17 February 2021; published 17 March 2021)

Detector blinding attacks have been proposed in the last few years, and they could potentially threaten the security of quantum key distribution systems. Even though such attacks are technically challenging to implement, it is important to consider countermeasures to avoid information leakage. In this paper, we present a countermeasure against these kinds of attacks based on the use of multipixel detectors. We show that with this method, we are able to estimate an upper bound on the information an eavesdropper could have on the key exchanged. Finally, we test a multipixel detector based on superconducting nanowire single-photon detectors to show it can fulfill all the requirements for our countermeasure to be effective.

DOI: 10.1103/PhysRevApplied.15.034052

I. INTRODUCTION

Since its first proposal by Bennett and Brassard in 1984 [1], quantum key distribution (QKD) has attracted a lot of interest for securing communications. Indeed, with QKD, two distant parties, Alice and Bob, can securely exchange a key to encrypt their communications. QKD does not require making assumptions on the computational power of the eavesdropper Eve, making this technology theoretically secure. However, imperfections of physical systems can potentially be exploited by Eve to break the security and obtain some information on the key without being noticed. Several attacks have already been proposed, such as the photon-number splitting attack [2], detector efficiency mismatch attack [3], and Trojan horse attack [4–6], as well as potential countermeasures such as the use of decoy states [7–9] to estimate the amount of information shared with Eve.

In this paper, we are interested in detector control attacks such as blinding attacks [10–13]. When no countermeasure is in place, this attack could possibly allow Eve to gain full information on the key exchanged by Alice and Bob without being noticed. Some protocols such as device-independent protocols [14–18] or measurement-device-independent protocols [19–28] are secure against these attacks but their current performances and certain technical challenges could hamper their deployment in a large-scale QKD network in the near future. For other protocols, like prepare-and-measure protocols, several potential countermeasures have been proposed like monitoring the state of the detector [29,30], measuring some statistical properties [31–33], bit-mapped gating [34], using a variable optical attenuator [35–37], or using a specially designed readout circuit [38–41]. These countermeasures are often designed for a specific type of detector or make assumptions on the attack that can be difficult to meet in practice, potentially compromising the effectiveness of the countermeasure. For example, a countermeasure based on the randomization of Bob's detectors' efficiency (using for example a variable-intensity modulator) was proposed in Ref. [42], but it was later shown to be ineffective against a modified version of the initial attack [43]. Here, we propose a method solely based on detection statistics using multipixel detectors to estimate the maximum information that Eve can have on the key exchanged.

In the next section, we detail the scheme of the attack considered and we present the security principle of our countermeasure using a simple case. Then, we give the results of our analysis in more realistic conditions. Finally, we test a two-pixel detector under blinding attack and show that it can fulfill the requirements for our countermeasure.

II. COUNTERMEASURE

Blinding attacks have been shown to potentially threaten the security of QKD. Indeed, they give the possibility to an adversary, Eve, to change the behavior of Bob's detectors such that she can send what is usually called a "faked state" that can only be detected if Bob chooses the same basis as hers [44]. In this way, Eve can reproduce her measurement outcome without introducing errors in the key. As a countermeasure, we propose to split Bob's detectors into two pixels. Other implementations such as a beam splitter with two detectors could be possible, but we show in Sec. III that the two-pixel detector is a good way to do it. As both pixels correspond to the detection of the same state, our main assumption is that Eve's faked state cannot be used to

*gaetan.gras@idquantique.com

control each pixel independently and that the coincidence detection probability in the presence of the faked states will inevitably increase, revealing Eve's attack. More precisely, we show that the measurement of the probabilities of single and coincidence gives enough information to Alice and Bob to estimate the maximum amount of information that an eavesdropper can have on the key.

The scheme of the attack is shown in Fig. 1.Alice sends weak coherent pulses with a mean photon number μ. Bob's measurement setup is composed of a basis choice (active or passive) and two detectors each split into two pixels. Eve is in the middle and can either perform the blinding attack or simply let the pulse from Alice go through to Bob. We note that p_a is the probability of attack. If Eve lets Alice's pulse go through, Bob's pixel $i \in \{1,2\}$ will click with a probability $p_{B1} = (1+\alpha)p_B$ or $p_{B2} = (1-\alpha)p_B$, where p_B is the average pixel detection probability and α is a coefficient known by Bob characterizing the efficiency mismatch between the pixels. If Eve chooses to intercept Alice's pulse, she measures it using a copy of Bob's setup (called "fake Bob") and she resends her faked state if she detected something. Bob's pixel i will detect this faked state with a probability p_{di} only if his basis choice is the same as Eve's. Otherwise, he will not detect anything. Therefore, the detection probability when Eve carries out her attack depends on the probability that Alice's pulse contains at least one photon $1-e^{-\mu t}$ (t being the transmission coefficient between Alice and Eve's detectors) and on the probability q that Bob and Eve choose the same basis. We call this probability p_E:

$$p_E = (1-e^{-\mu t})q. \tag{1}$$

By naming p_{s1} and p_{s2} the probabilities of detection of both pixels measured by Bob, we then can write

$$\begin{aligned} p_{s1} &= p_a p_E \sum_\lambda p^\lambda p_{d1}^\lambda + (1-p_a)(1+\alpha)p_B, \\ p_{s2} &= p_a p_E \sum_\lambda p^\lambda p_{d2}^\lambda + (1-p_a)(1-\alpha)p_B. \end{aligned} \tag{2}$$

We give Eve the possibility of using different strategies λ from one pulse to the other, each with a probability p^λ. We suppose both pixels are independent from each other. Thus, the probability that a faked state generates a coincidence is $p_{d1}p_{d2}$. The probability of coincidence for the two pixels is then

$$p_c = p_a p_E \sum_\lambda p^\lambda p_{d1}^\lambda p_{d2}^\lambda + (1-p_a)(1-\alpha^2)p_B^2. \tag{3}$$

By analyzing the coincidence probability between both pixels, we show how Alice and Bob can bound the information leaked to Eve.

A. Asymptotic case

In this section, we first want to convey the idea behind this countermeasure by considering a simple case where we are in the asymptotic limit and both pixels are perfectly identical ($p_{d1}^\lambda = p_{d2}^\lambda$ and $p_{B1} = p_{B2}$). The attack scenario defined by Eqs. (2) and (3) can be rewritten as

$$\begin{aligned} p_s &= p_a p_E \sum_\lambda p^\lambda p_d^\lambda + (1-p_a)p_B, \\ p_c &= p_a p_E \sum_\lambda p^\lambda (p_d^\lambda)^2 + (1-p_a)p_B^2. \end{aligned} \tag{4}$$

We define the ratio $r = p_c/p_s^2$ (note that this is similar to a second-order correlation measurement g_2; we call it r simply because, with the attack, it is not really a measurement of the photon statistics). In the limit $p_a = 0$, $r = 1$ as expected for coherent states. On the other hand, if $p_a = 1$, we have

$$r = \frac{p_c}{p_s^2} = \frac{\sum_\lambda p^\lambda (p_d^\lambda)^2}{p_E\left(\sum_\lambda p^\lambda p_d^\lambda\right)^2} \geq \frac{1}{p_E} > 1. \tag{5}$$

As we can see, the value of r induced by the attack is limited by the probability p_E, which depends on the vacuum probability in Alice's pulses and q. Let us now see how we can estimate Eve's information per bit I_E on the raw key in the case she attacks only a fraction of the pulses, i.e., $0 < p_a < 1$. As Eve knows the measurement outcome of Bob only when he detects a faked state, we want to maximize

$$I_E = \frac{p_a p_E \sum_\lambda p^\lambda p_d^\lambda}{p_s}, \tag{6}$$

given p_E, p_s, and p_c. Using the Lagrangian multiplier, we can show that Eve's best strategy is to always resend a pulse with the same probability of detection $p_d^\lambda = p_d$, $\forall\lambda$, and we find her maximum information is given by (see Appendix A 1)

$$I_{E,\max} = \frac{\sqrt{p_E}(\sqrt{p_c}-p_s)}{p_s(1-\sqrt{p_E})} = \frac{\sqrt{p_E}}{(1-\sqrt{p_E})}\left(\sqrt{r}-1\right). \tag{7}$$

As expected, Eve's information increases with the ratio $r = p_c/p_s^2$ measured by Bob and $I_{E,\max} = 1$ when $r = 1/p_E$.

In a more realistic scenario, Bob's pixels will not be perfectly identical. This is the scenario described by Eqs. (2) and (3). Without additional constraint on p_{d1}^λ and p_{d2}^λ, Eve can alternatively target pixel 1 ($p_{d1}^{(1)} \gg p_{d2}^{(1)}$) and pixel 2 ($p_{d2}^{(2)} \gg p_{d1}^{(2)}$) to reduce her coincidence probability and hide her presence from our countermeasure. On the other hand, a complete characterization of all detectors under all possible attack conditions in order to find bounds on

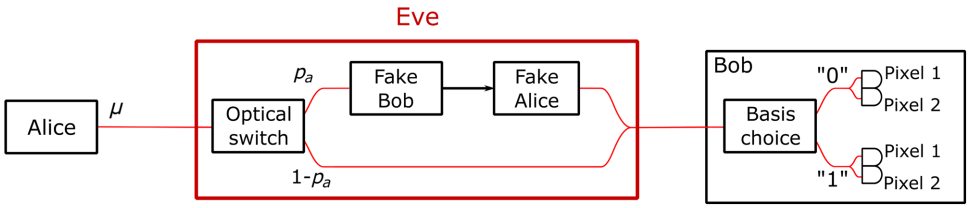

FIG. 1. Scheme of the attack. Alice sends pulses with a mean photon number per pulse μ. Eve intercepts the pulse with a probability p_a. If she gets a conclusive event with her "fake Bob," she resends a pulse to force Bob's detector to click; otherwise, she does nothing. Bob's apparatus is unchanged except for his detectors being split in two. Coincidences between the two pixels are kept to generate the key.

p_{d2} given p_{d1} seems an unpractical task. We circumvent this problem by adding the assumption that one pixel will always detect Eve's faked state with an equal or higher probability than the other. This constraint on the attack can be written as

$$p_{d2}^{\lambda} \geq p_{d1}^{\lambda}, \ \forall \lambda. \tag{8}$$

In this way, we prevent Eve from targeting preferably pixel 1. We show in Sec. III that this condition can be realized with a two-pixel detector. By applying the Lagrange multiplier with this additional constraint, we can calculate all the extrema of I_E to find the maximum of Eve's information $I_{E,\max}$. Here, we limit the number of strategies to two as increasing the number of strategies does not give much more information to Eve if the difference between p_{s1} and p_{s2} stays small. Indeed, in that case, Eve is forced to make both pixels click with the same probability most of the time to keep the probabilities of detection close. In a real system, the protocol can be aborted if the difference between p_{s1} and p_{s2} exceeds a certain threshold. Details of the calculations are given in Appendix A 2.

B. Finite key analysis

In order to take into account finite key length effects, we need to bound the probabilities of single and coincidence measured by Bob. Usually, QKD proofs rely on Hoeffding's inequality to calculate upper and lower bounds on measured values. However, in our countermeasure, the probability of coincidence will drop very quickly with the quantum channel length and in this case, Hoeffding's inequality is no longer tight. This would lead to an overestimation of Eve's information making our countermeasure usable only for short distances. In order to have a tighter bound on Bob's probabilities, we can use the equations given in Ref. [45]. The upper and lower bounds on p_{si} and p_c are given by

$$\begin{aligned} p_c^u &= 1 - I_\epsilon^{-1}[N(1-p_c), Np_c + 1], \\ p_{si}^l &= I_\epsilon^{-1}[Np_{si}, N(1-p_{si}) + 1], \end{aligned} \tag{9}$$

where N is the total number of pulses sent by Alice, ϵ our confidence factor, and I^{-1} the inverse incomplete beta function. By inserting these bounds in the calculation of $I_{E,\max}$, we obtain an upper bound on Eve's information $I_{E,\max}^u$, which can be reduced to zero after privacy amplification.

Figure 2shows simulations of $I_{E,\max}^u$ for a BB84 protocol. We run the simulations for different acquisition times (ATs) for Bob. As the quantum channel length increases, the probability of coincidence measured by Bob decreases rapidly requiring longer ATs to limit the uncertainty. If the uncertainty is too high, Alice and Bob may overestimate $I_{E,\max}^u$, which impacts the final secret key rate. Therefore, the factor ultimately limiting our countermeasure is the AT allowed by Alice and Bob. For most applications, an AT over 24 h becomes impractical [9], allowing our countermeasure to be efficient for distances of around 250 km, which is close to the limit of many current QKD implementations.

III. EXPERIMENTAL RESULTS

In this section, we show that actual detectors can fulfill the condition given by Eq. (8) for our countermeasure

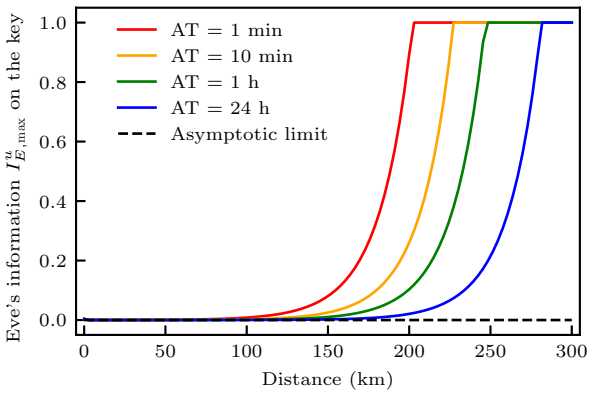

FIG. 2. Upper bound on Eve's information of the raw key as a function of the channel length between Alice and Bob for different AT and $\epsilon = 10^{-9}$. The protocol used is a BB84 with a passive basis choice. Alice sends pulses with a mean photon number $\mu = 0.5$ at a rate of 5 GHz. Losses in the channel are 0.2 dB/km. Bob's pixels have a quantum efficiency of 25% each giving a total efficiency of 50% for the whole detector.

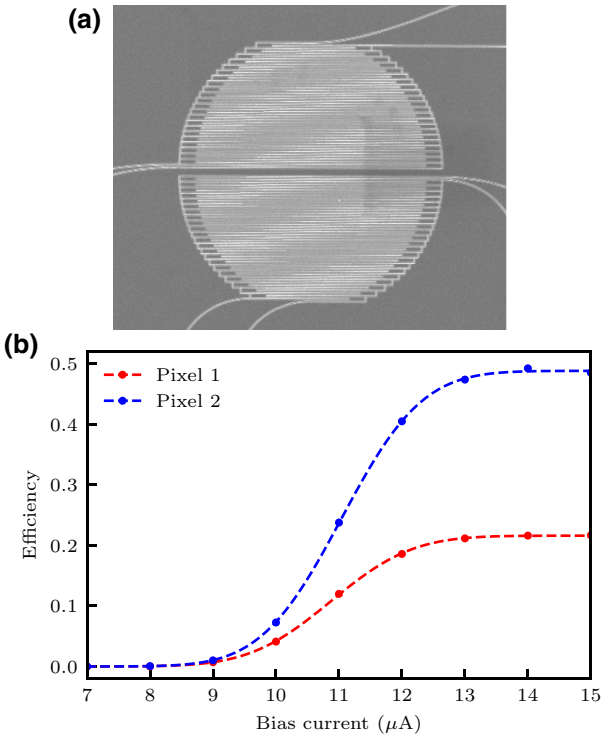

FIG. 3. (a) SEM image of a two-element molybdenum silicide SNSPD. Each pixel has its own bias current and readout circuit. The nanowire width is 100 nm with a fill factor of 0.6 [46]. The two pixels are separated by 600 nm to avoid thermal crosstalk between them. (b) Efficiency curves at 1550 nm of the two pixels of the detector operated at 0.8 K versus the bias current.

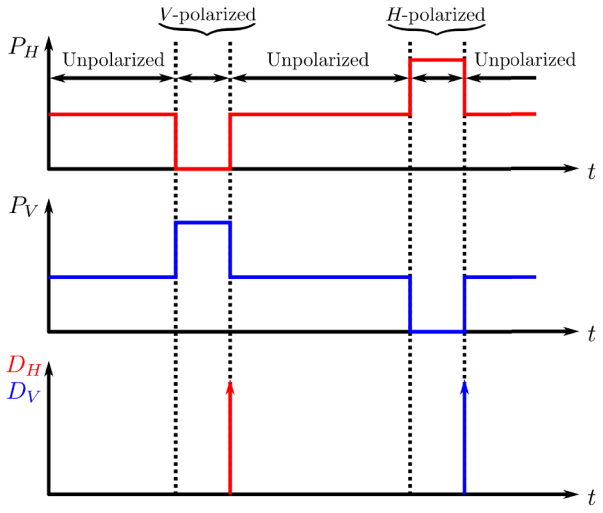

FIG. 4. Schematic representation of the blinding power distribution on detectors D_H and D_V during the attack on a BB84 QKD protocol based on polarization. By changing the polarization of her blinding light, Eve can let the detector of her choice partially recover its bias current to force it then to click.

against blinding attacks. To do so, we fabricate and test multipixel superconducting nanowire single-photon detectors (SNSPDs), as depicted in Fig. 3(a).The two pixels are separated by a gap of 600 nm in order to avoid thermal crosstalk. This gap has a small impact on the performances of the detector as we measure an overall quantum efficiency of 70% [see Fig. 3(b)]. We also note that both pixels have very similar efficiency curves (except for the optimum efficiency, which is probably due to a misalignment with the fiber). The main advantage of this design is that both pixels are illuminated by a single fiber, limiting the dependency of the light distribution on the wavelength used by Eve for her attack compared to an implementation with a beam splitter and two distinct detectors [47]. For even better security, the addition of a mode scrambler could prevent Eve from using smaller wavelengths where the fiber becomes multimode [48].

To illustrate how a blinding attack on a QKD system using this kind of detector works, we take as an example a BB84 protocol in polarization. In normal operation, when a photon hits the SNSPD, it will break the superconductivity inducing a rapid increase of the resistance of the nanowire. This sudden change of resistance will divert the bias current of the detector toward the readout circuit to generate a click. In order to blind Bob's detectors, Eve sends unpolarized light of a few hundreds of nanowatts inside Bob's setup such that her blinding power is equally distributed over all detectors. This forces the SNSPDs to stay in a resistive state where they are insensitive to single photons. When Eve wants to force Bob to detect the state of her choice, say $|H\rangle$, she polarizes her blinding light vertically for a time Δt. During this time, the optical power arriving on detector D_H will be greatly reduced (around 20 to 30 dB depending on Bob's components) while keeping the other detectors blinded.

By unpolarizing her blinding light after Δt, the optical power P_H arriving on the detector D_H will increase suddenly, forcing it to click as part of the current would have returned to the nanowire (see Fig. 4).Eve can control the probability p to force the detector to click by allowing more or less current to return to the detector via Δt. Many parameters have an influence on the probability of detection of the faked state. Some are controlled by Eve (blinding power P_{blind}, Δt) and some are controlled by Bob (bias current). However, as we mentioned in Sec. II A, if we can find a regime where one pixel always clicks with a probability greater than the second one (whatever are the parameters of the attack) then this gives enough constraints on Eve to ensure she cannot steal the key without being noticed. As the probability of click depends on the amount of current that returns to the nanowire, we want one pixel to recover its current more rapidly such that it will detect the faked state with a higher probability than the second pixel. For that, we set pixel 2 at its maximum bias current (15 μA) while pixel 1 is set at a bias current of 12.5 μA. This way, the current will return more rapidly

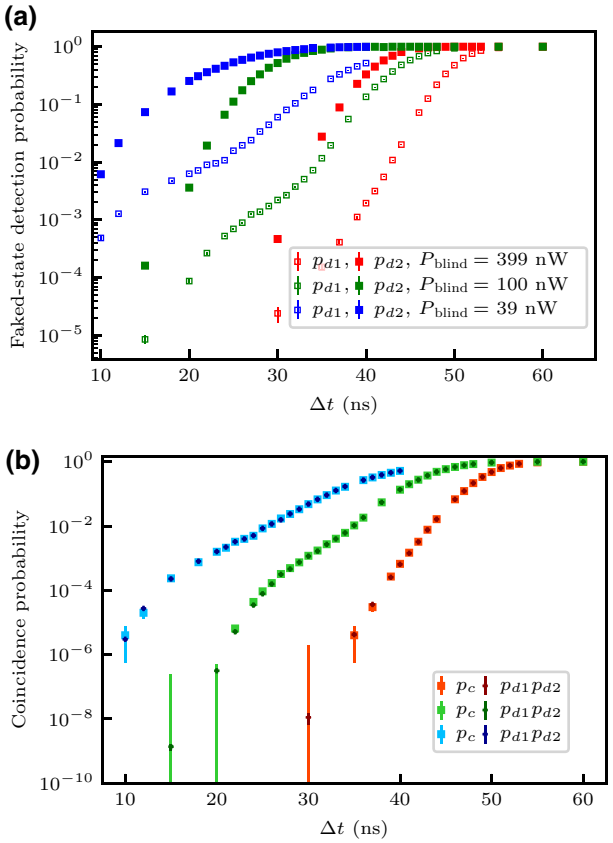

FIG. 5. (a) Probability of detection of the faked state as a function of Δt. Pixel 1: $I_{b1} = 12.5\ \mu$A; pixel 2: $I_{b2} = 15\ \mu$A. We vary the blinding power between 39 and 399 nW as it is the working range for the blinding attack. (b) Comparison between the measured coincidence probability and the coincidence probability calculated from the faked-state detection probabilities of both pixels.

to pixel 2 without impacting the overall efficiency of the detector [49].

We measure the probabilities of detection of both pixels as a function of Δt by sending the faked state at a frequency of 500 kHz and recording the detection rates with a counter. These measurements are made for blinding powers ranging from 39 nW (minimal blinding power) up to 399 nW. For higher P_{blind}, the pixels start to click in an uncontrolled way before Δt making the attack unfeasible as it would increase the error rate. We can see in Fig. 5(a)that $p_{d2} \geq p_{d1}$ for the whole range of working P_{blind} and Δt as we assume in our model. We then verify that the probabilities of detecting the faked state are uncorrelated. For that, we measure the coincidence probability p_c due to the faked state and compare it with the product of the individual detection probabilities $p_{d1}p_{d2}$ (value expected if the pixels are independent). Results are shown in Fig. 5(b). As we can see with the error bars, both values are in the uncertainty range of each other. No statistically significant signature of correlations is observable, validating the assumption made in our analysis. Thus, this multipixel detector fulfills all the requirements for our countermeasure.

This countermeasure could also work with single-photon avalanche diode detectors as the core idea behind our proposal does not rely on the working principle of the detectors. Further tests with this kind of detector need to be done to validate that it fulfills all the necessary conditions.

IV. CONCLUSION

In this paper, we propose a countermeasure against detector control attacks based on multipixel detectors, which, unlike previous works [31,32], does not assume a binary response of the pixels (i.e., p_{di} is equal to either 0 or 1) under the blinding attack. With this countermeasure, we take advantage of Eve's lack of knowledge on the state prepared by Alice when the incoming pulse contains zero photons. Because of this method, we are able to estimate an upper bound on the information leaked to the adversary solely using the single and coincidence probabilities measured by Bob. The effectiveness of our countermeasure over long distances is ultimately limited by the key exchange time between Alice and Bob. Nevertheless, we show that communications close to 250 km can be secured against attack with acquisition times of less than 24 h. Finally, we experimentally demonstrate that a multipixel SNSPD operated in the right conditions by Bob can satisfy the assumptions made in our analysis.

ACKNOWLEDGMENTS

This project has received funding from the research and innovation programme under Grant Agreement No. 675662. We thank Claire Autebert for designing and fabricating the detectors. We also thank Jean-Daniel Bancal and Nicolas Gisin for helpful discussions.

APPENDIX: LAGRANGE MULTIPLIER CALCULATIONS

1. Simple case

In order to find Eve's best strategy, we want to maximize the number of detections coming from faked states $n_a = Np_ap_E \sum_\lambda p^\lambda p_d^\lambda$ (with n being the total number of pulses sent by Alice) over the total number of detections n under the constraints given by Eq. (4). As n and N are fixed values, we can maximize the function f defined by

$$f = p_a p_E \sum_\lambda p^\lambda p_d^\lambda. \tag{A1}$$

We define the following equations representing our constraints:

$$\begin{aligned} g_1 &= p_a p_E \sum_\lambda p^\lambda p_d^\lambda + (1-p_a)p_B - p_s, \\ g_2 &= p_a p_E \sum_\lambda p^\lambda (p_d^\lambda)^2 + (1-p_a)p_B^2 - p_c, \\ g_3 &= \sum_\lambda p^\lambda - 1. \end{aligned} \tag{A2}$$

We can then define our Lagrange function:

$$\mathcal{L}(p_a, p^\lambda, p_d^\lambda, p_B, \Lambda_1, \Lambda_2, \Lambda_3) = f - \Lambda_1 g_1 - \Lambda_2 g_2 - \Lambda_3 g_3. \tag{A3}$$

The function f is maximum if

$$\nabla \mathcal{L} = 0. \tag{A4}$$

To show that Eve's best strategy is to always send the faked state with the same probability of detection, we take the derivatives:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial p_d^\lambda} &= p_a p_E p^\lambda - \Lambda_1 p_a p_E p^\lambda - 2\Lambda_2 p_a p_E p^\lambda p_d^\lambda \\ &= p_a p_E p^\lambda \left(1 - \Lambda_1 - 2\Lambda_2 p_d^\lambda\right) \\ &= 0. \end{aligned} \tag{A5}$$

This expression is valid only if $1 - \Lambda_1 - 2\Lambda_2 p_d^\lambda = 0,\ \forall\lambda$ (we neglect the case $p_a = 0$ as it would mean that Eve never does the attack and the case $p^\lambda = 0$ as it would be a strategy Eve never uses). Therefore, either p_d^λ is a constant or $\Lambda_1 = 1$ and $\Lambda_2 = 0$. The latter case is impossible as we can see by looking at an another derivative:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial p_B} &= -(1-p_a)(\Lambda_1 + 2\Lambda_2 p_B) \\ &= 0. \end{aligned} \tag{A6}$$

The solution $p_a = 1$ is possible only if $p_c/p_s^2 \geq 1/p_E$. Otherwise, $\Lambda_1 + 2\Lambda_2 p_B = 0$, which is incompatible with $(\Lambda_1, \Lambda_2) = (1, 0)$. Consequently, Eve's best strategy is to use the same $p_d^\lambda = p_d,\ \forall\lambda$. These results simplify our problem that we can rewrite as follows:

$$\begin{aligned} f &= p_a p_E p_d, \\ g_1 &= p_a p_E p_d + (1-p_a)p_B - p_s, \\ g_2 &= p_a p_E p_d^2 + (1-p_a)p_B^2 - p_c, \\ \mathcal{L} &= f - \Lambda_1 g_1 - \Lambda_2 g_2, \\ \nabla \mathcal{L} &= 0. \end{aligned} \tag{A7}$$

This system has a unique solution:

$$\begin{aligned} p_B &= \sqrt{p_c}, \\ p_d &= \sqrt{\frac{p_c}{p_E}}, \\ p_a &= \frac{\sqrt{p_c} - p_s}{\sqrt{p_c}(1 - \sqrt{p_E})}, \end{aligned} \tag{A8}$$

which finally gives us

$$\begin{aligned} I_{E,\max} &= \frac{n_a}{n} \\ &= \frac{\sqrt{p_E}(\sqrt{p_c} - p_s)}{p_s(1 - \sqrt{p_E})}. \end{aligned} \tag{A9}$$

2. General case

In the general case given by Eqs. (2) and (3), we can apply the same method where our problem is described by the following equations:

$$\begin{aligned} f &= p_a p_E \sum p^\lambda (p_{d1}^\lambda + p_{d2}^\lambda), \\ g_1 &= p_a p_E \sum p^\lambda p_{d1}^\lambda + (1-p_a)(1+\alpha)p_B - p_{s1}, \\ g_2 &= p_a p_E \sum p^\lambda p_{d2}^\lambda + (1-p_a)(1-\alpha)p_B - p_{s2}, \\ g_c &= p_a p_E \sum p^\lambda p_{d1}^\lambda p_{d2}^\lambda + (1-p_a)(1-\alpha^2)p_B^2 - p_c, \\ \mathcal{L} &= f - \Lambda_1 g_1 - \Lambda_2 g_2 - \Lambda_c g_c, \\ \nabla \mathcal{L} &= 0. \end{aligned} \tag{A10}$$

The optimization is done taking into account the physical constraints on the attack parameters: all probabilities must be between 0 and 1 and $p_{d2}^\lambda \geq p_{d1}^\lambda,\ \forall\lambda$. The resolution of the system gives us all the extrema of the function f. By discarding nonphysical solutions and taking the highest of the remaining values, we obtain the maximum of Eve's information on the key.

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

[2] B. Huttner, N. Imoto, N. Gisin, and T. Mor, Quantum cryptography with coherent states, Phys. Rev. A **51**, 1863 (1995).

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

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

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

[6] S. Sajeed, C. Minshull, N. Jain, and V. Makarov, Invisible trojan-horse attack, Sci. Rep. **7**, 8403 (2017).

[7] H.-K. Lo, X. Ma, and K. Chen, Decoy State Quantum Key Distribution, Phys. Rev. Lett. **94**, 230504 (2005).

[8] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, Practical decoy state for quantum key distribution, Phys. Rev. A **72**, 012326 (2005).

[9] D. Rusca, A. Boaron, F. Grünenfelder, A. Martin, and H. Zbinden, Finite-key analysis for the 1-decoy state QKD protocol, Appl. Phys. Lett. **112**, 171104 (2018).

[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 (2010).

[11] L. Lydersen, J. Skaar, and V. Makarov, Tailored bright illumination attack on distributed-phase-reference protocols, J. Mod. Opt. **58**, 680 (2011).

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

[13] M. G. Tanner, V. Makarov, and R. H. Hadfield, Optimised quantum hacking of superconducting nanowire single-photon detectors, Opt. Express **22**, 6734 (2014).

[14] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Device-Independent Security of Quantum Cryptography Against Collective Attacks, Phys. Rev. Lett. **98**, 230501 (2007).

[15] S. Pironio, A. Acín, N. Brunner, N. Gisin, S. Massar, and V. Scarani, Device-independent quantum key distribution secure against collective attacks, New J. Phys. **11**, 045021 (2009).

[16] N. Gisin, S. Pironio, and N. Sangouard, Proposal for Implementing Device-Independent Quantum Key Distribution Based on a Heralded Qubit Amplifier, Phys. Rev. Lett. **105**, 070501 (2010).

[17] L. Masanes, S. Pironio, and A. Acín, Secure device-independent quantum key distribution with causally independent measurement devices, Nat. Commun. **2**, 238 (2011).

[18] U. Vazirani and T. Vidick, Fully Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **113**, 140501 (2014).

[19] H.-K. Lo, M. Curty, and B. Qi, Measurement-Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **108**, 130503 (2012).

[20] 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).

[21] Y. Liu, T.-Y. Chen, L.-J. Wang, H. Liang, G.-L. Shentu, J. Wang, K. Cui, H.-L. Yin, N.-L. Liu, L. Li, X. Ma, J. S. Pelc, M. M. Fejer, C.-Z. Peng, Q. Zhang, and J.-W. Pan, Experimental Measurement-Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **111**, 130502 (2013).

[22] M. Curty, F. Xu, W. Cui, C. C. W. Lim, K. Tamaki, and H.-K. Lo, Finite-key analysis for measurement-device-independent quantum key distribution, Nat. Commun. **5**, 3732 (2014).

[23] Z. Tang, Z. Liao, F. Xu, B. Qi, L. Qian, and H.-K. Lo, Experimental Demonstration of Polarization Encoding Measurement-Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **112**, 190503 (2014).

[24] 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, W.-J. Zhang, H. Chen, M. J. Li, D. Nolan, F. Zhou, X. Jiang, Z. Wang, Q. Zhang, X.-B. Wang, and J.-W. Pan, Measurement-Device-Independent Quantum Key Distribution Over a 404 km Optical Fiber, Phys. Rev. Lett. **117**, 190501 (2016).

[25] 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 (2018).

[26] M. Minder, M. Pittaluga, G. Roberts, M. Lucamarini, J. Dynes, Z. Yuan, and A. Shields, Experimental quantum key distribution beyond the repeaterless secret key capacity, Nat. Photonics **13**, 1 (2019).

[27] 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).

[28] J.-P. Chen, C. Zhang, Y. Liu, C. Jiang, W. Zhang, X.-L. Hu, J.-Y. Guan, Z.-W. Yu, H. Xu, J. Lin, M.-J. Li, H. Chen, H. Li, L. You, Z. Wang, X.-B. Wang, Q. Zhang, and J.-W. Pan, Sending-Or-Not-Sending with Independent Lasers: Secure Twin-Field Quantum key Distribution Over 509 km, Phys. Rev. Lett. **124**, 070501 (2020).

[29] Ø. Marøy, V. Makarov, and J. Skaar, Secure detection in quantum key distribution by real-time calibration of receiver, Quantum Sci. Technol. **2**, 044013 (2017).

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

[31] T. Honjo, M. Fujiwara, K. Shimizu, K. Tamaki, S. Miki, T. Yamashita, H. Terai, Z. Wang, and M. Sasaki, Countermeasure against tailored bright illumination attack for DPS-QKD, Opt. Express **21**, 2667 (2013).

[32] T. Ferreira da Silva, G. C. do Amaral, G. B. Xavier, G. P. Temporão, and J. P. von der Weid, Safeguarding quantum key distribution through detection randomization, IEEE J. Sel. Top. Quantum Electron. **21**, 159 (2015).

[33] J. Wang, H. Wang, X. Qin, Z. Wei, and Z. Zhang, The countermeasures against the blinding attack in quantum key distribution, Eur. Phys. J. D **70**, 5 (2016).

[34] L. Lydersen, V. Makarov, and J. Skaar, Secure gated detection scheme for quantum cryptography, Phys. Rev. A **83**, 032306 (2011).

[35] A. Koehler-Sidki, M. Lucamarini, J. F. Dynes, G. L. Roberts, A. W. Sharpe, Z. Yuan, and A. J. Shields, Intensity modulation as a preemptive measure against blinding of single-photon detectors based on self-differencing cancellation, Phys. Rev. A **98**, 022327 (2018).

[36] M. Alhussein and K. Inoue, Differential phase shift quantum key distribution with variable loss revealing blinding and control side-channel attacks, Jpn. J. Appl. Phys. **58**, 102001 (2019).

[37] Y.-J. Qian, D.-Y. He, S. Wang, W. Chen, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han, Robust countermeasure against detector control attack in a practical quantum key distribution system, Optica **6**, 1178 (2019).

[38] Z. L. Yuan, J. F. Dynes, and A. J. Shields, Avoiding the blinding attack in QKD, Nat. Photonics **4**, 800 (2010).

[39] Z. L. Yuan, J. F. Dynes, and A. J. Shields, Resilience of gated avalanche photodiodes against bright illumination attacks in quantum cryptography, Appl. Phys. Lett. **98**, 231104 (2011).

[40] M. S. Lee, B. K. Park, M. K. Woo, C. H. Park, Y.-S. Kim, S.-W. Han, and S. Moon, Countermeasure against blinding attacks on low-noise detectors with a background-noise-cancellation scheme, Phys. Rev. A **94**, 062321 (2016).

[41] A. Koehler-Sidki, J. F. Dynes, M. Lucamarini, G. L. Roberts, A. W. Sharpe, Z. L. Yuan, and A. J. Shields, Best-Practice Criteria for Practical Security of Self-Differencing Avalanche Photodiode Detectors in Quantum Key Distribution, Phys. Rev. Appl. **9**, 044027 (2018).

[42] C. C. W. Lim, N. Walenta, M. Legré, N. Gisin, and H. Zbinden, Random variation of detector efficiency: A countermeasure against detector blinding attacks for quantum key distribution, IEEE J. Sel. Top. Quantum Electron. **21**, 192 (2015).

[43] A. Huang, S. Sajeed, P. Chaiwongkhot, M. Soucarros, M. Legré, and V. Makarov, Testing random-detector-efficiency countermeasure in a commercial system reveals a breakable unrealistic assumption, IEEE J. Quantum Electron. **52**, 1 (2016).

[44] V. Makarov and D. R. Hjelme, Faked states attack on quantum cryptosystems, J. Mod. Opt. **52**, 691 (2005).

[45] J.-D. Bancal, K. Redeker, P. Sekatski, W. Rosenfeld, and N. Sangouard, Self-testing with finite statistics enabling the certification of a quantum network link, Quantum **5**, 401 (2021).

[46] M. Caloz, M. Perrenoud, C. Autebert, B. Korzh, M. Weiss, C. Schönenberger, R. J. Warburton, H. Zbinden, and F. Bussières, High-detection efficiency and low-timing jitter with amorphous superconducting nanowire single-photon detectors, Appl. Phys. Lett. **112**, 061103 (2018).

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

[48] G. Gras and F. Bussières, Patent Publication No WO2019121783A1 (2019).

[49] C. Autebert, G. Gras, E. Amri, M. Perrenoud, M. Caloz, H. Zbinden, and F. Bussières, Direct measurement of the recovery time of superconducting nanowire single-photon detectors, J. Appl. Phys. **128**, 074504 (2020).