Practical Security Bounds Against the Trojan-Horse Attack in Quantum Key Distribution

M. Lucamarini,[1,2] I. Choi,[1] M. B. Ward,[1] J. F. Dynes,[1,2] Z. L. Yuan,[1,2] and A. J. Shields[1,2]

[1]*Toshiba Research Europe Limited, 208 Cambridge Science Park, Cambridge CB4 0GZ, United Kingdom*

[2]*Corporate Research & Development Center, Toshiba Corporation, 1 Komukai-Toshiba-Cho, Saiwai-ku, Kawasaki 212-8582, Japan*

(Received 15 March 2015; revised manuscript received 5 June 2015; published 9 September 2015)

In the quantum version of a Trojan-horse attack, photons are injected into the optical modules of a quantum key distribution system in an attempt to read information direct from the encoding devices. To stop the Trojan photons, the use of passive optical components has been suggested. However, to date, there is no quantitative bound that specifies such components in relation to the security of the system. Here, we turn the Trojan-horse attack into an information leakage problem. This allows us to quantify the system security and relate it to the specification of the optical elements. The analysis is supported by the experimental characterization, within the operation regime, of reflectivity and transmission of the optical components most relevant to security.

DOI: 10.1103/PhysRevX.5.031030

Subject Areas: Optoelectronics, Photonics, Quantum Information

I. INTRODUCTION

Since ancient times, the Trojan horse has been known as a stratagem for penetrating a securely protected space. It is therefore essential to consider Trojan-horse attacks in determining the boundaries of any supposedly secure space. This explains their ubiquitous presence in different fields where privacy is required, ranging from cryptography to computing and finance. In particular, for a cryptographic application like quantum key distribution (QKD) [1–4], as well as for its most recent developments showing full or partial independency from the specific devices used [5–10], the existence of a protected area is a fundamental assumption.

QKD allows two remote parties, usually called Alice (transmitter) and Bob (receiver), to share a common secret key with information theoretical security, over an insecure quantum channel and an authenticated or broadcast classical channel. QKD's security derives from the laws of quantum physics, and its implementation necessarily makes use of physical systems, whose correct behavior has to be characterized and guaranteed against unwanted imperfections. Any ignored deviation from the expected behavior can be exploited by an attacker (Eve) to compromise the system security. In Fig. 1, the Trojan-horse attack (THA) against an optical QKD setup is sketched. Eve uses the optical channel connecting Alice and Bob to launch a bright light pulse containing Trojan photons into Alice's supposedly secure module. The light pulse reaches the encoding device and is encoded with the same information φ as the photon normally prepared by Alice and then sent to Bob. The information φ is meant to be private. However, some of the Trojan photons are reflected back, and they deliver the information to Eve, thus compromising the security of the system.

This eavesdropping strategy was initially described in Ref. [11] and afterwards named "Trojan-horse attack" in Ref. [12]. Because of its apparent simplicity, the THA has often been considered easily tractable. However, to date, there is no quantitative analysis to mitigate it, and there is an increasing number of experiments showing its severity instead [12–16]. For example, it was shown in Ref. [12] that phase information can be extracted from a $LiNbO_3$-based encoding device using optical-frequency-domain reflectometry. More recently, it has been demonstrated that phase values from an encoding device can be discriminated with 90% success probability using only three photons [13].

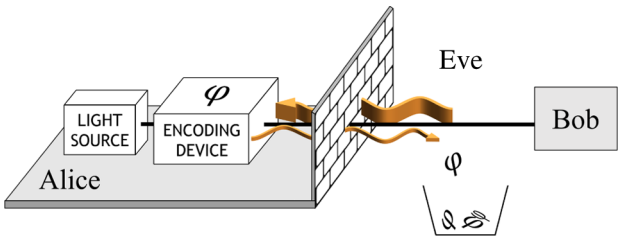

FIG. 1. Representation of the Trojan-horse attack against an optical QKD setup. Eve sends a large amount of Trojan photons (thick arrow) against Alice's defensive structure. Some of the photons reach the encoding device, are encoded with the private information φ, and are reflected back to Eve (thin arrow), who retrieves the information by measuring the photons.

To counteract the THA, different solutions have been proposed. On the one hand, we have active countermeasures, similar to the ones used to ensure the security of the "plug-and-play" QKD setup [17–20]. Alice could be endowed with an active phase randomizer [12,21] and a watchdog detector [17,18] to remove the phase reference from Eve's hands and bound the energy of the incoming light pulses. However, active components usually add extra complexity to the setup and may offer more options to the eavesdropper [14]. For instance, it has been shown recently that a monitoring detector of a commercial QKD system can be bypassed easily [15]. On the other hand, passive countermeasures can be realized with much simpler elements, e.g., optical fiber loops, filters, and isolators, which leave fewer degrees of freedom for the eavesdropper. Furthermore, they are often inexpensive, and simple to implement and to characterize experimentally. However, in this case, powerful resources like the phase randomization and the watchdog detector cannot be used to prove the security of the system.

As a result, the security analysis of the THA remains elusive, and no security-proof solution has been derived to date. The only provably secure countermeasures are for users endowed with a teleportation filter [22] or for the receiver in a system running the BB84 protocol [1,11]. In the former case, the solution is not practical, and it entails considerable changes in the setup that could open additional loopholes. In the latter case, a delay line installed at the entrance of Bob's module prevents Eve from reading the basis information before the qubit has entered Bob's protected territory. However, the same measure is ineffective in protecting the transmitting side of the QKD system, nor does it apply to other protocols such as the B92 [23] and the SARG04 [24]. Hence, it cannot be considered a general solution against the THA.

In this work, we analyze an entirely passive architecture to counteract the THA. We provide quantitative bounds that connect the values of the passive optical components to the security of the QKD system. The key element is interpreting the THA as a side channel. Normally, Alice is unaware of it and treats her preparation as ideal. This causes undetected leakage of information from her module to Eve's territory. However, if Alice characterizes the relevant optical components in her apparatus, she can bound the information leakage and attain security through an adequate level of privacy amplification.

II. THEORETICAL DESCRIPTION

Let us consider the transmitter module [25] in the unidirectional, fiber-based, phase-modulated QKD setup depicted in Fig. 2. In the THA, Eve injects light into Alice's apparatus through the same optical fiber that serves as a quantum channel between the users (thick arrow in the figure). The goal is to reach the phase modulator that encodes the private information φ_A. A concrete possibility

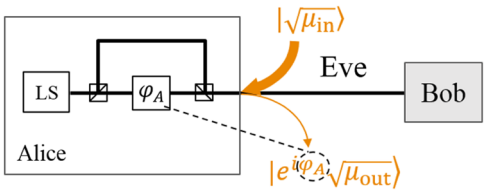

FIG. 2. Schematics of the transmitting unit of a unidirectional fiber-based QKD setup and Eve's THA. LS is a generic light source. The square with φ_A is the encoding device. It writes the phase information φ_A on photons traveling in the short arm of the interferometer. Eve injects a bright light pulse in the coherent state $|\sqrt{\mu_{\text{in}}}\rangle$ into Alice's module. A fraction of it is encoded by Alice and back-reflected to Eve, emerging as $|e^{i\varphi_A}\sqrt{\mu_{\text{out}}}\rangle$, i.e., attenuated by a factor γ ($\mu_{\text{out}} = \gamma\mu_{\text{in}}$) but containing the phase information φ_A (dashed line).

for Eve is to use a laser emitting pulses with average photon number μ_{in}, prepared in a coherent state $|\sqrt{\mu_{\text{in}}}\rangle$ [26]. The pulses acquire the phase modulation information φ_A and return to Eve as $|e^{i\varphi_A}\sqrt{\mu_{\text{out}}}\rangle$ (thin arrow in Fig. 2), where $\mu_{\text{out}} = \gamma\mu_{\text{in}}$, with $\gamma \ll 1$ the optical isolation of the transmitting unit. The light pulse retrieved by Eve is correlated to the phase φ_A, and this compromises the security of the system.

To prevent the THA, Eve's action has to be bounded by a physical mechanism. In particular, it is clear that if the intensity μ_{in} is unbounded, no solution can exist against the THA. On the contrary, when μ_{in} is bounded, Alice can adjust the value of the optical isolation γ to make μ_{out}, and therefore Eve's information, arbitrarily small. In this work, we consider the laser induced damage threshold (LIDT) as the main physical mechanism limiting Eve's action. The LIDT provides an estimate of the energy, thence of the number of photons, that Eve can inject into Alice's module in a characteristic time interval without damaging it. Details about the LIDT are given in Sec. III. For the moment, we call N the maximum number of photons that Eve is allowed to inject in the transmitter module in the time unit (1 second) without violating the LIDT condition. This parameter will be used to provide a security argument against the THA.

A. Preliminary quantities

In a THA, Eve first prepares M groups of photons and then uses each group to probe a different value of Alice's phase modulator (PM). To fix ideas, we can imagine that each group of photons physically corresponds to one pulse of Eve's light source and that each pulse is prepared in a pure coherent state [27]. The resulting structure is a tensor product of coherent states:

$$|\sqrt{\mu_1}\rangle \otimes |\sqrt{\mu_2}\rangle \otimes \ldots \otimes |\sqrt{\mu_M}\rangle. \tag{1}$$

In Eq. (1), μ_i ($i = 1, \ldots, M$) is the mean photon number of the ith coherent state. In order to not overcome the

LIDT threshold N, Eve has to guarantee the following condition:

$$\sum_{i=1}^{M} \mu_i = M\mu_{\text{in}} < N, \tag{2}$$

where we have introduced the overall mean photon number of Eve's light μ_{in}. In general, it is possible for Eve to vary each μ_i to enhance her strategy. However, it turns out that this gives her no advantage, as we show later. The convexity of the key rate as a function of μ_i makes it always better for Eve to set μ_i equal to a constant value. Therefore, we have

$$\mu_i = \mu_{\text{in}}. \tag{3}$$

It can be noted that Eq. (3) rules out a whole class of attacking strategies by Eve, where she redistributes her initial Trojan photons in a fewer number of pulses. Intuitively, this could increase Eve's information on a subset of Alice's states, but it can never increase her total information about the whole key. It is beneficial to Eve to distribute her photons evenly among the available pulses to maximize her total information. We will reach the same conclusion in Sec. III A, but from a physical point of view. In that section, an even distribution of the Trojan photons will allow Eve to keep the LIDT of an optical component close to its minimum value.

Each of Eve's Trojan pulses is sent in the transmitting unit to probe a different phase value φ_A of Alice's PM. After that, the pulses are retrieved by Eve, and their mean photon number amounts to $\mu_{\text{out}} = \gamma\mu_{\text{in}}$. Let us call f_A the total number of phase values encoded by Alice's PM in 1 second. This is equal to the PM clock rate, expressed in Hz. Because Alice knows f_A, the maximum number of Trojan photons per second N, and the optical isolation γ, she can bound the mean photon number of the Trojan pulses emerging from her module. We call μ_{out} the upper bound. It amounts to

$$\mu_{\text{out}} = \frac{N\gamma}{f_A}. \tag{4}$$

μ_{out} is a crucial parameter in the security argument because it is directly controllable by Alice. It can be interpreted as the mean photon number of the Trojan pulses retrieved by Eve.

In the next section, we proceed from these preliminary observations to derive the secure key rate of the BB84 protocol, assuming that Alice is endowed with an ideal single-photon source. Then, in Sec. II C, we extend the security argument to the BB84 protocol implemented with a laser source and decoy states.

B. Key rate of single-photon BB84 protocol

Let us suppose that Alice prepares ideal single-photon BB84 states and that the only source of information leakage from Alice's system to Eve is from the THA on the PM. Eve shall execute the THA using coherent states of constant intensity as per Eqs. (1) and (3). We assume the worst-case scenario where Eve can retrieve her states back from the quantum channel with 100% fidelity, even though, in practice, this may not be fully permitted by the laws of physics. In this description, the THA can be executed without adding any noise to the communication channel. Despite this, secure keys can still be extracted if the QKD system is well characterized. This is quite counterintuitive as it challenges the common view of QKD as an *eavesdropping detection* system, while promoting it as an *eavesdropping prevention* system [28].

The characterization of the QKD system proceeds as follows. With reference to Alice's interferometer (see Fig. 2), we define the states in the computational basis Z as $|0_Z\rangle \coloneqq |1\rangle_l|0\rangle_s$, $|1_Z\rangle \coloneqq |0\rangle_l|1\rangle_s$, where $|n\rangle_l$ ($|n\rangle_s$) is the n-photon state traveling in the long (short) arm of the interferometer. Then, we write the four BB84 protocol states as $|0_X\rangle$, $|1_X\rangle$ and $|0_Y\rangle$, $|1_Y\rangle$ for the X and Y bases, respectively, corresponding to setting the phase φ_A equal to $\{0, \pi\}$ and $\{\pi/2, 3\pi/2\}$, respectively, in the qubit state $(|0_Z\rangle + e^{i\varphi_A}|1_Z\rangle)/\sqrt{2}$.

Eve's task is to determine φ_A using the light back-reflected from Alice's apparatus. However, the states prepared by Alice and sent to Bob (labeled below with "B") are single photons and do not give any phase reference to Eve. Also, the states sent and retrieved by Eve (labeled below with "E") originate from an external independent source. Therefore, the resulting states emerging from Alice's module can be written as tensor products:

$$\begin{aligned}
|\psi_{0X}\rangle_{BE} &= |0_X\rangle_B \otimes |+\sqrt{\mu_{\text{out}}}\rangle_E,\\
|\psi_{1X}\rangle_{BE} &= |1_X\rangle_B \otimes |-\sqrt{\mu_{\text{out}}}\rangle_E,\\
|\psi_{0Y}\rangle_{BE} &= |1_Y\rangle_B \otimes |+i\sqrt{\mu_{\text{out}}}\rangle_E,\\
|\psi_{1Y}\rangle_{BE} &= |0_Y\rangle_B \otimes |-i\sqrt{\mu_{\text{out}}}\rangle_E.
\end{aligned} \tag{5}$$

The above states justify an alternative interpretation of μ_{out}, i.e., an excess mean photon number exiting Alice's module. If $\mu_{\text{out}} = 0$, only true single-photon states leave the transmitting unit, whereas if $\mu_{\text{out}} > 0$, a hidden side channel, created by the THA, provides Eve with additional information via the excess photons contained in the states of Eq. (5).

It is natural to ask how Eve can use the information obtained in the THA. One possibility is for her to wait until the basis reconciliation step of QKD, in order to measure the back-reflected Trojan photons in the correct basis and learn the bit encoded by Alice. In this case, Eve simply prepares and retrieves the Trojan photons and causes no disturbance on the quantum channel. However, she only gains the information carried by the Trojan photons and makes no use of the photons prepared by Alice. A more powerful strategy is to use the Trojan photons during the

quantum transmission, without waiting for the basis reconciliation step. Eve could first measure the Trojan photons and then decide whether to stop or transmit Alice's qubits conditional on the result from her measurement. Finally, Eve could glean information about the basis chosen by Alice and use it to measure Alice's qubits, thus making optimal use of all the sources of information available to her. This is a convenient framework, as it allows us to prove the security of QKD against the most general attack by Eve [29–31]. We analyze all the above attacking strategies, from the weakest one to the most general. The first and second THA are analyzed in Appendixes C 1 and C 2, respectively, while the third, most general, THA is outlined here and detailed in Appendix B.

To bound the security in the general case, we resort to the so-called "GLLP approach" [29]. More precisely, we use the refinement of GLLP based on the qubit distillation protocol by Koashi [30]. In Appendix B, we apply this approach to the states in Eq. (5) and derive the secure key rate of the efficient BB84 protocol [32,33]. There, it is shown that if the key is distilled from the X basis and the phase error rate is estimated in the Y basis, the asymptotic key rate of a QKD system endowed with a single-photon source is

$$R = Q_X[1 - h(e'_Y) - f_{\mathrm{EC}} h(e_X)]. \tag{6}$$

In Eq. (6), Q_X is the single-photon detection rate in the X basis, i.e., the joint probability that a single-photon pulse is emitted by Alice and detected by Bob and both users measure in the X basis; h is the binary entropy function, f_{EC} is the error correction efficiency [34], and e_X is the (single-photon) quantum-bit error rate (QBER) measured in the X basis. The term e'_Y is the (single-photon) error rate estimated in a virtual protocol, where the users measure in the Y basis and Alice announces the X basis [30]. It is given by the following equations:

$$\begin{aligned} e'_Y &= e_Y + 4\Delta'(1-\Delta')(1-2e_Y) \\ &\quad + 4(1-2\Delta')\sqrt{\Delta'(1-\Delta')e_Y(1-e_Y)}, \\ \Delta' &= \frac{\Delta}{\mathcal{Y}}, \\ \Delta &= \frac{1}{2}[1 - \exp(-\mu_{\mathrm{out}})\cos(\mu_{\mathrm{out}})], \end{aligned} \tag{7}$$

where we conservatively defined $\mathcal{Y} \coloneqq \min[\mathcal{Y}_X, \mathcal{Y}_Y]$, with $\mathcal{Y}_X$ and $\mathcal{Y}_Y$ the single-photon yields in the X and Y bases, respectively, i.e., the conditional probabilities that a single-photon state emitted by Alice causes a click in Bob's detector, when the users measure in the same basis, X or Y. The presence of μ_{out} in the last line of Eq. (7) shows how the THA affects the key rate in Eq. (6).

The key rate R has been plotted in Fig. 3 as a function of the distance between the users, for different values of the output mean photon number μ_{out}, using parameters close to existing real systems [35]. From the figure, it can be seen that the key rate corresponding to $\mu_{\mathrm{out}} = 10^{-6}$ is indistinguishable from $\mu_{\mathrm{out}} = 0$ (no THA) over distances up to 100 km, i.e., about 60% of the maximum distance (170 km in Fig. 3). For $\mu_{\mathrm{out}} = 10^{-8}$, the key rates in the presence and absence of a THA overlap over nearly the whole range. In this case, a negligible amount of additional privacy amplification is required to guarantee security against the THA. The key rate remains positive also for $\mu_{\mathrm{out}} = 10^{-2}$, but the maximum distance is limited to 9 km in this case and the key rate is severely affected by the THA. The largest value of μ_{out} showing a positive key rate is 0.015.

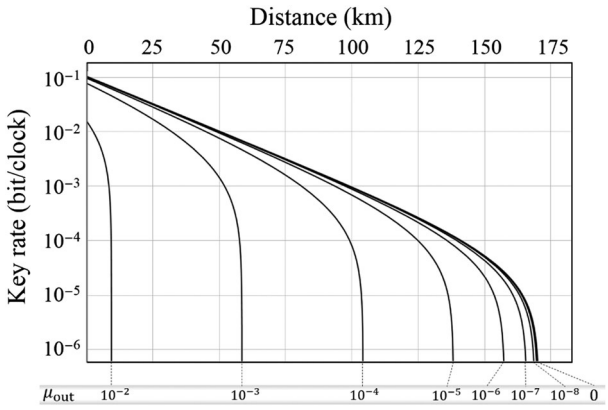

FIG. 3. Asymptotic key rate R versus distance for the single-photon efficient BB84 protocol under a THA. The rate is plotted for different values of the output mean photon number μ_{out}. Other parameters in the simulation are as follows: fiber loss coefficient 0.2 dB/km, total detection efficiency 12.5%, optical error rate 1%, dark count probability per gate 10^{-5}, and error correction inefficiency 20% above the Shannon limit.

It is worth remarking that the entire effect of the THA is condensed in the parameter μ_{out}, as it is apparent from Eq. (7). Therefore, the obtained key rate is equally applicable to any QKD setup capable of guaranteeing an upper bound to the mean number of Trojan photons reflected by the transmitter back to Eve.

C. Key rate of decoy-state BB84 protocol

The key rate in Eq. (6) has been derived assuming that a single-photon source is available to Alice. However, it is well known that security can still be guaranteed without a single-photon source if a phase-randomized attenuated laser [36] is combined with the decoy-state technique [37,38]. Actually, such a solution is currently more efficient than a single-photon source because of the limited generation rates of existing single-photon sources [39,40].

To extend the result to a decoy-state source, we assume that the decoy-state execution is not affected by the THA. This is equivalent to saying that Eve's only target in the THA considered here is Alice's PM, and the devices used by Alice to implement the decoy-state technique are not

touched by the THA (see assumption 3 in Appendix A and the accompanying discussion). Under this assumption, the decoy-state key rate is a straightforward generalization of Eq. (B5) along the lines described, e.g., in Ref. [38]. Indicating with a tilde the quantities to be estimated via the decoy-state technique and with s the mean photon number of the signal pulse in the decoy set of states, we obtain

$$\tilde{R} = \tilde{Q}_X^{(1)}\{1 - h[\tilde{e}_Y^{\prime(1)}]\} - Q_X^{(s)} f_{\mathrm{EC}} h[e_X^{(s)}], \tag{8}$$

where

$$\begin{aligned} \tilde{e}_Y^{\prime(1)} &= \tilde{e}_Y + 4\tilde{\Delta}'(1-\tilde{\Delta}')(1-2\tilde{e}_Y) \\ &\quad + 4(1-2\tilde{\Delta}')\sqrt{\tilde{\Delta}'(1-\tilde{\Delta}')\tilde{e}_Y(1-\tilde{e}_Y)}, \\ \tilde{\Delta}' &= \frac{\Delta}{\tilde{\mathcal{Y}}}. \end{aligned} \tag{9}$$

In Eq. (8), $\tilde{Q}_X^{(1)}$ is the decoy-state estimation of the single-photon detection rate Q_X in Eq. (6), while $Q_X^{(s)}$ is the detection rate of the signal pulse measured in the X basis. In Eq. (9), we conservatively defined $\tilde{\mathcal{Y}} \coloneqq \min[\tilde{\mathcal{Y}}_X, \tilde{\mathcal{Y}}_Y]$, with $\tilde{\mathcal{Y}}_X$ and $\tilde{\mathcal{Y}}_Y$ the single-photon yields in the X and Y bases, respectively, estimated via the decoy-state technique.

The key rate $\tilde{R}$ is plotted in Fig. 4. Although rate and maximum distance are smaller than in the single-photon case (Fig. 3), as expected, it is remarkable that the key rate corresponding to a value $\mu_{\mathrm{out}} = 10^{-7}$ remains indistinguishable from the ideal rate ($\mu_{\mathrm{out}} = 0$) over nearly the whole distance range. A 10 times larger value, $\mu_{\mathrm{out}} = 10^{-6}$, which is easier to achieve in practice, generates a key rate that closely follows the ideal one up to 100 km, i.e., 70% of the maximum distance achievable (146 km in Fig. 4), and remains positive up to 140 km, i.e., 96% of the maximum distance. This motivates our choice of $\mu_{\mathrm{out}} = 10^{-6}$ for the case study in Sec. IV A. Finally, it is worth noting that the key rate remains positive even for larger values of μ_{out}, up to 0.012.

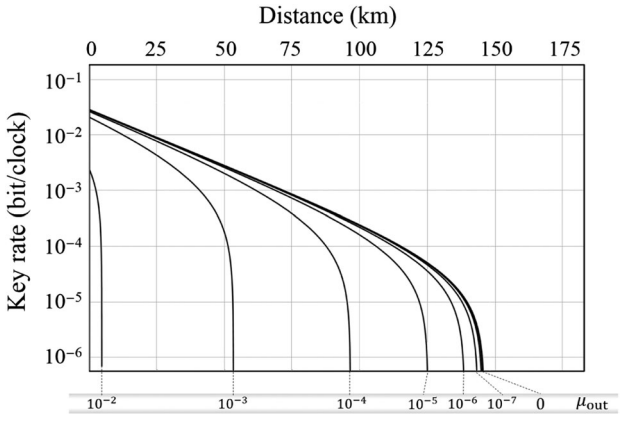

FIG. 4. Asymptotic key rate $\tilde{R}$ versus distance for the decoy-state efficient BB84 protocol under a THA, for various values of the output mean photon number μ_{out}. Experimental parameters in the simulation are as in Fig. 3. The average photon number of the signal states in the decoy-state technique is $s = 0.5$.

Before concluding this section, a couple of remarks are in order. First, it has been convincingly proven that the key rate achieved with only three decoy states is very close to that obtained with an infinite amount of decoy states [41]. We have run simulations that confirm this result. Therefore, the key rate in Fig. 4 is achievable in a real system. Second, the rate equations provided so far have been derived using coherent states of constant intensity. Here, we show that this setting is actually advantageous to Eve. Suppose that Eq. (3) does not hold, i.e., $\mu_i \neq \mu_{\mathrm{in}}$. With this setting, Eve is trying to distribute her N Trojan photons unevenly among the M pulses, in an attempt to enhance her information gain. Suppose that Eve distributes the N photons in only two classes of pulses, such that the first (second) class features an average photon number μ_1 (μ_2) and $\mu_1 < \mu_2$, $\mu_{\mathrm{in}} = (\mu_1 + \mu_2)/2$. Then, for each of the key rates given in this work, represented by a generic symbol $\mathcal{R}$, we have numerically verified that

$$\mathcal{R}(\mu_{\mathrm{in}}) \leq \frac{\mathcal{R}(\mu_1) + \mathcal{R}(\mu_2)}{2}. \tag{10}$$

We have used the explicit expressions of the key rates and their dependence on μ_{out}, which is related to the input photon number by the linear equation $\mu_{\mathrm{out}} = \gamma\mu_{\mathrm{in}}$. In other words, the key rates are convex functions of μ_{out} and thence of μ_{in}. According to Eq. (10), the rate distilled by the users under Eve's new strategy (right-hand side) is larger than the one pertaining to the old strategy (left-hand side), so the new strategy is less effective and not advantageous to Eve. More general strategies by Eve that account for more than two classes of photons with different mean photon numbers can be treated as a trivial extension of Eq. (10).

III. BOUNDS ON INPUT PHOTONS

In our security argument, the quantity N plays an important role. Here, we provide more details about this limiting threshold and describe a way to quantify it. We adopt a pragmatic approach, motivated by the results of the previous sections. In particular, we have established that Eve conducts the THA using coherent states of constant intensity. Therefore, we can conveniently think that such states are generated by a single-mode laser operated well above threshold [42]. This view naturally leads to considering the laser-induced damage threshold (LIDT). From a security perspective, the LIDT can only provide a general indication of the bounds to be used in a security analysis. The actual response to thermal damage of the real components of a QKD system should be experimentally measured.

A. Laser-induced damage threshold

A single-mode optical fiber is arguably the most common component of a fiber-based QKD setup. It is used mainly to transmit information in the third telecom window (wavelength $\lambda = 1.55$ μm) because of its small attenuation coefficient. Its typical core diameter is 8–10 μm, corresponding to a core area of 50–80 μm^2. If the laser power used by Eve is sufficiently high, it creates an accumulation of energy in this small region of the core and increases the temperature of the medium beyond the tolerance level, inducing fiber thermal damage [43,44].

Such a damage threshold is usually quantified by the LIDT, defined in the 2011 international standard ISO 21254-1 as follows [45]: "the highest quantity of laser radiation incident upon the optical component for which the extrapolated probability of damage is zero, where the quantity of laser radiation may be expressed in energy density, power density or linear power density" [46]. The smaller the LIDT of the component, the larger the probability of damaging it. This subject is well studied, and values for the LIDT of a silica-based optical fiber, which is the component we are interested in, can be obtained. However, before discussing the absolute values, it is worth examining the qualitative behavior of the LIDT, which is determined by the underlying thermal damaging mechanism. The purpose is to investigate how features of Eve's laser, like the repetition rate or the pulse width, can affect the LIDT and, as a consequence, the system security. This provides useful indications for setting a proper LIDT value.

One prominent feature of the LIDT is that it increases with the pulse width of the incident laser; i.e., a wide light pulse causes less damage to the optical component than a narrow one. This result makes narrow pulses more detectable to Alice and Bob than wide ones. This can be formalized using the well-known square-root dependence of the LIDT on the pulse width [47–50]:

$$\frac{\mathrm{LIDT}(\tau_1)}{\mathrm{LIDT}(\tau_2)} = \sqrt{\frac{\tau_1}{\tau_2}}. \tag{11}$$

Here, τ_1 and τ_2 are two different pulse widths for the same pulse energy. Equation (11) suggests that Eve's laser pulse should be the widest possible, compatible with Alice's phase modulator.

A similar rule applies to the laser wavelength, resulting in the shorter wavelength causing more damage to the optical component than the longer one (see, e.g., Ref. [51]):

$$\frac{\mathrm{LIDT}(\lambda_1)}{\mathrm{LIDT}(\lambda_2)} = \sqrt{\frac{\lambda_1}{\lambda_2}}. \tag{12}$$

Equation (12) suggests that Eve's optimal laser's wavelength should be as large as possible, even larger, if necessary, than the typical wavelength used in the QKD setup. However, it also entails that the LIDT remains reasonably constant for all the wavelengths possibly transmitted in the fiber. A standard optical fiber cannot transmit by total internal reflection beyond the so-called "bend-edge" wavelength, which is only a few hundred of nanometers away from the fiber cutoff wavelength (see, e.g., Ref. [52]). As an example, we can consider a bend-edge wavelength of 1850 nm for an optical fiber transmitting at 1550 nm [53,54]. According to Eq. (12), this would increase the LIDT by less than 10%, showing that the wavelength of Eve's laser is not crucial in determining the efficacy of the THA. To compensate for this effect in the theory, it suffices to increase the LIDT value by 10%.

To upper bound the input photon number N used in the security argument, we need to estimate the LIDT of Alice's optical module. This is arguably given by the LIDT of the most fragile component in the module. However, we consider the LIDT of just one of the components in Alice's unit, the one most exposed to Eve's light. The other components are assumed to either work in their normal operation regime or fail in a way that is detectable by the users (see assumption 1 in Appendix A and the accompanying discussion). In Sec. IV, we describe the architecture of Alice's setup against the THA. The component most exposed to Eve's light is a loop of standard optical fiber placed at the main entrance of the transmitting box. Hence, we are interested in the LIDT of a standard single mode optical fiber. One possible way to estimate it is to consider the geometry of the fiber and the material it is made of. As already mentioned, a typical fiber has a core area of about 50 μm^2 and is made of fused silica. The LIDT of fused silica is determined by the softening point of the material [47] and amounts to 1.1×10^7 J/cm^2 [55]. For a longer time, the silica-based medium starts dissipating heat and the threshold increases linearly with the pulse width. For a shorter time, the square root law in Eq. (11) applies, decreasing the LIDT accordingly.

The above-cited LIDT value corresponds to an average power of 5.5×10^4 W over 50 μm^2. For a typical wavelength of $\lambda = 1.55$ μm, this means that 4.3×10^{23} photons impinge every second onto the fiber core area. Before such a large number of photons can damage the fiber core, other highly detectable damages are likely to occur at the fiber interfaces, causing, e.g., a net reduction of the transmission or an increase in the noise figure. Also, the LIDT value mentioned above relates to a homogeneous medium. In reality, large temperature gradients can occur in the proximity of a defect, or at the connection between two segments of fiber, or at the interface between the fiber core and the cladding. Some of these properties can even be artificially enhanced by acting on the number of connectors, the bending radius, and the doping levels of the fiber. These considerations lead to the conclusion that the given LIDT value is an overly conservative estimation of the real LIDT of an optical fiber. In the next section, we obtain a

different LIDT value by combining the findings of Sec. II with the results from experiments performed on real optical fibers.

B. Fiber thermal fuse-induced LIDT

In Sec. II, we have shown, from an information theory point of view, that Eve's optimal strategy is to distribute her photons into a number of pulses M that is equal to Alice's PM clock rate (in Hz) f_A, so as to maximize her total information gain. In the description of Eve's laser, the above strategy translates into setting $f_E = f_A$, where f_E is Eve's laser repetition rate. Moreover, in the previous section, we have shown that the LIDT depends only weakly on the laser pulse width and that the larger the width, the larger the damaging threshold. In the description of Eve's laser, this translates into having a laser pulse width τ_E as large as possible, compatible with Alice's PM. Let us call τ_A the time window of Alice's PM. If $\tau_E > \tau_A$, a fraction of Eve's photons fall outside the PM gate and deliver no information to Eve. Therefore, the optimality condition for Eve is $\tau_E = \tau_A$. This condition on the pulse width represents an additional constraint for Eve and an extra parameter under Alice's control. After γ and f_A, Alice can now act on τ_A to make Eve's strategy less effective. In particular, by reducing τ_A, Alice reduces the damaging threshold of her module, hence N.

Let us draw a worst-case scenario from the above considerations. We conservatively assume that Alice's PM is driven by a perfectly rectangular wave. This assumption helps Eve match the condition $\tau_E = \tau_A$ and simultaneously keep the damaging threshold high. As a consequence, the amplitude of Alice's PM is assumed to be flat in time. The amplitude is selected at random among the four equally spaced values of the BB84 protocol. The way these values are selected depends on the logic driving the PM. If a non-return-to-zero (NRZ) logic is used, the PM duty cycle is 100%; i.e., the PM is always active, transiting from a given phase value directly to the next one, and we have, in this case, $\tau_A = 1/f_A$. If a return-to-zero (RZ) logic is used, the modulator is reset after each encoded phase value. In this case, the duty cycle is less than 100% and the PM time duration is $\tau_A < 1/f_A$. We note that in the particular case where Alice's PM is driven according to a NRZ logic (100% duty cycle), Eve's laser coincides with a continuous-wave (CW) laser, as it emits a seamless sequence of rectangle pulses, all of the same amplitude, sitting one next to each other. A deeper thought reveals that this is actually a worst-case scenario because, when the condition $\tau_E = \tau_A$ is matched, τ_E takes on its maximum value ($1/f_A$), thus minimizing the risk of optical damage, while leaving Eve's information unchanged. Therefore, we can always imagine that Alice's PM is driven by a NRZ logic, even if it is RZ and, accordingly, Eve uses a CW laser to probe the PM.

Experiments performed with CW lasers on real optical fibers have demonstrated that an average power around 2–5 W causes catastrophic thermal damage in a standard single-mode silica fiber [56–58]. This effect is known as "self-propelled self-focusing" or "fiber thermal fuse" [56,57,59–64]. The high power of the laser generates a heating point in the fiber where the local temperature overcomes the melting point of the medium. From there, the damage propagates along the fiber, eventually making it unusable. This effect has also been exploited to build an "optical fuse" that breaks by 1.2–5.3 W of incident light at wavelengths around 1500 nm [62]. For a wavelength of $\lambda = 1550$ nm, 2 W correspond to 1.6×10^{19} photons crossing a 50-μm^2-fiber core area (a_{50}) every second. In order to have an easy reference for the LIDT value, we set it equal to $N = 10^{20}$ photons/s/a_{50}. The new LIDT value is 4.3×10^3 smaller than the previous one. Still, it corresponds to 12.8 W from a CW laser, which is much larger than the power threshold reported in the fiber thermal fuse experiments. We adopt this number to draw an example where the values of the optical components in Alice's apparatus are connected to the security requirements. However, the more conservative threshold for N could be adopted instead to arrange a different use for an application that requires a stronger bound, independent of the fabrication details of the fiber and relying only on the softening point of silica.

IV. EXPERIMENTAL CHARACTERIZATION

A. Passive architecture against the THA

An entirely passive architecture against the THA is drawn schematically in Fig. 5. It is based on a sequence of components that actualize the security argument described so far. A silica-based optical fiber loop (OFL) of length L defines the LIDT of the transmitter and is followed by a filtering block F, an optical isolator I, and an attenuator A. We also indicate with R the total reflectivity of the optical elements to the left of the dot-dashed line [not to be confused with the key rate R given in Eq. (6) and plotted in Fig. 3]. The line for the reflectivity R is conservatively drawn to also include the first beam splitter as seen by Eve to allow an easier experimental implementation

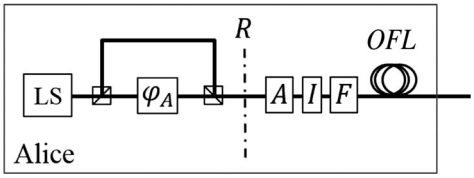

FIG. 5. Architecture of a QKD transmitter to mitigate the THA. LS is a generic light source and the square with φ_A is the encoding device. OFL: optical fiber loop determining the LIDT; F: optical filter; I: optical isolator; A: attenuator; R: total reflection from all components to the left of the dot-dashed line.

(Sec. IV B). In the figure, all the components are presumed to either work as expected or fail in a way that is detectable by the users.

The OFL acts as a regulator for high-power input light and as a filter for wavelengths longer than the bend-edge point. Together with the optical filter F, which is tuned to let the wavelength of the quantum channel pass and stop all the others, it limits the maximum number N of photons that Eve can inject into Alice's module in the chosen time unit. In other words, it represents the optical component to which the LIDT should apply. To fix ideas, we can imagine a length for the OFL greater than 1 m, in line with what was reported in the experiments about the fiber thermal fuse effect. A longer OFL can only be beneficial to the users, as it increases the probability of thermal damage, which increases as well if a few interfaces are present in the OFL.

The optical isolator strongly attenuates the input light from Eve, enforcing the unidirectionality condition in the module. A typical dual-stage optical isolator features an isolation value of 10^{-5} or smaller. It is convenient to measure the isolation in decibels, or dB, rather than in absolute value. If x is the absolute value of the optical isolation of a given component, we use the following notation,

$$\dot{x} = 10\log_{10} x, \tag{13}$$

to indicate its value in decibels. For example, the optical isolator mentioned above would feature an isolation of −50 dB.

The attenuator box in Fig. 5 is already present in the schematics of various QKD systems using an attenuated laser as a light source, while it is not present in systems using a single-photon source, as it would entail major losses in the system. If used, it helps to avert the THA, as it contributes to the optical isolation of Alice's module, γ. The following equation quantifies the contributions of each single conceptual block in Fig. 5 to γ:

$$\gamma = F^2 \times I^n \times A^2 \times R. \tag{14}$$

Equation (14) can be conveniently rewritten in dB:

$$\dot{\gamma} = 2\dot{F} + n\dot{I} + 2\dot{A} + \dot{R}. \tag{15}$$

In Eqs. (14) and (15), the typical double pass of a THA through Alice's components has been considered, which leads to explicit corrections for the filter and the attenuator terms. For the isolator term, there is no such correction because one direction of the double pass features zero attenuation. However, there is a factor n that represents the number of optical isolators present in the system.

To relate the isolation γ to the system security, we need to connect it to the parameter μ_{out} via Eq. (4). Therefore, we introduce the dimensionless ratio $\chi := N/f_A$ and rewrite Eq. (4) in dB notation:

$$\dot{\mu}_{\text{out}} = \dot{\chi} + \dot{\gamma}. \tag{16}$$

To give an example of how Eqs. (15) and (16) can be used to meet the security criterion, let us start by setting a target value for the excess average photon number μ_{out}. We have seen from Figs. 3 and 4 that a value $\mu_{\text{out}} = 10^{-6}$ (i.e., $\dot{\mu}_{\text{out}} = -60$ dB) can guarantee security against the THA with only a negligible (limited) amount of additional privacy amplification over short-range and middle-range (long-range) QKD transmissions. Therefore, we choose this value as the target. We consider the threshold value $N = 10^{20}$ photons/s/a_{50} discussed in Sec. III B and a system clock rate $f_A = 10^9$ Hz. These values give $\chi = 10^{11}$ ($\dot{\chi} = 110$ dB). From Eq. (16), we then get $\dot{\gamma} = \dot{\mu}_{\text{out}} - \dot{\chi} = (-60\text{–}110)$ dB $= -170$ dB. This result is the total optical isolation required in Alice's module in order to guarantee security. Alice can try and match this value by using well-characterized components and then applying Eq. (15).

Table I contains some possible combinations of f_A, $\dot{R}$, $\dot{A}$, and $\dot{I}$ to match the target value $\dot{\mu}_{\text{out}} = -60$ dB. For convenience, we report the absolute values of the components. In the table, we set $\dot{F} = 0$ because the filter insertion loss is typically close to zero at its central wavelength, and we assume that the filter is centered at the operational wavelength of the QKD setup. In the first column, we have considered three interesting and feasible regimes, 1 kHz, 1 MHz, and 1 GHz. The lines with the asterisk are for situations where attenuation cannot be used, e.g., if the transmitter uses a single-photon source or at the receiver side. It is worth noting that single-photon sources up to the MHz range are currently available (see, e.g., Refs. [39,40]). In all cases, we have reported what we believe to be the most practical combination of components. For the optical reflectivity $\dot{R}$, we have considered a typical absolute value of 40 dB, which comes from a common fiber connector.

TABLE I. Practical combinations of system components to meet the target $\mu_{\text{out}} = 10^{-6}$ when $N = 10^{20}$ photons/s/a_{50} and $\dot{F} = 0$ dB. All dotted quantities are in decibels and are given in absolute value. Lines with the asterisk are cases in which attenuation cannot be used, e.g., when the transmitter uses a single-photon source or at the receiver side. The feasibility of the values for 1-GHz clock rate has been confirmed experimentally using the QKD setup described in Ref. [65].

Clock rate	f_A (Hz)	$\lvert\dot{\gamma}\rvert$	$\lvert\dot{R}\rvert$	$\lvert\dot{A}\rvert$	$\lvert\dot{I}\rvert(n)$
1 GHz	10^9	170	40	35	60(1)
1 GHz*	10^9	170	50	0	60(2)
1 MHz	10^6	200	40	30	50(2)
1 MHz*	10^6	200	50	0	50(3)
1 kHz	10^3	230	40	35	60(2)
1 kHz*	10^3	230	50	0	60(3)

However, an absolute value of 50 dB is possible if angled connectors or splicing are used for the fiber-integrated optics in the module. This latter option is worth considering, especially for the lines with the asterisks in Table I. For the optical isolator, its absolute value is set in the factory and cannot be varied by the users. We set it equal to either 50 dB or 60 dB in Table I, according to the most convenient configuration. The former value is common in dual-stage optical isolators. The latter value is less common, but it can be obtained by properly sampling a set of isolators and selecting the best one (see Sec. IV B). Finally, for the attenuator, we avoided using absolute values larger than 35 dB, as that would commit the transmitter to unusually high-power lasers.

From Table I, it can be seen that two or more optical isolators might be necessary to meet the security target $\mu_{\text{out}} = 10^{-6}$. However, if the clock rate is high enough, a single isolator is sufficient (first line of the table). In any case, given the low cost and the low insertion loss of filters, attenuators, and optical isolators, all the options in Table I can be considered feasible and relatively inexpensive.

B. Components characterization

To prove the attainability of the values reported in the first line of Table I, we have experimentally characterized reflectivity and transmission of the components most relevant to security in the transmitting unit of a unidirectional GHz-clocked QKD system [65], within their operational range. A full-range characterization of the real components in the setup is necessary to guarantee their behavior against unwanted deviations, as required by the security argument (see assumption 1 in Appendix A).

As a first step, we have used single-photon optical time-domain reflectometry (ν-OTDR, Ref. [66]) to quantify the reflectivity R of Alice's apparatus. The measurement setup and the resulting traces are shown in Fig. 6, on the top and bottom diagrams, respectively. In the ν-OTDR setup, a 1-MHz pulsed laser at 1550 nm is connected to Alice via a circulator. Polarization controllers are used to align the pulses to the long or short path of Alice's interferometer to obtain the output patterns of the orthogonal polarizations. These are shown as blue and red traces in Fig. 6. The two patterns have been added together to upper bound the total reflectivity, and this is indicated by the black trace in the figure. The upper bound to R is obtained assuming the linearity of the reflectivity, as follows: $R(a|s\rangle + b|l\rangle) = aR(|s\rangle) + bR(|l\rangle) \leq R(|s\rangle) + R(|l\rangle)$, where the vector $|s\rangle$ ($|l\rangle$) represents the polarization traveling in the short (long) arm and a, b are complex numbers with modulo squared adding to 1. The traces are plotted from the entering point of Alice's module, which is connector J1 in Fig. 6. However, only the peaks pertaining to the components included in the shaded region of the top diagram have to be considered in the estimation of R (see also dot-dashed line in Fig. 5).

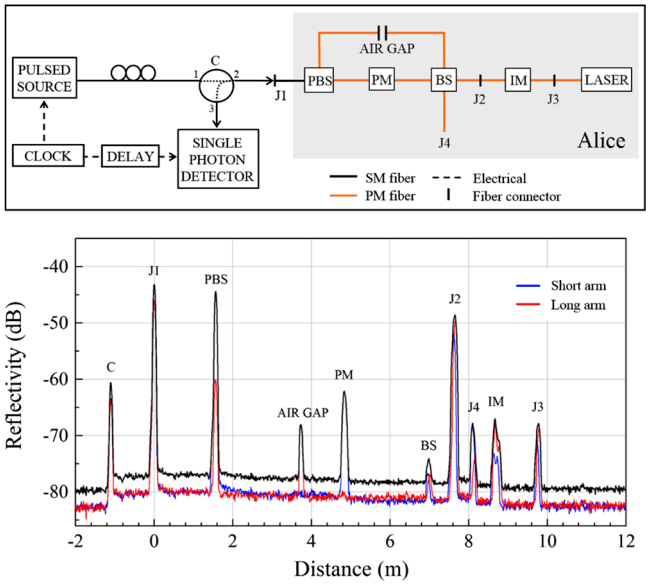

FIG. 6. Top panel: Schematics of the QKD transmitter module and of the ν-OTDR setup used for characterizing its reflectivity. Bottom panel: Reflection peaks of the transmitting unit. The distance is measured from the connector J1 placed at the entrance of the module. The traces are acquired for two orthogonal polarizations, aligned to maximize the transmission through the short (blue traces) and long (red traces) arms of the interferometer. The peaks of the reflectivity are added to obtain a worst-case estimation (black). Only the peaks from the components included in the shaded region of the top diagram have to be considered in the estimation of R.

The sum of all the peaks relevant to R gives a total reflectivity of -42.87 dB. This value meets the requirement $\dot{R} < -40$ dB set in the first line of Table I. Also, the characterized QKD system includes an attenuator set to -35 dB. To match the $|\dot{\gamma}| = 170$ dB condition, additional optical isolation of at least -60 dB is needed. Dual-stage isolators specifying typical isolation at this level are commercially available from a number of manufacturers. For demonstrative purposes, we tested isolators M-IS/M-II from FOCI Fiber Optic Communications, Inc. One of the isolators featured an absolute isolation larger than 65 dB in the proximity of the main transmission wavelength of the system, 1550 nm, as shown in Fig. 7. Across the S, C, and L bands, the isolation value varies, until it reaches a minimum of about 40 dB. However, in this regime, the optical filter takes over and provides high optical isolation so that a typical suppression of more than 80 dB is obtainable across the entire C band. Because the filter is crossed twice by Eve's light, this leads to more than 160 dB additional optical isolation to the system whenever the wavelength is different from 1550 nm. This result demonstrates that the values reported in the first line of Table I are feasible when devices are operated in their working regime.

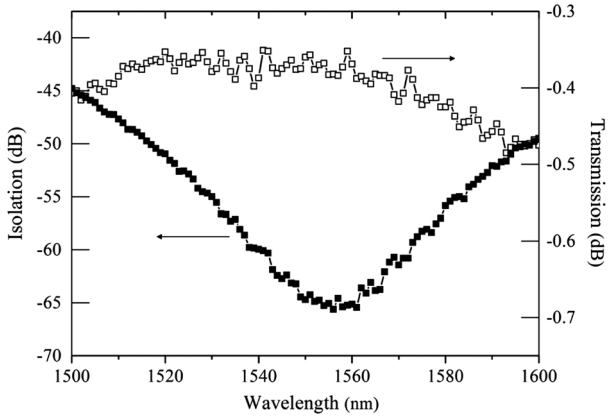

FIG. 7. Spectral characterization of a dual-stage optical isolator. The isolator shows less than 0.36 dB insertion loss in the forward direction (right axis, empty squares) and more than 65 dB isolation in the backward direction (left axis, filled squares) around the central wavelength of 1550 nm.

V. DISCUSSION

In the first part of the work, we derived our main result, i.e., the secure key rate of a QKD system in the presence of a THA, under reasonable assumptions (see Appendix A for a summary and a discussion of the assumptions). The result depends on Alice's ability to limit the number of the incoming photons N and to reliably upper bound the mean number of Trojan photons μ_{out} exiting from her module. However, the curves plotted in Figs. 3 and 4 are independent of N and can be applied to different QKD systems, provided that the assumptions in the theory are met. From the key rates, we have shown that a value of the mean output photon number $\mu_{\text{out}} \sim 10^{-6}$ allows one to approach the situation with no THA for nearly any distance between the users. For distances up to 70% of the maximum working distance, this can be achieved without any additional privacy amplification.

In Sec. III, we have drawn an example of how to set a value on N using the thermal damage point of an optical fiber. The most conservative value for N is on the order of 10^{23} photons injected every second on the core area of the fiber. This value has been obtained from the softening point of a homogeneous medium made of fused silica and is independent of future advancements in technology, provided that the composition of fused silica remains unchanged. A lower value for N, equal to 10^{20} photons/s/a_{50}, has been drawn from recent experiments on the thermal damage of real fibers, after taking into account the presence of inhomogeneities in the fiber and the qualitative behavior of the LIDT in response to the laser pulse width. Using this lower value, we showed the feasibility of our passive architecture in a practical scenario. We related the key rate of a QKD system to its clock, detection rate, and reflectivity and to the properties of a sequence of fiber loop, filter, and optical isolator, as depicted in Fig. 5. In Table I, we devised various combinations of these components to meet the security condition against the THA. According to the table, most of the existing QKD systems can potentially be protected from THA's, provided that a sufficient number of optical isolators are used and that the real components behave as expected.

Some elements in our security argument may appear optimistic: for instance, the use of coherent states by Eve. However, we believe that overall our analysis is conservative. The considered LIDT threshold corresponds to a light power of 12.8 W from a CW laser and is larger than the power required to activate the fiber thermal fuse effect in a standard single-mode fiber. It is reasonable to think that before this large number of photons can melt the fiber core, some other mechanism would make Eve detectable. We also assumed a noise-free retrieval of quantum states by Eve, while it is well known that the retrieval is physically limited by Raman and Rayleigh scattering. We decided not to consider the fact that other QKD components, already present in Alice's module, could have a lower LIDT than the optical fiber. Finally, we ignored the fact that monitoring detectors are already present in most of the QKD systems, mainly for stabilization purposes. Such devices additionally constrain Eve's action and can be beneficial to improve our solution.

VI. CONCLUSION

In this work, we studied the security of a fiber-based QKD setup endowed with passive optical components against the long-standing Trojan-horse attack. In the framework of Ref. [29], we provided quantitative security bounds, easily applicable in practice, against a general THA. With the proof method of Ref. [67], we analyzed two specific examples of a THA, giving useful insights into the THA mechanism and the method employed to prove security against it (Appendix C). In our analysis, we focused on a particular unidirectional QKD setup, in which light flows from the transmitter to the receiver and the reverse direction is forbidden. This architecture is similar to that of the transmitters used in Ref. [10] to guarantee the measurement-device-independent security of the decoy-state BB84 protocol. Hence, we expect that our results can be applied to that system after minor modifications. The unidirectional configuration allows the use of optical isolators, whose proper behavior has to be tested against undesired deviations. The resulting protection measure against the THA is entirely passive, thus preventing the loop holes inherent to active, more sophisticated countermeasures. We believe it will become a standard tool in all quantum-secured optical systems that need to guarantee the protection of a private space.

ACKNOWLEDGMENTS

We are indebted to Norbert Lütkenhaus for his scientific support throughout the work preparation and to

Kiyoshi Tamaki for noticing an ill-posed generalization of Ref. [67] during his critical reading of the manuscript. Useful discussions with Bernd Fröhlich are gratefully acknowledged.

APPENDIX A: SECURITY-RELATED ASSUMPTIONS

In the main text, we considered the unidirectional, fiber-based, phase-modulated QKD setup depicted in Fig. 2 and studied its resistance to the THA. We have made a number of assumptions that we summarize in the following list:

(1) Alice has the ability to bound N, the number of Trojan photons entering her setup. She can characterize the components in her setup and test whether they behave as expected under all relevant conditions.
(2) Eve uses a tensor product of coherent states to execute the THA. The intensity of the coherent states does not need to be constant, but it is advantageous for Eve to choose it to be constant.
(3) Alice's light source emits either single-photon states or phase-randomized coherent states that are perfectly encoded into the states of the BB84 protocol. Imperfect encoding of the initial states as studied, e.g., in Ref. [67] is excluded. The only side channel in the QKD setup is the THA against Alice's phase modulator described in the main body of this work.
(4) The detection efficiency of the receiver is independent of the basis choice, and the basis is randomly chosen by the users.
(5) The key rate is worked out in the asymptotic scenario, assuming that Alice and Bob have infinitely many signals and decoy states to generate the key.
(6) The reflectivity measured via the OTDR experiment is a linear function of the input polarization.

Without assumption 1, it would be impossible to prove security. If the quantity N cannot be bounded, there is no private space for the encoding of the classical information onto the quantum systems, and the quantum protection is circumvented. In the main text, N was bounded using the LIDT of the OFL in Fig. 5. It is natural to ask whether a power monitor or a watchdog detector, placed at the entrance of Alice's unit to actively monitor the input power, can provide an alternative, better, bound to N. There are reasons this option might not work. First, an additional detector would add extra cost and complexity to the setup, opening up additional potential loop holes. For example, it has been shown in Ref. [15] that a power monitor can be easily bypassed if not properly engineered. Second, we used Eq. (11) for the LIDT, according to which narrow pulses of light create larger damage to the optical component than continuous-wave light, so they are more easily detectable by the users. This result let us draw a worst-case scenario for Eve's laser. We are not aware of a similar law applicable to a power monitor. Finally, even if the power monitor solution worked fine and allowed us to reduce the input photon number N by several orders of magnitude, it should still be compared to the 6 or more orders of magnitude guaranteed by the addition of a single inexpensive and nearly loss-free component like an optical isolator.

As for the second part of assumption 1, if Alice cannot characterize her components, she cannot work out the value of the optical isolation γ to relate N and μ_{out} via Eq. (4). The characterization should consider the physical limits imposed on Eve's laser. For example, Eve's laser's power is constrained by the LIDT of the OFL in Fig. 5. Therefore, the behavior of the components should be tested up to the LIDT value of the OFL. This excludes hacking strategies leveraging on an unexpected behavior of the real components, passively or actively triggered by the eavesdropper. The characterization step could be simplified if an optical fuse with a LIDT value lower than the lowest tolerance threshold of the components in Alice's setup were available [68].

Assumption 2 allows us to write the states leaving Alice's apparatus as in Eq. (5). It is possible, in principle, that phase-sensitive states of light, e.g., squeezed states [42], could provide Eve with more information than coherent states. However, as Table I shows, the value of the attenuation in Alice's setup is at least 170 dB. It seems unlikely that the fragile squeezed state can survive in this lossy environment. The second part of this assumption descends from the convexity of the secure key rate as a function of μ_{out}, which has been verified for all the key rates presented in this work.

Assumption 3 is necessary to remove additional side channels that could, in principle, enhance the THA, e.g., encoding states that are different from the ideal ones prescribed by the BB84 protocol. Also, it guarantees that the rate equations derived for the decoy-state BB84 protocol hold because Eve's tampering with Alice's decoy state estimation would represent an additional side channel and would contradict the assumption. Extending the security argument to decoy states without making use of assumption 3 could be a trivial task, and a separate detailed study is required. However, we would like to speculate on this point further.

For simplicity, we assume that the light emitted by Alice is phase randomized. In some cases, this is simple to guarantee, e.g., when phase randomization is an intrinsic feature of the light source [36]. In other cases, when phase randomization is committed to a separate active component [21], it could be more difficult to show that Eve cannot access this extra component with a more refined THA. With phase randomization on hand, the decoy-state technique requires that the intensity of the emitted light is varied in a random way, known to Alice. This can be achieved by adding an intensity modulator (IM) to the setup of Fig. 2,

between the interferometer and the laser source. If there is no additional optical isolation between the IM and the interferometer, the optical isolation γ that shields Alice's PM from Eve applies to the IM too, and the coherent state sent by Eve to probe the IM returns to her with an average photon number not larger than μ_{out}. However, if there is a perfect optical isolator between the IM and the interferometer, then the Trojan photons retrieved by Eve are only informative about Alice's PM, whereas the IM is perfectly shielded from Eve. This latter case is an example of how assumption 3 can be enforced. However, because perfect isolation is impossible in practice, we considered how the key rates of the decoy-state BB84 would change if a *single real* optical isolator, guaranteeing 50 dB isolation, were used instead. In this case, the μ_{out} back-reflected to Eve from the IM would be 5 orders of magnitude smaller than the one back-reflected from the PM. Applying to this realistic scenario an argument similar to the one described in the forthcoming Appendix C 2, we found key rates that are indistinguishable from the ones presented in this work.

Assumptions 4 and 5 are related to the proof methods adopted by us to draw the key rates in the presence of a THA [29–31,67]. There, security was proven in the asymptotic scenario leveraging on the fact that the measurement performed by the receiver is equivalent to a basis-independent filter followed by a two-valued positive-operator valued measure (POVM). In Ref. [69], it was shown that this assumption can be enforced if Bob's single-photon detectors have equal efficiency and if their dark counts and efficiencies are carefully modeled. The detectors can be threshold detectors, and in this case, a specific value of the key bit must be assigned whenever both detectors click to guarantee the basis-independence condition.

Assumption 6 is necessary during the characterization stage to upper bound the reflectivity of the transmitter, as shown in Sec. IV A. To meet this assumption, we put particular care into the OTDR experiment to avoid nonlinear effects [70] due to a high power from the laser, which is the only source of light in the experiment. The intensity of the laser was set to about 6 nW. Let us notice that Eve is not playing any role here because the characterization of the QKD setup is accomplished in a protected environment. Therefore, we can safely assume that the reflectivity depends linearly on the polarization, as in ordinary Fresnel equations.

APPENDIX B: RATE EQUATIONS FOR THE TROJAN-HORSE ATTACK-GENERAL CASE

With the assumptions of the previous section on hand, let us describe the security argument in more detail. In the GLLP-Koashi approach [29,30], an entanglement-based description of the preparation stage is adopted. The states to be prepared are given in Eq. (5). We rewrite them here for convenience:

$$\begin{aligned}
|\psi_{0X}\rangle_{BE} &= |0_X\rangle_B \otimes |+\sqrt{\mu_{\text{out}}}\rangle_E,\\
|\psi_{1X}\rangle_{BE} &= |1_X\rangle_B \otimes |-\sqrt{\mu_{\text{out}}}\rangle_E,\\
|\psi_{0Y}\rangle_{BE} &= |1_Y\rangle_B \otimes |+i\sqrt{\mu_{\text{out}}}\rangle_E,\\
|\psi_{1Y}\rangle_{BE} &= |0_Y\rangle_B \otimes |-i\sqrt{\mu_{\text{out}}}\rangle_E.
\end{aligned} \tag{B1}$$

The X basis states of Eq. (B1) can be prepared by Alice by measuring in the basis $\{|0_X\rangle_A, |1_X\rangle_A\}$ the following entangled state:

$$|\Psi_X\rangle = \frac{|0_X\rangle_A|\psi_{0X}\rangle_{BE} + |1_X\rangle_A|\psi_{1X}\rangle_{BE}}{\sqrt{2}}. \tag{B2}$$

Similarly, the Y basis states of Eq. (B1) can be prepared by measuring in the basis $\{|0_Y\rangle_A, |1_Y\rangle_A\}$ the state

$$|\Psi_Y\rangle = \frac{|0_Y\rangle_A|\psi_{0Y}\rangle_{BE} + |1_Y\rangle_A|\psi_{1Y}\rangle_{BE}}{\sqrt{2}}. \tag{B3}$$

If the state preparation stage was perfect, the two states $|\Psi_X\rangle$ and $|\Psi_Y\rangle$ would be indistinguishable, as can be verified from the above equations in the limit $\mu_{\text{out}} \to 0$. In this case, we know that the secure key rate of the single-photon efficient BB84 protocol with data basis X and test basis Y would be

$$R_{\text{ideal}} = Q_X[1 - h(e_Y) - f_{\text{EC}}h(e_X)], \tag{B4}$$

where Q_X is the single-photon detection rate in the X basis, e_Y (e_X) is the error rate measured from single photons in the Y (X) basis, and $f_{EC} \geq 1$ is the inefficiency of error correction [71]. Because we are considering a single-photon source here, all the quantities in Eq. (B4) refer to the single-photon case.

When the preparation is not perfect, or part of the basis information leaks out of the transmitting unit, the states $|\Psi_X\rangle$ and $|\Psi_Y\rangle$ are different and the above key rate has to be replaced by the following one [31]:

$$R = Q_X[1 - h(e'_Y) - f_{\text{EC}}h(e_X)]. \tag{B5}$$

In Eq. (B5), the phase error rate e_Y has been replaced by a larger error rate, $e'_Y \geq e_Y$. It was shown in Ref. [30] that the term e'_Y is an upper bound to the error rate that the users would find if they measured the X-basis state $|\Psi_X\rangle$ in the basis Y.

To find the relation between the error rates e'_Y and e_Y, we can imagine that Alice owns a private bidimensional quantum system, a "quantum coin" [29], and prepares the following state:

$$|\Phi\rangle = \frac{|0_Z\rangle_C|\Psi_X\rangle + |1_Z\rangle_C|\Psi_Y\rangle}{\sqrt{2}}, \tag{B6}$$

where the subscript C refers to the quantum coin. The states in Eqs. (B1) can then be prepared by Alice by first

measuring the quantum coin in the basis $\{|0_Z\rangle_C, |1_Z\rangle_C\}$ and then, depending on the outcome, measuring the resulting state $|\Psi_X\rangle$ or $|\Psi_Y\rangle$ in the basis $\{|0_X\rangle_A, |1_X\rangle_A\}$ or $\{|0_Y\rangle_A, |1_Y\rangle_A\}$, respectively. Because Eve has no access to the quantum coin, she cannot distinguish this virtual preparation from the real preparation executed in the actual protocol. Therefore, we are allowed to think that Alice prepares her initial states using the quantum coin. Also, she can delay her measurement until after Bob has measured the states received from Alice. In this scenario, by noting that Eve's information about Alice's key does not change if Bob measures $|\Psi_X\rangle$ in the basis Y, Koashi obtained e'_Y from e_Y using a complementarity argument, by applying the "Bloch sphere bound" [72] to the quantum coin [30].

Let us quantify the basis dependence of Alice's states in terms of the quantum coin imbalance [29]. By rewriting Eq. (B6) in the X basis of the quantum coin, we find

$$|\Phi\rangle = \frac{|0_X\rangle_C(|\Psi_X\rangle + |\Psi_Y\rangle) + |1_X\rangle_C(|\Psi_X\rangle - |\Psi_Y\rangle)}{2}. \quad \text{(B7)}$$

To quantify the basis dependence of Alice's states, we need to evaluate the probability that the two states $|\Psi_X\rangle$ and $|\Psi_Y\rangle$ are different. From the above equation, it amounts to the probability that Alice obtains the outcome $X = -1$, associated with the state $|1_X\rangle_C$, when she measures the quantum coin in the basis X. We call this probability Δ:

$$\Delta = \mathrm{Prob}(X_C = -1) = \frac{1 - \mathrm{Re}(\langle\Psi_X|\Psi_Y\rangle)}{2}. \quad \text{(B8)}$$

Let us estimate this probability for the states prepared by Alice. From Eqs. (B2) and (B3), we can calculate

$$\Delta = \frac{1}{2}[1 - \exp(-\mu_{\mathrm{out}})\cos(\mu_{\mathrm{out}})]. \quad \text{(B9)}$$

When $\mu_{\mathrm{out}} = 0$, $\Delta = 0$ and the states emitted by Alice are basis independent. However, when $\mu_{\mathrm{out}} > 0$, Δ is positive and the states carry some basis information out of Alice's enclosure. The basis information can be exploited by Eve to enhance her strategy, acting on the channel losses, which are entirely under her control. Specifically, she can replace the real channel with another, loss-free channel. Then, she selectively stops all the states that are not favorable to her, until the loss rate measured by the users is matched. To account for this possibility, the users must consider the worst case, where all the nondetected events are coming from $X = 1$ eigenstates of the quantum coin, and renormalize Δ accordingly:

$$\Delta' = \frac{\Delta}{\mathcal{Y}}. \quad \text{(B10)}$$

In Eq. (B10), $\mathcal{Y} = \min(\mathcal{Y}_X, \mathcal{Y}_Y)$, with $\mathcal{Y}_X$ and $\mathcal{Y}_Y$ the single-photon yields measured in the X and Y bases, respectively. Finally, using the Bloch sphere bound [72] and the effective coin imbalance Δ', the relation between the phase error rates e'_Y and e_Y is obtained as [30,31]

$$e'_Y = e_Y + 4\Delta'(1-\Delta')(1-2e_Y) + 4(1-2\Delta')\sqrt{\Delta'(1-\Delta')e_Y(1-e_Y)}. \quad \text{(B11)}$$

When the single-photon source is replaced by a decoy-state source, and under assumption 3 of Appendix A, the resulting rate is a straightforward generalization of Eq. (B5) along the lines described in Ref. [38]. Indicating with a tilde the quantities to be estimated via the decoy-state technique, we have

$$\tilde{R} = \tilde{Q}_X^{(1)}\{1 - h[\tilde{e}_Y'^{(1)}]\} - Q_X^{(s)} f_{\mathrm{EC}} h[e_X^{(s)}], \quad \text{(B12)}$$

where

$$\tilde{e}_Y'^{(1)} = \tilde{e}_Y + 4\tilde{\Delta}'(1-\tilde{\Delta}')(1-2\tilde{e}_Y) + 4(1-2\tilde{\Delta}')\sqrt{\tilde{\Delta}'(1-\tilde{\Delta}')\tilde{e}_Y(1-\tilde{e}_Y)},$$
$$\tilde{\Delta}' = \frac{\Delta}{\tilde{\mathcal{Y}}}. \quad \text{(B13)}$$

In Eq. (B12), $\tilde{Q}_X^{(1)}$ is the decoy-state estimation of the single-photon detection rate Q_X [see Eq. (B5)], and $Q_X^{(s)}$ is the detection rate of the signal pulse measured in the X basis. In Eq. (B13), we conservatively defined $\tilde{\mathcal{Y}} := \min[\tilde{\mathcal{Y}}_X, \tilde{\mathcal{Y}}_Y]$, with $\tilde{\mathcal{Y}}_X$ and $\tilde{\mathcal{Y}}_Y$ the single-photon yields in the X and Y bases, respectively, estimated via the decoy-state technique.

APPENDIX C: RATE EQUATIONS FOR TWO SPECIFIC TROJAN-HORSE ATTACKS

1. Trojan-horse attack with passive use of the Trojan photons

We analyze the security of the BB84 protocol against a different, less general THA. This serves a twofold purpose: It provides an upper bound to the key rate achievable in the presence of a THA and gives us a chance to use a different proof method to study the THA.

In the specific THA of this section, Eve uses the information leaked from Alice in a passive way. She extracts from the quantum channel the states labeled with E in Eq. (B1) and stores them in a perfect quantum memory. This causes no disturbance in the quantum channel connecting Alice and Bob. Then, during the basis reconciliation stage of the BB84 protocol, Eve learns the basis information communicated by the users on a public channel. This information allows her to measure the stored states in the same bases as the users and learn the resulting key bit every time the result of her measurement is

conclusive. The conclusiveness of her results depends on the magnitude of the parameter μ_{out} in the stored states.

We analyze this situation using the *loss-tolerant* proof method described by Tamaki *et al.* in Ref. [67]. We can use the equations of the previous section up until Eq. (B5). The difference starts with the estimation of the phase error rate in the virtual protocol, e'_Y, which is more direct than in the GLLP-Koashi approach.

We consider a *real* protocol, in which Alice prepares the state in Eq. (B1) and sends it to Bob (and Eve), and a *virtual* protocol, in which an entanglement-based view is adopted. In both cases, we assume that Bob's measurement does not depend on the basis choice. This is guaranteed by assumption 4 discussed in Appendix A.

In the virtual protocol, Alice prepares the states to be sent to Bob by measuring her half of an entangled state. This is the same state as in Eq. (B2), which we rewrite here both in the X and in the Y basis:

$$|\Psi\rangle = \frac{|0_X\rangle_A|\psi_{0X}\rangle_{BE} + |1_X\rangle_A|\psi_{1X}\rangle_{BE}}{\sqrt{2}}, \tag{C1}$$

$$|\Psi\rangle = \frac{|0_Y\rangle_A|\phi_{1Y}\rangle_{BE} + |1_Y\rangle_A|\phi_{0Y}\rangle_{BE}}{\sqrt{2}}. \tag{C2}$$

In Eqs. (C1) and (C2), we have defined

$$|\phi_{1Y}\rangle_{BE} \coloneqq \frac{-i|0_Y\rangle_B|\epsilon_-\rangle_E + |1_Y\rangle_B|\epsilon_+\rangle_E}{\sqrt{2}}, \tag{C3}$$

$$|\phi_{0Y}\rangle_{BE} \coloneqq \frac{|0_Y\rangle_B|\epsilon_+\rangle_E + i|1_Y\rangle_B|\epsilon_-\rangle_E}{\sqrt{2}}, \tag{C4}$$

$$|\epsilon_\pm\rangle \coloneqq \frac{|\sqrt{\mu_{\text{out}}}\rangle \pm |-\sqrt{\mu_{\text{out}}}\rangle}{\sqrt{2}}. \tag{C5}$$

Notice that when $\mu_{\text{out}} \to 0$, $|\psi_{0X}\rangle_{BE} \to |0_X\rangle_B|v\rangle_E$, $|\psi_{1X}\rangle_{BE} \to |1_X\rangle_B|v\rangle_E$, $|\phi_{1Y}\rangle_{BE} \to |1_Y\rangle_B|v\rangle_E$ and $|\phi_{0Y}\rangle_{BE} \to |0_Y\rangle_B|v\rangle_E$, where $|v\rangle$ is the vacuum state, thus recovering from Eqs. (C1) and (C2) two maximally entangled states in a two-dimensional Hilbert space tensor product with the vacuum state. This situation is secure against the THA and constitutes a reference for our later argument in Appendix C 2. However, when $\mu_{\text{out}} > 0$, the effective Hilbert space's dimension becomes larger than two, favoring the THA. Note also that the states in Eq. (C5) are orthogonal but not normalized, while the states in Eqs. (C3) and (C4) are normalized. More specifically, $\langle\epsilon_\pm|\epsilon_\pm\rangle = 1 \pm \exp(-2\mu_{\text{out}})$, $\langle\epsilon_\pm|\epsilon_\mp\rangle = 0$, $\langle\phi_{wY}|\phi_{wY}\rangle = 1$, with $w = \{0, 1\}$.

Let us assume for the moment that the system E is not accessible either to Alice or to Eve. Under this assumption, the proof method in Ref. [67] applies. The reason for this is twofold. First, the security argument in Ref. [67] is based on the description given in Refs. [30,31], which allows for an enlarged dimension of Alice's Hilbert space. Second, Eve cannot perform a basis-dependent selection of the states emitted by Alice because the basis information is contained in the system E, which is not accessible to her. Therefore, as shown in Ref. [67], she cannot modify the transmission rates of Alice's states using the basis information potentially leaked from Alice's module. We notice that, also in the specific THA considered here, Eve has no chance of modifying the transmission rates during the quantum transmission due to the fact that Eve is allowed to access the system E only *after* the basis information has been publicly disclosed by the users.

Suppose that Alice measures the ancillary states of $|\Psi\rangle$ in the Y basis. Because the states in Eqs. (C3) and (C4) are normalized, she will obtain with probability $1/2$ the state $|0_Y\rangle$ and with probability $1/2$ the state $|1_Y\rangle$, thus projecting $|\Psi\rangle$ into one of the following two states, respectively:

$$\begin{aligned}\rho_B^{(0)} &= \text{Tr}_E(|\phi_{1Y}\rangle_{BE}\langle\phi_{1Y}|) \\ &= c_-|0_Y\rangle_B\langle 0_Y| + c_+|1_Y\rangle_B\langle 1_Y| \\ &= \frac{1}{2}[\hat{\sigma}_0 - \exp(-2\mu_{\text{out}})\hat{\sigma}_2],\end{aligned} \tag{C6}$$

$$\begin{aligned}\rho_B^{(1)} &= \text{Tr}_E(|\phi_{0Y}\rangle_{BE}\langle\phi_{0Y}|) \\ &= c_+|0_Y\rangle_B\langle 0_Y| + c_-|1_Y\rangle_B\langle 1_Y| \\ &= \frac{1}{2}[\hat{\sigma}_0 + \exp(-2\mu_{\text{out}})\hat{\sigma}_2],\end{aligned} \tag{C7}$$

where we have defined $c_\pm \coloneqq \langle\epsilon_\pm|\epsilon_\pm\rangle/2$ and introduced the identity operator in the two-dimensional Hilbert space $\hat{\sigma}_0$ and the Pauli matrix $\hat{\sigma}_2 = [(0, -i), (i, 0)]$. These operators are necessary to connect the Y-basis states of the virtual protocol, Eqs. (C3) and (C4), to the Y-basis states of the real protocol, contained in the third and fourth lines of Eq. (B1). Because any qubit state can be written as a linear combination of identity and Pauli matrices, its transmission rate can be obtained directly from the Pauli matrices' transmission rates [67]. Accordingly, we define the transmission rate of $\hat{\sigma}_k$, $k = \{0, 2\}$, as $q_{s_Y|k} \coloneqq \text{Tr}(\hat{D}_{s_Y}\hat{\sigma}_k)/2$, with $\hat{D}_{s_Y} \coloneqq \sum_l \hat{A}_l^\dagger \hat{M}_{s_Y}\hat{A}_l$, $\hat{A}_l$ an arbitrary operator associated with Eve's action and $\hat{M}_{s_Y}$ the operator representing Bob's POVM in the Y basis associated with the bit value s. We can then obtain the transmission rates in the virtual and real protocols as combinations of the $q_{s_Y|k}$'s.

Let us call p_y the probability that Alice and Bob both select the Y basis. In the real protocol (superscript r), the joint probability $\mathcal{P}_{j_Y,i_Y}$ that Alice sends out the state $|i_Y\rangle$ and Bob detects $|j_Y\rangle$ $(i, j = 0, 1)$ is for each pair of states:

$$\mathcal{P}^{(r)}_{0_Y,1_Y} = \frac{p_y^2}{2}(q_{0_Y|0} + q_{0_Y|2}),$$
$$\mathcal{P}^{(r)}_{1_Y,1_Y} = \frac{p_y^2}{2}(q_{1_Y|0} + q_{1_Y|2}),$$
$$\mathcal{P}^{(r)}_{0_Y,0_Y} = \frac{p_y^2}{2}(q_{0_Y|0} - q_{0_Y|2}),$$
$$\mathcal{P}^{(r)}_{1_Y,0_Y} = \frac{p_y^2}{2}(q_{1_Y|0} - q_{1_Y|2}). \tag{C8}$$

The corresponding probabilities in the virtual protocol (superscript v) are

$$\mathcal{P}^{(v)}_{0_Y,1_Y} = \frac{p_y^2}{2}(q_{0_Y|0} + e^{-2\mu_{\text{out}}} q_{0_Y|2}),$$
$$\mathcal{P}^{(v)}_{1_Y,1_Y} = \frac{p_y^2}{2}(q_{1_Y|0} + e^{-2\mu_{\text{out}}} q_{1_Y|2}),$$
$$\mathcal{P}^{(v)}_{0_Y,0_Y} = \frac{p_y^2}{2}(q_{0_Y|0} - e^{-2\mu_{\text{out}}} q_{0_Y|2}),$$
$$\mathcal{P}^{(v)}_{1_Y,0_Y} = \frac{p_y^2}{2}(q_{1_Y|0} - e^{-2\mu_{\text{out}}} q_{1_Y|2}). \tag{C9}$$

In order to define the phase error rate, we need to identify the error event in the virtual protocol. This can be done using Eqs. (B1), (C2), (C3) in the limit $\mu_{\text{out}} \to 0$. When there is no THA on the channel, Bob obtains the correct state $|1_Y\rangle$ ($|0_Y\rangle$) when Alice measures $|0_Y\rangle$ ($|1_Y\rangle$) on her ancillary states. Hence, we associate an error with both Alice and Bob obtaining the same state, $|0_Y\rangle$ or $|1_Y\rangle$. So the phase error rate can be written as

$$e'_Y = \frac{\mathcal{P}^{(v)}_{0_Y,0_Y} + \mathcal{P}^{(v)}_{1_Y,1_Y}}{\mathcal{P}^{(v)}_{0_Y,0_Y} + \mathcal{P}^{(v)}_{0_Y,1_Y} + \mathcal{P}^{(v)}_{1_Y,0_Y} + \mathcal{P}^{(v)}_{1_Y,1_Y}}. \tag{C10}$$

From Eqs. (C8) and (C9), we can rewrite the phase error rate in terms of the rates measured in the real protocol. The result is

$$e'_Y = \frac{1}{2}[1 - a^{(r)}_{\mathcal{P}} \exp(-2\mu_{\text{out}})], \tag{C11}$$

where we have set

$$a^{(r)}_{\mathcal{P}} = \frac{\sum_{i,j=\{0,1\}} (-)^{i+j+1} \mathcal{P}^{(r)}_{j_Y,i_Y}}{\sum_{i,j=\{0,1\}} \mathcal{P}^{(r)}_{j_Y,i_Y}}. \tag{C12}$$

The secure key rate is obtained by replacing the phase error of Eq. (C11) in Eq. (B5):

$$R^* = Q_X[1 - h(e'_Y) - f_{EC} h(e_X)]. \tag{C13}$$

The key rate in Eq. (C13) applies to slightly more general THA than the specific one considered in this section. It applies to all THA in which Eve cannot interact with the auxiliary Trojan-horse states [labeled with E in Eq. (B1)] *during* the transmission of the qubit states [labeled with B in Eq. (B1)]. We already noted that if Eve cannot access the auxiliary system E during the quantum transmission, she cannot selectively modify the transmission rates $\mathcal{P}$. Here, we additionally note that even if Eve changed her action, described by the operators $\hat{A}_l$, according to whether or not she will own the auxiliary system E after the basis reconciliation step, she would not gain more information about the final key. This descends from Koashi's proof method [30], upon which the proof described in Ref. [67] is built. There, it was shown that irrespective of who owns the auxiliary system, whether it is Alice or Eve, if the users can obtain a faithful estimation of the phase error rate e'_Y, they can, in principle, distill a perfect qubit in a Y eigenstate. When measured by Alice in the data basis X, the Y eigenstate always provides her with a fully random key bit, not predictable by Eve. Therefore, even if Eve tunes her choice of the operators $\hat{A}_l$ on the auxiliary system E, her knowledge of the final key does not increase. The only condition required is that Eve accesses the auxiliary system E *after* the quantum transmission has been completed by the users.

To adapt the key rate in Eq. (C13) to the decoy-state estimation technique, we exploit assumption 3 in Appendix A, according to which Eve cannot use the THA to modify the decoy-state estimation. Then, we need to show which quantities have to be estimated using decoy states. We use the tilde to explicitly indicate such quantities in the key rate:

$$\tilde{R}^* = \tilde{Q}^{(1)}_X \{1 - h[\tilde{e}'^{(1)}_Y]\} - Q^{(s)}_X f_{EC} h[e^{(s)}_X], \tag{C14}$$

where s is the mean photon number of the signal in the decoy-state set; $\tilde{Q}^{(1)}_X$ is the overall single-photon detection rate in the X basis; estimated using the decoy-state technique; and $Q^{(s)}_X$ and $e^{(s)}_X$ are, respectively, the measured detection and error rates for the signal in the X basis. Furthermore, we have set

$$\tilde{e}'^{(1)}_Y = \frac{1}{2}[1 - a^{(r)}_{\tilde{\mathcal{P}}} \exp(-2\mu_{\text{out}})],$$
$$a^{(r)}_{\tilde{\mathcal{P}}} = \frac{\sum_{i,j=\{0,1\}} (-)^{i+j+1} \tilde{\mathcal{P}}^{(r)}_{j_Y,i_Y}}{\sum_{i,j=\{0,1\}} \tilde{\mathcal{P}}^{(r)}_{j_Y,i_Y}}, \tag{C15}$$

which gives a straightforward generalization of Eqs. (C11) and (C12).

The key rates R^* and $\tilde{R}^*$ are plotted in Figs. 8 and 9. Neither of the resulting key rates shows strong dependence on the mean Trojan photon number μ_{out}. The key rates are coincident with the ideal rate corresponding to no THA for all values of μ_{out} smaller than $\sim 10^{-2}$, and they remain

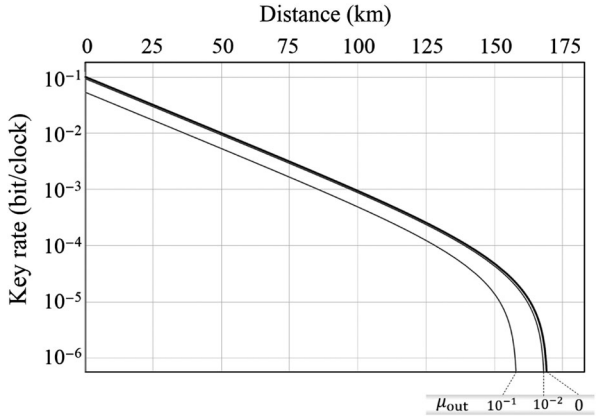

FIG. 8. Asymptotic key rate R^* versus distance for the single-photon efficient BB84 protocol, under a passive THA. The rate is plotted for various values of the parameter μ_{out}. Parameters in the simulation are as in Fig. 3.

positive up to values 0.5 (0.38) in the case of a single-photon (decoy-state) source. This represents an improvement of several orders of magnitude over the key rates for the general THA presented in Sec. II and suggests that the power of a THA comes from Eve's capability of selectively introducing losses in the transmission channel, conditional on the information gained from the THA. This observation motivates the study of a more involved THA, in which the shield system E is actively used.

2. Trojan-horse attack with active unambiguous state discrimination of the Trojan photons

We consider a particular THA in which Eve can access the ancillary system E, generated by the THA, during the quantum transmission, i.e., *before* the bases are revealed by the users. However, she can only measure it using a specific measurement described later on. This is more powerful than the THA considered in the previous section, but it is less powerful than the most general THA discussed in Sec. II and Appendix B.

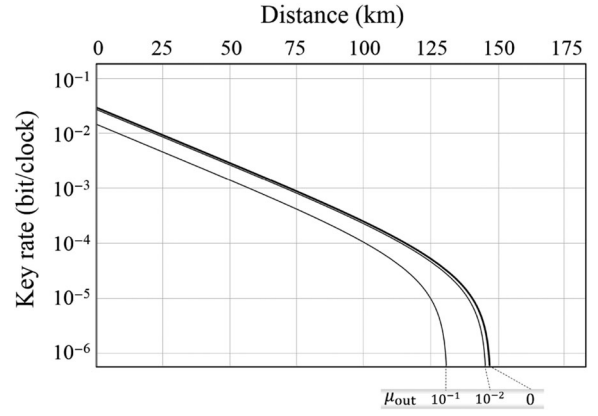

FIG. 9. Asymptotic key rate $\tilde{R}^*$ versus distance for the decoy-state efficient BB84 protocol, under a passive THA. The rate is plotted for various values of the parameter μ_{out}. Parameters in the simulation are as in Fig. 4.

In this THA, Eve accesses the space E of the Trojan photons during the quantum transmission stage. She then uses unambiguous state discrimination (USD) [73] to distinguish $|+\sqrt{\mu_{out}}\rangle_E$ from $|-\sqrt{\mu_{out}}\rangle_E$. These states correspond to Alice's states in the data basis (X), as per Eq. (B1). Therefore, whenever the USD succeeds, Eve knows the key bit without measuring, and hence perturbing, Alice's qubit.

However, the USD measurement does not always provide Eve with a conclusive result, and Eve's strategy can be improved as follows. When the result is conclusive, Eve transmits Alice's pulse to Bob without modification; when it is inconclusive, Eve stops Alice's pulse and introduces a loss in the communication channel. Later, during the basis reconciliation step, Eve will learn the bases chosen by Alice and Bob. After discarding the outcomes of the USD performed on Alice's Y-basis states, Eve will be left, ideally, with the same key bits as the users, distilled from the X basis, without having caused any noise in the communication channel.

Let us add more details to this scenario. When Alice prepares a Y-basis state, Eve's retrieved Trojan pulse is in a state $|\pm i\sqrt{\mu_{out}}\rangle_E$. This state cannot help Eve decide between the two outcomes related to the X basis because it is equally likely to be projected on either of the two X basis states $|\pm\sqrt{\mu_{out}}\rangle_E$. Therefore, Eve's decision to retain or transmit Alice's state is not related to an increased information gained by Eve and does not require an increase of the privacy amplification performed by the users. On the contrary, when Alice prepares a X-basis state, Eve can modify the transmission rates in a way that affects the security of the system. In a worst-case scenario, we then assume that all the counts detected by Bob come from the X basis and from a conclusive outcome of Eve's USD measurement.

Let us call p_{con} and $p_{inc} = 1 - p_{con}$ the probabilities of a conclusive and inconclusive outcome, respectively, from the USD of X-basis states. A lower bound on p_{inc} is given by the Ivanovic-Dieks-Peres bound [73–75]:

$$p_{inc} \geq |\langle\sqrt{\mu_{out}}| - \sqrt{\mu_{out}}\rangle| = \exp(-2\mu_{out}). \tag{C16}$$

Then, according to the above-described THA, the fraction of detected events in the X basis that have been transmitted conditional on a conclusive result by Eve is at most

$$\delta \leq \frac{1 - p_{inc}}{\mathcal{Y}_X} \leq \frac{1 - \exp(-2\mu_{out})}{\mathcal{Y}}, \tag{C17}$$

with $\mathcal{Y} \coloneqq \min[\mathcal{Y}_X, \mathcal{Y}_Y]$ and $\mathcal{Y}_X$ ($\mathcal{Y}_Y$) the single-photon yield in the X (Y) basis. The fraction δ (respectively $1-\delta$) contains insecure (secure) bits distilled by the users because they come from Eve's conclusive (inconclusive)

measurement. When the USD is inconclusive, Eve cannot selectively modify Alice's pulses using the system E. Therefore, following the same reasoning as in the previous section, we can apply the proof of Ref. [67] to this fraction of pulses to estimate the key rate.

In the present THA, whenever the USD provides a conclusive outcome, Eve forwards Alice's pulse to Bob without perturbing it. Therefore, only a fraction $1-\delta$ of the counts provide a faithful estimation of the error rate. After bounding the phase error rate as $e'_Y/(1-\delta)$, we can follow similar steps as in Ref. [29] to show that secure key bits can be extracted from the single-photon efficient BB84 protocol in the presence of the THA described here at a rate

$$R^{**} = Q_X\left\{(1-\delta)\left[1-h\left(\frac{e'_Y}{1-\delta}\right)\right]-f_{\mathrm{EC}}h(e_X)\right\}. \quad \text{(C18)}$$

Equations (C17) and (C18) can be easily generalized to the case of a decoy-state source under assumption 3 of Appendix A:

$$\tilde{R}^{**} = \tilde{Q}_X^{(1)}(1-\tilde{\delta}^{(1)})\left[1-h\left(\frac{\tilde{e}_Y^{\prime(1)}}{1-\tilde{\delta}^{(1)}}\right)\right]-Q_X^{(s)}f_{\mathrm{EC}}h[e_X^{(s)}],$$
$$\tilde{\delta}^{(1)} = \frac{1-\exp(-2\mu_{\mathrm{out}})}{\tilde{\mathcal{Y}}^{(1)}}, \quad \text{(C19)}$$

with $\tilde{\mathcal{Y}}^{(1)} = \min[\tilde{\mathcal{Y}}_X^{(1)}, \tilde{\mathcal{Y}}_Y^{(1)}]$. In Eq. (C19), the tilde indicates quantities to be estimated via the decoy-state technique. $Q_X^{(s)}$ and $e_X^{(s)}$ are the same as in Eq. (C14). The expression of the phase error rate $\tilde{e}_Y^{\prime(1)}$ is the same as in Eq. (C15) because it is estimated in the Y basis which, in this specific THA, does not allow Eve to selectively modify the transmission rate conditional on her information on Alice's state.

The key rates in Eqs. (C18) and (C19) are plotted in Figs. 10 and 11, respectively. Even though they are better

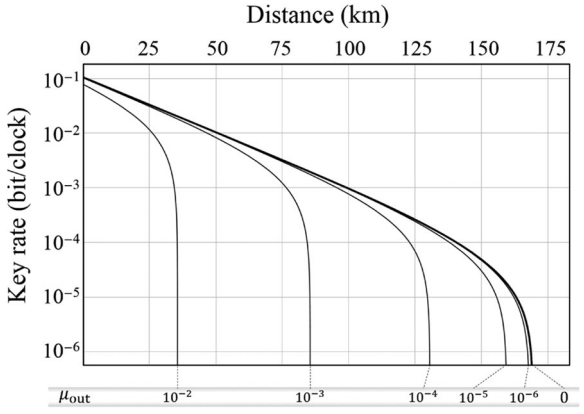

FIG. 10. Asymptotic key rate R^{**} versus distance for the single-photon efficient BB84 protocol, under a THA with unambiguous state discrimination by Eve. The rate is plotted for various values of the parameter μ_{out}. Parameters in the simulation are as in Fig. 3.

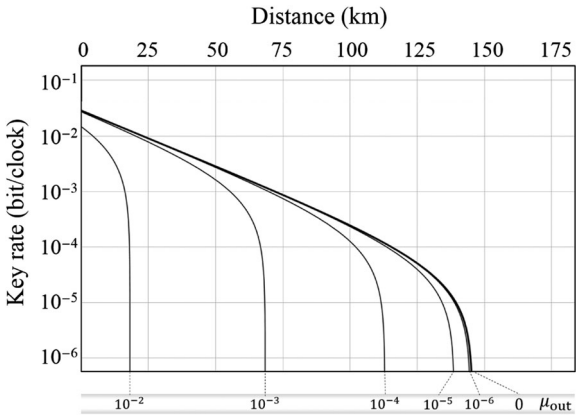

FIG. 11. Asymptotic key rate $\tilde{R}^{**}$ versus distance for the decoy-state efficient BB84 protocol, under a THA with unambiguous state discrimination by Eve. The rate is plotted for various values of the parameter μ_{out}. Parameters in the simulation are as in Fig. 4.

than the key rates in Figs. 3 and 4, drawn for the most general THA from the GLLP proof method [29], there is no wide gap between the two situations. This suggests that the particular THA described here catches the main features of the general attack described in Sec. II. It also suggests that the real-time use of the auxiliary system E is the main source of trouble in a THA. This seems to be particularly detrimental in the framework of Ref. [67], which heavily relies on Eve's inability to change the transmission rates of the states emitted by Alice.

It would be interesting to extend the proof method of Ref. [67] to more general Trojan-horse attacks than the one described in this section. However, such a generalization is not straightforward, and a separate analysis is required [76].

[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*, Vol. 175, pp. 175–179, http://researcher.watson.ibm.com/researcher/files/us-bennetc/BB84highest.pdf.

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

[3] 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 (2009).

[4] N. Lütkenhaus and A. J. Shields, *Focus on Quantum Cryptography: Theory and Practice*, New J. Phys. **11**, 045005 (2009).

[5] A. K. Ekert, *Quantum Cryptography Based on Bell's Theorem*, Phys. Rev. Lett. **67**, 661 (1991).

[6] D. Mayers and A. C.-C. Yao, *Quantum Cryptography with Imperfect Apparatus*, in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society, Washington, DC, 1998), pp. 503–509.

[7] J. Barrett, L. Hardy, and A. Kent, *No Signaling and Quantum Key Distribution*, Phys. Rev. Lett. **95**, 010503 (2005).

[8] 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).

[9] S. Braunstein and S. Pirandola, *Side-Channel-Free Quantum Key Distribution*, Phys. Rev. Lett. **108**, 130502 (2012).

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

[11] A. Vakhitov, V. Makarov, and D. R. Hjelme, *Large Pulse Attack as a Method of Conventional Optical Eavesdropping in Quantum Cryptography*, J. Mod. Opt. **48**, 2023 (2001).

[12] 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).

[13] N. Jain, E. Anisimova, I. Khan, V. Makarov, Ch. Marquardt, and G. Leuchs, *Trojan-Horse Attacks Threaten the Security of Practical Quantum Cryptography*, New J. Phys. **16**, 123030 (2014).

[14] N. Jain, B. Stiller, I. Khan, V. Makarov, C. Marquardt, and G. Leuchs, *Risk Analysis of Trojan-Horse Attacks on Practical Quantum Key Distribution Systems*, IEEE J. Sel. Top. Quantum Electron. **21**, 168 (2015).

[15] S. Sajeed, I. Radchenko, S. Kaiser, J.-P. Bourgoin, 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).

[16] I. Khan, N. Jain, B. Stiller, P. Jouguet, S. Kunz-Jacques, E. Diamanti, Ch. Marquardt, and G. Leuchs, *Trojan-Horse Attacks on Practical Continuous-Variable Quantum Key Distribution Systems*, http://2014.qcrypt.net/wp-content/uploads/2014-09-QCRYPT-2014-ImranKhan-3.pptx.

[17] A. Muller, T. Herzog, B. Hutter, W. Tittel, H. Zbinden, and N. Gisin, *Plug and Play Systems for Quantum Cryptography*, Appl. Phys. Lett. **70**, 793 (1997).

[18] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, *Quantum Key Distribution over 67 km with a Plug&Play System*, New J. Phys. **4**, 41 (2002).

[19] Y. Zhao, B. Qi, and H.-K. Lo, *Quantum Key Distribution with an Unknown and Untrusted Source*, Phys. Rev. A **77**, 052327 (2008).

[20] Y. Zhao, B. Qi, H.-K. Lo, and L. Qian, *Security Analysis of an Untrusted Source for Quantum Key Distribution: Passive Approach*, New J. Phys. **12**, 023024 (2010).

[21] Y. Zhao, B. Qi, and H.-K. Lo, *Experimental Quantum Key Distribution with Active Phase Randomization*, Appl. Phys. Lett. **90**, 044106 (2007).

[22] H.-K. Lo and H. F. Chau, *Unconditional Security of Quantum Key Distribution over Arbitrarily Long Distances*, Science **283**, 2050 (1999) (see Note 21, Proposition 1).

[23] C. H. Bennett, *Quantum Cryptography Using Any Two Nonorthogonal States*, Phys. Rev. Lett. **68**, 3121 (1992).

[24] V. Scarani, A. Acín, G. Ribordy, and N. Gisin, *Quantum Cryptography Protocols Robust Against Photon Number Splitting Attacks for Weak Laser Pulse Implementations*, Phys. Rev. Lett. **92**, 057901 (2004).

[25] Most of the results hold for the receiver side too, but the connection with security depends on the particular protocol used. In the standard BB84 protocol, the receiver is already secure against the THA (see Ref. [11]).

[26] R. J. Glauber, *Coherent and Incoherent States of the Radiation Field*, Phys. Rev. **131**, 2766 (1963).

[27] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic Publishers, Dordrecht, 1993).

[28] T. Sasaki, Y. Yamamoto, and M. Koashi, *Practical Quantum Key Distribution Protocol without Monitoring Signal Disturbance*, Nature (London) **509**, 475 (2014).

[29] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, *Security of Quantum Key Distribution with Imperfect Devices*, Quantum Inf. Comput. **4**, 325 (2004).

[30] M. Koashi, *Simple Security Proof of Quantum Key Distribution Based on Complementarity*, New J. Phys. **11**, 045018 (2009).

[31] H.-K. Lo and J. Preskill, *Security of Quantum Key Distribution Using Weak Coherent States with Nonrandom Phases*, Quantum Inf. Comput. **7**, 431 (2007).

[32] H.-K. Lo, H. F. Chau, and M. Ardehali, *Efficient Quantum Key Distribution Scheme and Proof of Its Unconditional Security*, J. Cryptol. **18**, 133 (2005).

[33] V. Scarani and R. Renner, *Quantum Cryptography with Finite Resources: Unconditional Security Bound for Discrete-Variable Protocols with One-Way Postprocessing*, Phys. Rev. Lett. **100**, 200501 (2008).

[34] G. Brassard and L. Salvail, *Advances in Cryptology, Eurocrypt 93* (Springer-Verlag, Berlin, 1993), pp. 410–423.

[35] M. Lucamarini, K. A. Patel, J. F. Dynes, B. Fröhlich, A. W. Sharpe, A. R. Dixon, Z. L. Yuan, R. V. Penty, and A. J. Shields, *Efficient Decoy-State Quantum Key Distribution with Quantified Security*, Opt. Express **21**, 024550 (2013).

[36] Z. L. Yuan, M. Lucamarini, J. F. Dynes, B. Fröhlich, A. Plews, and A. J. Shields, *Robust Random Number Generation Using Steady-State Emission of Gain-Switched Laser Diodes*, Appl. Phys. Lett. **104**, 261112 (2014).

[37] X.-B. Wang, *Beating the Photon-Number-Splitting Attack in Practical Quantum Cryptography*, Phys. Rev. Lett. **94**, 230503 (2005).

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

[39] A. J. Bennett, D. C. Unitt, P. See, A. J. Shields, P. Atkinson, K. Cooper, and D. A. Ritchie, *Electrical Control of the Uncertainty in the Time of Single Photon Emission Events*, Phys. Rev. B **72**, 033316 (2005).

[40] L. A. Ngah, O. Alibart, L. Labonté, V. D'Auria, and S. Tanzilli, *Ultra-Fast Heralded Single Photon Source Based on Telecom Technology*, Laser Photonics Rev. **9**, L1 (2015).

[41] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, *Practical Decoy State for Quantum Key Distribution*, Phys. Rev. A **72**, 012326 (2005).

[42] R. Loudon, *The Quantum Theory of Light* (Oxford University Press, New York, 1973).

[43] S. W. Allison, G. T. Gillies, D. W. Magnuson, and T. S. Pagano, *Pulsed Laser Damage to Optical Fibers*, Appl. Opt. **24**, 3140 (1985).

[44] L. W. Tutt and T. F. Boggess, *A Review of Optical Limiting Mechanisms and Devices Using Organics, Fullerenes,*

SemiConductors and Other Materials, Prog. Quantum Electron. **17**, 299 (1993).
[45] See http://www.iso.org., The standard ISO 21254 (2011) replaces the standard ISO 11254 (2000).
[46] By definition, LIDT establishes a point where zero damage occurs on the optical component. Hence, the resulting damage threshold is a loose bound. A tighter bound would be one where damage certainly occurs on the component.
[47] R. M. Wood, *Laser-Induced Damage of Optical Materials* (Taylor & Francis, 2003).
[48] D. Du, X. Liu, G. Korn, J. Squier, and G. Mourou, *Laser-Induced Breakdown by Impact Ionization in* SiO_2 *with Pulse Widths from 7 ns to 150 fs*, Appl. Phys. Lett. **64**, 3071 (1994).
[49] B. C. Stuart, M. D. Feit, A. M. Rubenchik, B. W. Shore, and M. D. Perry, *Laser-Induced Damage in Dielectrics with Nanosecond to Subpicosecond Pulses*, Phys. Rev. Lett. **74**, 2248 (1995).
[50] M. Mero, J. Liu, W. Rudolph, D. Ristau, and K. Starke, *Scaling Laws of Femtosecond Laser Pulse Induced Breakdown in Oxide Films*, Phys. Rev. B **71**, 115109 (2005).
[51] C. W. Carr, H. B. Radousky, and S. G. Demos, *Wavelength Dependence of Laser-Induced Damage: Determining the Damage Initiation Mechanisms*, Phys. Rev. Lett. **91**, 127402 (2003).
[52] http://fibercore.com/expertise/fiberpaedia/bend-edge.
[53] Q. Wang, G. Farrell, and T. Freir, *Theoretical and Experimental Investigations of Macro-Bend Losses for Standard Single Mode Fibers*, Opt. Express **13**, 4476 (2005).
[54] The value 1850 nm has been extrapolated from Fig. 3 of Ref. [53]. It entails an attenuation of 178 dB for 0.66 m of standard single-mode optical fiber wound on a spool of 1 cm radius.
[55] This is the amount of energy transferred to the medium before it starts dissipating heat. The characteristic dissipation time for fused silica is 100 μs, where we used $D = 0.75\ \mathrm{mm}^2/\mathrm{s}$ for the thermal diffusivity in fused silica [47] and the mentioned value of the core area.
[56] R. Kashyap, *Self-Propelled Self-Focusing Damage in Optical Fibers*, in *Proceedings of the XXth International Conference on Lasers, Nevada, USA*, edited by F. J. Duarte (STS Press, 1987), pp. 859–866.
[57] R. Kashyap and K. J. Blow, *Observation of Catastrophic Self-Propelled Self-Focusing in Optical Fibers*, Electron. Lett. **24**, 47 (1988).
[58] The laser used in Ref. [57] featured $\lambda = 1.064\ \mu$m. The corresponding LIDT at $\lambda = 1.55\ \mu$m is only a factor 1.2 larger, in line with the square-root dependence of LIDT on the wavelength [51].
[59] D. D. Davis, S. C. Mettler, and D. J. DiGiovani, *Experimental Data on the Fiber Fuse*, in *Proceedings of SPIE 2714, 27th Annual Boulder Damage Symposium: Laser-Induced Damage in Optical Materials, 1996*, p. 202, http://dx.doi.org/10.1117/12.240382.
[60] S. Yanagi, S. Asakawa, and R. Nagase, *Characteristics of Fibre-Optic Connector at High-Power Optical Incidence*, Electron. Lett. **38**, 977(2002).
[61] P. S. André, A. M. Rocha, F. Domingues, and A. Martins, *Thermal Model for Optical Fiber Coating under Tight Bending Diameters*, Electron. Lett. **46**, 695 (2010).
[62] S. Todoroki and S. Inoue, *Optical Fuse Made of Silica Glass Optical Fibers Spliced through Low-Melting Glass with Carbon-Coating*, Proceedings of the XXth International Congress on Glass, Report No. O-14-010, Kyoto, Japan (2004).
[63] R. Kashyap, *The Fiber Fuse—From a Curious Effect to a Critical Issue: A 25th Year Retrospective*, Opt. Express **21**, 6422 (2013).
[64] S. Todoroki, *Quantitative Evaluation of Fiber Fuse Initiation Probability in Typical Single-Mode Fibers*, in Optical Fiber Communication Conference, OSA Technical Digest (online) (Optical Society of America, 2015), http://dx.doi.org/10.1364/OFC.2015.W2A.33.
[65] A. R. Dixon, J. F. Dynes, M. Lucamarini, B. Fröhlich, A. W. Sharpe, A. Plews, S. Tam, Z. L. Yuan, Y. Tanizawa, H. Sato, S. Kawamura, M. Fujiwara, M. Sasaki, and A. J. Shields, *High Speed Prototype Quantum Key Distribution System and Long Term Field Trial*, Opt. Express **23**, 7583 (2015).
[66] P. Eraerds, M. Legré, J. Zhang, H. Zbinden, and N. Gisin, *Photon Counting OTDR: Advantages and Limitations*, IEEE J. Lightwave Techn. **28**, 952 (2010).
[67] K. Tamaki, M. Curty, G. Kato, H.-K. Lo, and K. Azuma, *Loss-Tolerant Quantum Cryptography with Imperfect Sources*, Phys. Rev. A **90**, 052314 (2014).
[68] It would be useful to have an optical fuse at less than 300 mW, which is a common tolerance level specified for fiber-based components. However, despite some recent advertisement, to our knowledge such a device is not yet commonly available.
[69] M. Koashi, *Efficient Quantum Key Distribution with Practical Sources and Detectors*, arXiv:quant-ph/0609180.
[70] R. W. Boyd, *Nonlinear Optics*, 3rd ed. (Academic Press, New York, 2008).
[71] Although we are in the asymptotic scenario, where error correction can, in principle, reach the Shannon limit ($f_{\mathrm{EC}} = 1$), we follow Ref. [38] in maintaining a factor $f_{\mathrm{EC}} > 1$ in the rate equation to provide a more realistic prevision on the final rate.
[72] K. Tamaki, M. Koashi, and N. Imoto, *Unconditionally Secure Key Distribution Based on Two Nonorthogonal States*, Phys. Rev. Lett. **90**, 167904 (2003).
[73] I. D. Ivanovic, *How to Differentiate between Non-orthogonal States*, Phys. Lett. A **123**, 257 (1987).
[74] D. Dieks, *Overlap and Distinguishability of Quantum States*, Phys. Lett. A **126**, 303 (1988).
[75] A. Peres, *How to Differentiate between Non-orthogonal States*, Phys. Lett. A **128**, 19 (1988).
[76] K. Tamaki (private communication).