


Trojan-horse attack on a real-world quantum key distribution system: Theoretical and experimental security analysis

Ivan S. Sushchev,^{1,2,*} Daniil S. Bulavkin,¹ Kirill E. Bugai,¹ Anna S. Sidelnikova,^{1,2} and Dmitriy A. Dvoretzkiy¹

¹*SFB Laboratory, Ltd, Moscow 127273, Russia*

²*Quantum Technology Centre and Faculty of Physics, Lomonosov Moscow State University, Moscow 119991, Russia*

 (Received 9 April 2024; revised 28 June 2024; accepted 6 August 2024; published 12 September 2024)

This paper presents a theoretical and experimental demonstration of a security analysis of a Trojan-horse attack (THA) on a real-world quantum key distribution (QKD) system. We show that the upper bound on the information leakage depends solely on the fidelity between the states of the adversary. We find the lower bound for fidelity between THA states in both the polarization- and phase-coding BB84 protocols, considering both pure and mixed states. Our bounds depend only on the mean photon number per pulse available to an adversary. We also present an experimental analysis of a QKD system, including optical time-domain reflectometry measurements with centimeter resolution and spectral transmittance measurements for optical defense elements ranging from 1100 to 1800 nm with a noise floor lower than -100 dB. Finally, by considering the optimal attack, we obtain the value of the mean photon number per pulse available to an adversary and calculate the key leakage that needs to be eliminated during the privacy amplification procedure.

DOI: [10.1103/PhysRevApplied.22.034032](https://doi.org/10.1103/PhysRevApplied.22.034032)

I. INTRODUCTION

Quantum key distribution (QKD) systems are theoretically capable of providing absolute security in key exchange between two parties, commonly referred to as Alice and Bob, which is ensured by the principles of quantum mechanics [1]. Security of QKD protocols is usually achieved through the privacy amplification procedure, as long as the information gain of an eavesdropper (Eve) is limited by an upper bound, which can be expressed as a function of the quantum bit error rate (QBER) estimated by legitimate users during the session. This bound is set by the entropic uncertainty relations [2,3]; however, there are practical imperfections in real-world implementations of QKD systems. The first is the employment of weak coherent pulses (WCPs) instead of the single-photon radiation for which security proofs were initially provided. They can usually be extended to the WCP case by placing an intensity modulator inside Alice's setup and running the

decoy-state method [4,5], which guarantees secrecy provided that accurate modulation takes place [6] or methods to counter statistical fluctuations are used [7]. Other problems include the imperfections of QKD apparatus leading to different types of fake-state attacks [8–11], vulnerabilities to external influences [12–16], and side channels [17–21]. Separately or in combination, these loopholes may be successfully exploited by an adversary to steal the secret key. It is therefore necessary to implement robust mitigation strategies through both protocol modifications and additional optical elements, as well as to carry out a comprehensive security analysis for a particular QKD system, considering every possible loophole.

The Trojan-horse attack (THA) on a QKD system is a well-known approach to stealing the secret key using an optical side channel [17–19,22–29]. It consists of the following: Eve injects a bright optical pulse through the communication channel into Alice or Bob's setup, where it passes through a state-preparation device (for Alice) or an active-basis-choice device (for Bob), which is usually a phase modulator (PM). Inhomogeneous regions inside optical devices or optical connections will lead to Fresnel reflections of Eve's pulse back to her station, where she can measure its state and, possibly, gain the information about the bit encoded Fig. 1. An experimental demonstration of the Trojan-horse attack has been presented previously using 1536 nm radiation [23] (this, however, failed due to

*Contact author: sushchev.is16@physics.msu.ru

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

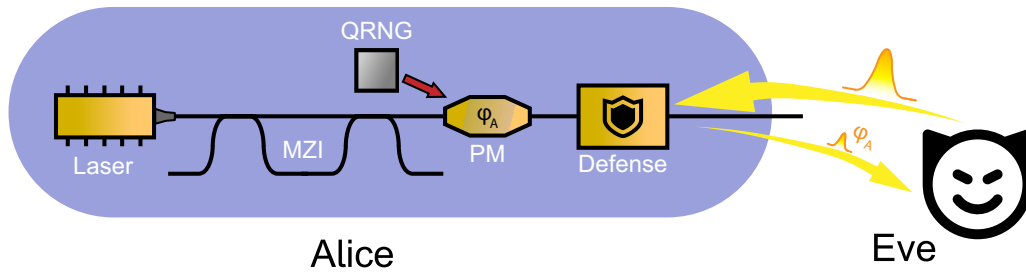


FIG. 1. Schematic of a Trojan-horse attack on a phase-coding fiber-based QKD system on the Alice side. Laser, source of signal pulses; MZI, Mach-Zehnder interferometer; QRNG, quantum random number generator for state choice; PM, phase modulator; Defense, the components used to attenuate Eve’s radiation (attenuators, optical isolators, circulators, band-pass filters, etc.).

a rise in QBER caused by Eve-induced afterpulsing) and 1924 nm [18] radiation. This emphasizes the necessity of designing a defense layout against THAs. As we show in Sec. II, Eve’s information gain is limited by the indistinguishability of vacuum terms in the states returned to Eve; therefore, it would be preferable to minimize the intensity of radiation returned to Eve, i.e., the mean photon number in Eve’s pulse μ_{Eve} . This can be achieved by equipping the system with attenuators (which are usually presented in Alice’s setup for WCP preparation) to suppress light in both directions, optical isolators and circulators to suppress light in the nonlegitimate direction, and band-pass filters and wavelength-division multiplexers to suppress light outside the legitimate spectral region. Attenuators appear to be impractical for Bob’s setup, which, however, often contains a watchdog detector to limit the maximum input radiation. Nonetheless, as an active device, it may introduce additional loopholes [30]. An alternative solution is fully passive QKD without active devices that are vulnerable to THA [31].

With the proper functioning of the defense, the μ_{Eve} value drops to an extremely low level, so that $\mu_{\text{Eve}} \ll 1$. It is also important to estimate a certain value of μ_{Eve} to nullify the information accessible to Eve using privacy amplification [32]. As demonstrated in Ref. [17], in case of side-channel attacks, Eve’s information gain is associated with the Holevo quantity [33]. The remarkable result is that it depends only on the fidelity or scalar product between two Trojan-light states corresponding to different bits in a known basis; however, Eve can prepare an arbitrary input state limited only by maximum input power. In this paper, we extend the results given in Ref. [17] to an arbitrary state case, either pure or mixed, and demonstrate that the fidelity between THA states has a lower bound both for polarization and phase coding depending only on μ_{Eve} . We also describe a complete experimental approach to estimating the exact μ_{Eve} value, even when defense elements with up to 100-dB attenuation are used. We apply our methods to a real phase-coding QKD system and, for the first time, present a comprehensive security analysis of a THA.

II. THEORETICAL ANALYSIS

A. Holevo bound for THA

In this section, we describe a theoretical approach to estimation of Eve’s information gain by bounding the Holevo quantity for her quantum ensemble. We show that this only depends on the fidelity between THA states. After that, considering the BB84 protocol, we demonstrate that the fidelity can be bounded for polarization and phase coding. We examine the cases of pure states and mixed states and derive that the fidelity is always limited by the probability of the vacuum component which, in turn, can be bounded with a known value of μ_{Eve} , which is consistent with the previous results [22,24,25].

General theoretical THA security analyses have been presented in Refs. [17,25]. Both papers focused on coherent states returned to Eve based on the assumption that any state tends to exhibit Poissonian statistics under high losses; however, this is not a common case, especially for setups without highly attenuative elements. A more general bound was reported in Ref. [24], yet the authors made several model assumptions (such as particular Kraus operators and their Taylor expansion), and their Eve did not possess knowledge of the selected basis. Here, we abandon every model assumption except that the phase modulator only rotates the phases of Fock components without affecting their absolute values. Moreover, we suggest that Eve can wait until the basis reconciliation procedure finishes and eventually discriminate the states in a known basis.

A well-known result for the secret key rate is [17]:

$$\ell = 1 - \chi_{\text{Eve}} - \text{leak}, \quad (1)$$

where χ_{Eve} is the mutual information between Alice and Eve, which can be expressed as the Holevo quantity [33], and “leak” is the information spent on error correction. The density matrix for a general attack exploiting side channels represents an equiprobable mixture of tensor products in which each multiplier is responsible for a side channel. Here, we examine a single side channel case related to THA, but our results can be easily extended to a general

case by constructing density matrices for the other side channels (see the example in Ref. [17]).

The density matrix of the Alice-Eve system in the presence of the THA channel is [17]

$$\begin{aligned} \rho_{XE} &= \frac{1}{2} |0\rangle_X \langle 0| \\ &\otimes [(1-Q) |\Phi_0\rangle_Q \langle \Phi_0| + Q |\Theta_0\rangle_Q \langle \Theta_0|] \otimes \rho_{\text{THA}}^0 \\ &+ \frac{1}{2} |1\rangle_X \langle 1| \\ &\otimes [(1-Q) |\Phi_1\rangle_Q \langle \Phi_1| + Q |\Theta_1\rangle_Q \langle \Theta_1|] \otimes \rho_{\text{THA}}^1, \end{aligned} \quad (2)$$

where $|0\rangle_X$ and $|1\rangle_X$ are the signal states in a chosen basis (for instance, the Z basis), $|\Phi_{0,1}\rangle_Q$ and $|\Theta_{0,1}\rangle_Q$ are the perturbed states of Eve's auxiliary subsystem (see details in Ref. [17]), $\rho_{\text{THA}}^{0,1}$ are the states of the THA channel, and Q is the QBER between Alice and Bob. It can be shown [17] that for optimal attack:

$$\begin{aligned} \langle \Phi_0 | \Phi_1 \rangle_Q &= \langle \Theta_0 | \Theta_1 \rangle_Q = \varepsilon, \\ \langle \Phi_{0,1} | \Theta_{0,1} \rangle_Q &= 0, \end{aligned} \quad (3)$$

where $\varepsilon = 1 - 2Q$.

The density matrix of the Eve system is given by the partial trace over all Alice's states:

$$\begin{aligned} \rho_E &= \frac{1}{2} [(1-Q) |\Phi_0\rangle_Q \langle \Phi_0| + Q |\Theta_0\rangle_Q \langle \Theta_0|] \otimes \rho_{\text{THA}}^0 \\ &+ \frac{1}{2} [(1-Q) |\Phi_1\rangle_Q \langle \Phi_1| + Q |\Theta_1\rangle_Q \langle \Theta_1|] \otimes \rho_{\text{THA}}^1 \\ &\equiv \frac{1}{2} \rho_Q^0 \otimes \rho_{\text{THA}}^0 + \frac{1}{2} \rho_Q^1 \otimes \rho_{\text{THA}}^1. \end{aligned} \quad (4)$$

One can calculate Eve's information gain via the Holevo quantity of the resulting quantum ensemble, but this seems challenging without knowing the structure of the THA channel states. Nonetheless, as we will show, it is sufficient to know only the Uhlmann fidelity [34] between ρ_{THA}^0 and ρ_{THA}^1 , which is connected to the probability of the vacuum component and Eve's mean photon number per THA pulse.

Let us make use of the result obtained in Ref. [35]. For the quantum ensemble with two density matrices ρ_1 and ρ_2 occurring with probabilities $1/2$, the Holevo quantity has an upper limit:

$$\begin{aligned} \chi_{\text{Eve}} &\leq \bar{\chi}_{\text{Eve}} \equiv h \left(\frac{1 - \sqrt{\mathcal{F}}}{2} \right) \\ &\equiv -\frac{1 + \sqrt{\mathcal{F}}}{2} \log_2 \frac{1 + \sqrt{\mathcal{F}}}{2} \\ &\quad - \frac{1 - \sqrt{\mathcal{F}}}{2} \log_2 \frac{1 - \sqrt{\mathcal{F}}}{2}, \end{aligned} \quad (5)$$

where $\sqrt{\mathcal{F}} = \text{tr} \sqrt{\sqrt{\rho_1} \rho_2 \sqrt{\rho_1}}$ is the square root of the Uhlmann fidelity. In our case, the density matrices are $\rho_Q^0 \otimes \rho_{\text{THA}}^0$ and $\rho_Q^1 \otimes \rho_{\text{THA}}^1$. It is convenient to use the multiplicativity property of fidelity [36]:

$$\begin{aligned} \mathcal{F}(\rho_Q^0 \otimes \rho_{\text{THA}}^0, \rho_Q^1 \otimes \rho_{\text{THA}}^1) \\ = \mathcal{F}(\rho_Q^0, \rho_Q^1) \cdot \mathcal{F}(\rho_{\text{THA}}^0, \rho_{\text{THA}}^1). \end{aligned} \quad (6)$$

Thus, fidelities can be calculated separately. The fidelity square root for Eve's auxiliary subsystem can be calculated directly (see Appendix A):

$$\sqrt{\mathcal{F}(\rho_Q^0, \rho_Q^1)} = \varepsilon. \quad (7)$$

This result is in good agreement with a standard BB84 case without side channels (i.e., when the THA fidelity equals 1), as

$$\bar{\chi}_{\text{Eve}} = h \left(\frac{1 - \varepsilon}{2} \right) = h \left(\frac{1 - 1 + 2Q}{2} \right) = h(Q), \quad (8)$$

which corresponds perfectly to the entropic uncertainty relations [3].

Let us now explore the $\eta = \sqrt{\mathcal{F}(\rho_{\text{THA}}^0, \rho_{\text{THA}}^1)}$ value. For pure states, the fidelity square root is reduced to the scalar product of the state vectors, and the result matches the definite solution in Ref. [17]. In our general case, $\bar{\chi}_{\text{Eve}}$ acts as an upper bound, which can be estimated regardless of particular pure or mixed THA states, as shown below.

B. Vacuum probability estimation

As a first step, let us examine the connection between the mean photon number μ and the probability of the vacuum component:

$$\mu = \sum_{n=0}^{\infty} p_n n = \sum_{n=1}^{\infty} p_n n \geq \sum_{n=1}^{\infty} p_n = 1 - p_0, \quad (9)$$

where p_n is the probability of the n -photon component. Thus, we express the lower bound for the probability of the vacuum component, which is state-independent (both for pure and mixed states):

$$p_0 \geq 1 - \mu. \quad (10)$$

This estimate only makes sense for $\mu < 1$, which, however, corresponds to our case, as we imply the proper functioning of the THA defense.

It can be shown (see Appendices B, C, E, and F) that the η value for both polarization and phase-coded states can be bounded when p_0 is known, which can be intuitively

TABLE I. Lower bounds for fidelity square roots for the polarization and phase-coding cases.

	Polarization coding	Phase coding
Pure states	$\eta = p_0 \geq 1 - \mu_{\text{Eve}}$	$\eta \geq 2p_0 - 1 \geq 1 - 2\mu_{\text{Eve}}$
Mixed states	$\eta \geq p_0 \geq 1 - \mu_{\text{Eve}}$	$\eta \geq \sqrt{2p_0^2 - 1}$ $\geq \sqrt{2(1 - \mu_{\text{Eve}})^2 - 1}$

interpreted as the distinguishability of different states being determined by the indistinguishable vacuum component fraction.

In real QKD systems, it is mandatory to consider the worst-case scenario, that is, the mixed-states case. However, a question arises: is it beneficial for Eve to use nonpure states? To answer this question, we constructed the bounds for both pure and mixed states. The particular results are presented in Table I.

We found that using mixed polarization states instead of pure states gives no advantage to Eve; i.e., the pure states are optimal, as implied in Refs. [17,22,25] based on an intuitive understanding of quantum coherence. The same hypothesis was neither proved nor disproved for phase encoding, as we obtained different bounds for the pure and mixed states. The states reported in Ref. [24] as optimal are also intrinsically mixed, although some special assumptions were made. We can conclude that there is room for further research to find the optimal THA states for phase encoding and tighten the bound. Until then, our conservative mixed-states bound can be used.

III. EXPERIMENTAL ANALYSIS

This experimental analysis of a THA in the context of a QKD system seeks to determine the maximum value of μ_{Eve} . The optimal strategy for Eve is to apply optical pulses that coincide with electrical gates on the PM under attack [25]. Therefore, she adjusts the time delay of her pulses to maximize their reflection and reach the PM at the optimal moment. It is consistent for Eve to set the pulse repetition frequency f_{Eve} to be equal to the legitimate pulse repetition rate f . In this case, her maximum information gain corresponds to the uniform distribution of the mean photon number per pulse due to the concavity of the $\bar{\chi}_{\text{Eve}}(\mu_{\text{Eve}})$ function [even for convex $\eta(\mu_{\text{Eve}})$ dependence (F7)]; hence, the mean photon number in each pulse is the same and equals μ_{Eve} . More effectively, she can decrease the attacking frequency and block the legitimate pulses that are not affected by the THA. Finally, Eve can tune the attacking wavelength λ to maximize μ_{Eve} . Thus, Eve's mean photon number per pulse can be calculated as

follows:

$$\mu_{\text{Eve}}(\lambda) = \frac{P_{\text{Eve}}(\lambda) \cdot \lambda}{f_{\text{Eve}} h c}, \quad (11)$$

where P_{Eve} is the average power of Eve's radiation, λ is the attacking wavelength, f_{Eve} is the attack repetition frequency, $h \approx 6.63 \times 10^{-34}$ J Hz⁻¹ is the Planck constant, and $c \approx 3 \times 10^8$ m/s is the speed of light in vacuum.

We now consider some realistic boundaries for expression (11). First, there might be a lower limit for f_{Eve} , despite the fact that the BB84 protocol implies only QBER estimation and does not call for monitoring of the detection frequency. In realistic QKD devices, a low key-generation rate will induce a timeout error, and the QKD session will fail. The specific value of the maximum allowable QKD session period depends on a particular system realization, but it is reasonable to set it no more than 10–100 times higher than the regular one; however, Eve can replace the lossy communication channel with an ideal one, resulting in a key-generation frequency boost. For a conventional SMF-28 optical fiber with approximately 0.2-dB/km losses and 100-km communication length, Eve can gain an additional 20 dB. Therefore, the effective attack repetition frequency is $f_{\text{Eve}} = \alpha f$, where α is typically no less than 10^{-4} .

A straightforward THA implies the use of a wavelength equal to the legitimate one, which usually lies in the telecommunications region (1260–1625 nm) for fiber QKD and near-infrared for free space; however, it has been shown that THAs can be effective both in the visible [26] and mid-infrared regions [18,19,27–29]. In practice, only two spectral regions are feasible for security analysis: 200–1100 nm (Si detector sensitivity) and 1100–2600 nm (InGaAs detector sensitivity). They can nonetheless be extended by the employment of superconducting nanowire single-photon detectors, although this is still challenging and has not yet been demonstrated in the THA analysis context.

Finally, the existence of the upper bound for Eve's radiation power is the primary limiting factor for μ_{Eve} . It can be set up either by the sensitivity level of a watchdog detector or, more generally, by the laser-induced damage threshold (LIDT) of an optical component inside the device under attack [25]. Exceeding the LIDT during the QKD session will disrupt the link between Alice and Bob, leading to failure of both the communication and the attack. The LIDT for a realistic QKD setup is considered to be approximately 10 W (40 dBm), with insignificant variations inside the attainable spectral region; however, it is important to note that the influence of high-power radiation can be substantially different for various optical elements, and a laser damage attack (LDA) is possible [12–15]. Hence, the susceptibility to LDA of every optical component should be verified. The actual LIDT value for a

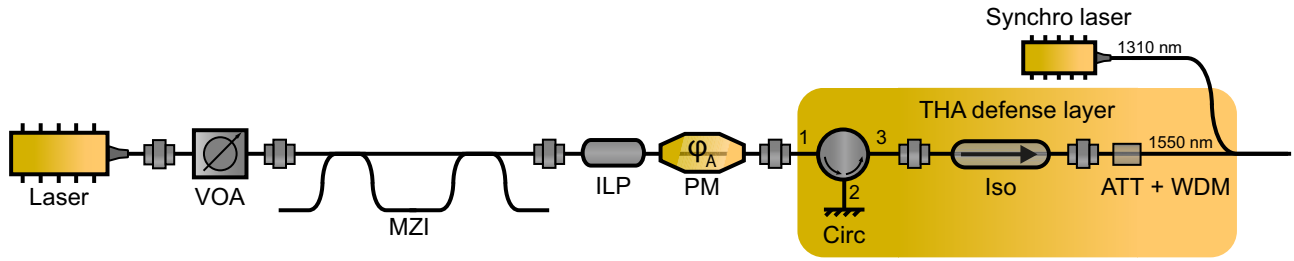


FIG. 2. The QKD system under test (Alice's side). Laser, the source of signal pulses; VOA, variable optical attenuator; MZI, Mach-Zehnder interferometer; ILP, inline polarizer; PM, phase modulator; Circ, circulator with the mirror connected to port 2; Iso, isolator; ATT+WDM, wavelength division multiplexer combined with the fixed attenuator; Synchro laser, source of synchronization pulses.

particular QKD system should also be measured during LDA security analysis.

The power of Eve's radiation can be expressed as a sum of three terms that can be revealed experimentally:

$$P_{\text{Eve}} [\text{dBm}] = P_{\text{max}} [\text{dBm}] + T[\text{dB}] + R [\text{dB}], \quad (12)$$

where P_{max} is the maximum power of input Eve's radiation, T is the transmittance of all passive optical defense components inside the setup under test, and R is the maximum reflectivity inside the setup under test in the absence of defense components. Here, P_{max} is either the sensitivity of the watchdog detector or the LIDT value for the system under test, which can be obtained by LDA security analysis [13–15]. Next, we will describe the methods for experimental estimation of T and R and present the results for Alice's side of a real phase-coding QKD system [37] with a THA defense layer (Fig. 2).

A. OTDR

The primary physical principle underlying the THA is the Fresnel reflection of light from the optical interfaces. QKD systems consist of a set of interconnected fiber-optic elements. Typically, the connection is made using optical connectors, such as FC-type connectors. Due to the presence of an air gap between the ferrules of optical connectors, an interface arises, leading to reflection with a magnitude varying from -60 dB to -20 dB depending on the type of connector [22,23,25,38]. The best performance is provided by angled physical contact connectors with angled polishing, so their use in QKD represents a simple but quite effective measure against THA. Reflection from the connections can be suppressed to a minimum using fusion splicing of the pigtails of various optical elements; this, however, complicates the assembly, debugging, and maintenance of the setup.

The connections of various optical elements are not the only source of parasitic reflections. Usually, QKD systems contain a number of optical elements with a complex and heterogeneous internal structures. Significant return losses (of the order of 30–60 dB) are often observed from phase modulators, detector surfaces, and lasers. Moreover, many

QKD systems contain mirrors, leading to almost 100% reflection. This is the case for two-pass [39] and single-pass [37] plug-and-play QKD systems, which contain Faraday mirrors at Alice's and Bob's sides, respectively, to correct polarization distortion. Some implementations of optical modulators in QKD (i.e., polarization and intensity modulators) contain Faraday mirrors to prevent polarization mode dispersion [40,41]. Therefore, it is necessary to measure the reflection picture along the entire optical path of the scheme after the state preparation device (the PM in our case). This procedure is well known in classical optics and is called optical reflectometry [42]. So-called optical time-domain reflectometry (OTDR) is carried out using a setup consisting of a laser and a synchronized photodetector. A laser pulse is sent into the optical channel to reflect and arrive at the detector inside the OTDR setup. The distance from where the pulse was reflected is calculated by the time of its arrival. This method is often used to find breaks and damage in fiber-optic links, which are usually tens of kilometers long. In the case of QKD systems, the OTDR setup should meet the following requirements: first, its spatial resolution should be no more than a few tens of centimeters, which is provided by the picosecond durations of the laser pulses; second, it should guarantee precision measurements in a sufficiently large dynamic range, which is provided by the use of sensitive single-photon detectors (the so-called ν -OTDR method [43]). Such OTDR devices operating at a wavelength of 1550 nm with ultrahigh resolution are commercially available today.

The OTDR setup for reflectivity measurement in a QKD system is depicted in Fig. 3. A polarization controller (PC) was connected to the OTDR setup, as the QKD system under test contained polarization-sensitive elements (a PM). The defense elements were extracted from the DUT to be characterized separately. This is dictated by the high attenuation level of the defense components, which significantly raises the noise floor. The PC was set to maximum forward transmittance, which was determined by an optical power meter connected to the free output of the Mach-Zehnder interferometer (MZI).

The accumulated photon counts N over the distance l within the measurement time t (5–10 min) were

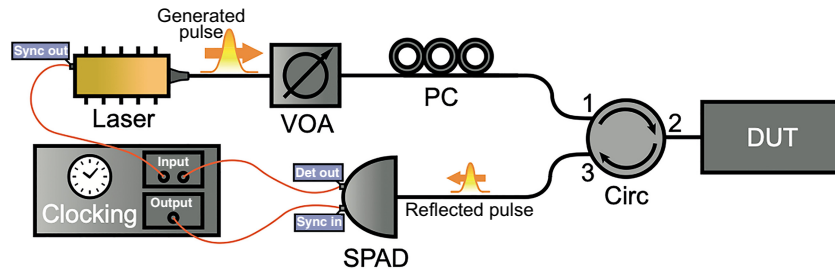


FIG. 3. OTDR setup: Laser, picosecond pulsed laser; VOA, variable optical attenuator; PC, polarization controller; Cloning, time controller; SPAD, single-photon avalanche photodiode; Circ, optical circulator; DUT, device under test.

recalculated into optical power in Watts using the following expression:

$$P(l) = \frac{N(l)hc}{t\eta_r\lambda}, \quad (13)$$

where $h \approx 6.63 \times 10^{-34} \text{ J Hz}^{-1}$ is the Planck constant, $c \approx 3 \times 10^8 \text{ m/s}$ is the speed of light in vacuum, $\eta_r \approx 10\%$ is the quantum efficiency of the single-photon avalanche diode (SPAD) inside the OTDR setup, and $\lambda = 1550 \text{ nm}$ is the wavelength under test. Finally, the R value over the distance was calculated:

$$R(l) = 10 \log_{10} \frac{P(l)}{P_{\text{in}}}, \quad (14)$$

where P_{in} is the input radiation power, which was measured preliminarily.

The OTDR results are depicted in Fig. 4. The reflections from each element of the optical scheme before the variable optical attenuator (VOA) can be clearly distinguished. Further peaks are negligibly small, as the radiation is suppressed by the VOA. The set of peaks around 7–10 m correspond to reflection from the optical connector and the VOA being tripled by the MZI.

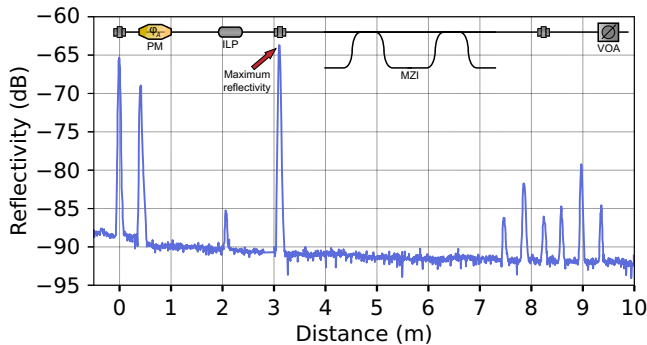


FIG. 4. The OTDR picture of the QKD system under test (Alice). The peak interpretations are as follows: PM, phase modulator; ILP, inline polarizer; MZI, Mach-Zehnder interferometer; VOA, variable optical attenuator. The reference point is set to the input optical connector.

It was revealed empirically that the OTDR precision corresponds to 1–3 dB. The flaws lower than 10 dB are, however, not crucial for the THA security analysis when sufficient protection measures are applied, as shown in Sec. IV. The maximum reflectivity corresponds to the optical connector between the inline polarizer (ILP) and the MZI. The peak value $R = -63 \text{ dB}$.

B. Spectral measurements

To characterize the transmittance of the defense components, such as optical isolators, circulators, attenuators, spectral filters, etc. (hereinafter each referred to as the device under test, or DUT), measurements across the widest possible spectral range are required. The radiation passes through the DUT and goes to the receiver. Typically, one should expect ultralow transmittance in the vicinity of the legitimate wavelength, reaching -100 dB and below. Moreover, their behavior can be unpredictable outside this spectral region; the transmittance can drop or increase to values close to 0 dB [19,26–29]. This behavior imposes quite stringent requirements for the dynamic range of measurements, i.e., the noise floor of the receiver. On the other hand, measurements across a wide spectral range require a broadband or widely tunable radiation source and a receiver (or a set of receivers) with wide spectral sensitivity. Thus, a high-power white-light source is required to ensure a wide dynamic range inside a broad spectral region.

The supercontinuum laser (SCL) stands out for its extended spectral range, which surpasses that of standard lasers, and for its superior beam quality compared to incoherent white-light sources. A set of tunable filters (for example, acousto-optic filters) should be added for spectral selection. SPADs are preferable as receivers due to their high sensitivity, especially when compared to traditional optical power meters and spectrum analyzers, which are constrained by electrical noise, limiting their sensitivity threshold to around -80 dBm or higher. In contrast, SPADs operating in gated mode with sufficiently narrow electrical gates and adequate dead time exhibit very low dark count rates (around 10 counts per second at a gate frequency of 10 MHz). This ensures precision transmittance

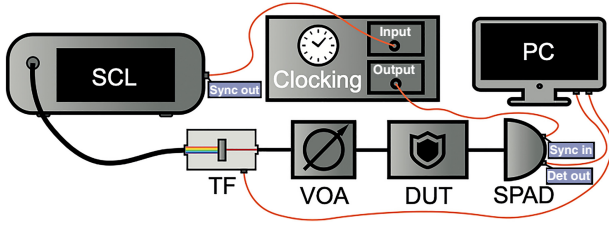


FIG. 5. Spectral measurement setup: SCL, supercontinuum laser; TF, tunable filter; VOA, variable optical attenuator; DUT, device under test; SPAD, single-photon avalanche photodiode; Clocking, time controller; PC, personal computer.

measurements across a wide dynamic range up to levels around -120 dB. For the near-infrared spectral region, an InGaAs-based SPAD is suitable, while for wavelengths of 400–1100 nm, a silicon SPAD should be used.

For the SPAD to run in photon-counting mode, it is necessary for the mean photon number per pulse μ to be much less than 1, and this is ensured by means of a variable attenuator. Otherwise, the probability of a multiple-photon count will not be negligible. This is dictated by the Poissonian statistics of the number of photons in a coherent state:

$$P_{\mu}(n=1) = \mu e^{-\mu} \approx \mu, \quad (15)$$

$$P_{\mu}(n > 1) = 1 - P_{\mu}(n=0) - P_{\mu}(n=1) \\ \approx \frac{3\mu^2}{2} \ll P_{\mu}(n=1), \quad (16)$$

where we have considered $\mu \ll 1$ to neglect the higher terms of series expansion. The experimental setup is depicted in Fig. 5.

The forward and reverse transmittance of several defense components were measured. We also recorded the reference spectrum without a DUT and the level of dark counts to express the transmittance value for each wavelength as follows:

$$T[\text{dB}] = 10 \log_{10} \frac{N - N_{\text{dark}}}{N_{\text{ref}} - N_{\text{dark}}} + A_{\text{ref}}[\text{dB}] - A[\text{dB}], \quad (17)$$

where N is the photon count for DUT spectrum measurements, N_{ref} is the reference photon count, N_{dark} is the dark count, A_{ref} is the attenuation value of the VOA for the reference spectrum, and A is the VOA attenuation value for the DUT spectrum measurements. Notably, expression (17) automatically takes spectral nonuniformity of the SCL into account, as well as the fiber transmittance and the spectral dependence of the SPAD's quantum efficiency, since their contributions to photon counts are linear and identical for both DUT and reference spectrum measurements. We applied the setup to measure the transmittance spectra of the isolator (forward and reversed), the circulator with a mirror connected to port 2 (port 3 to port 1 and port 1 to port 3), the variable attenuator, and the wavelength division multiplexer (WDM) combined with the fixed attenuator. The spectra spanned from 1100 to 1800 nm with a 1-nm step, which resulted in moving averaging of the data, as the bandwidth of the tunable filter

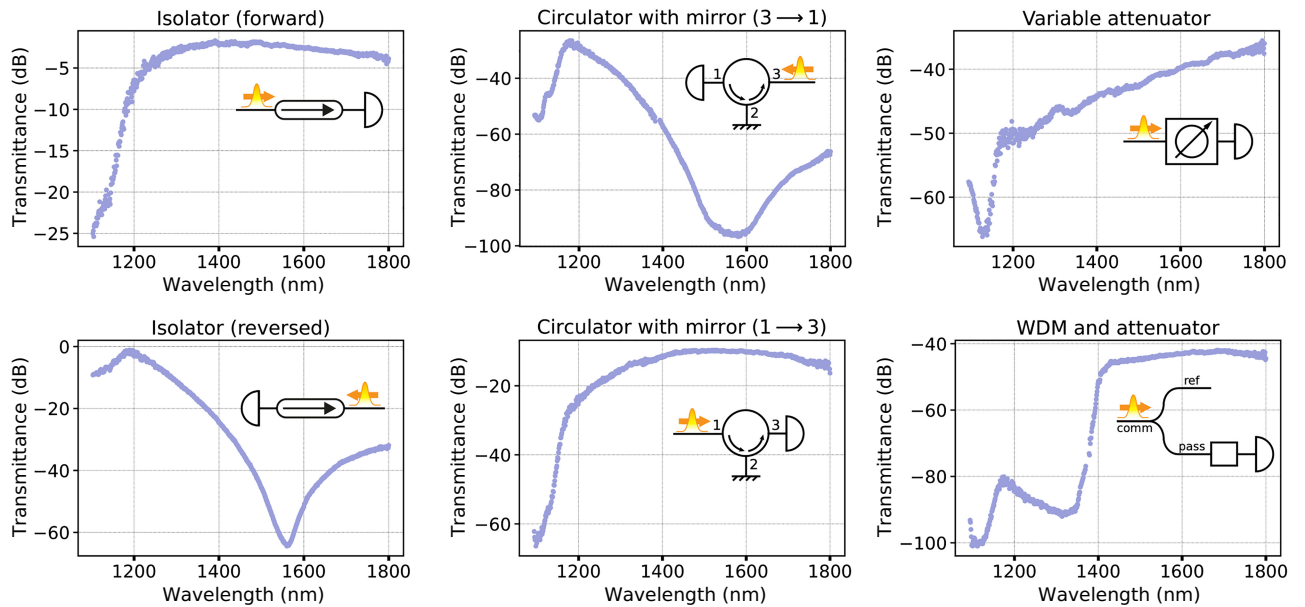


FIG. 6. Measured transmittance spectra of various optical defense components: optical isolator (forward and reverse transmittance), optical circulator with mirror on port 2 (port 1 to port 3 and port 3 to port 1 transmittance), variable attenuator, and WDM with fixed attenuator.

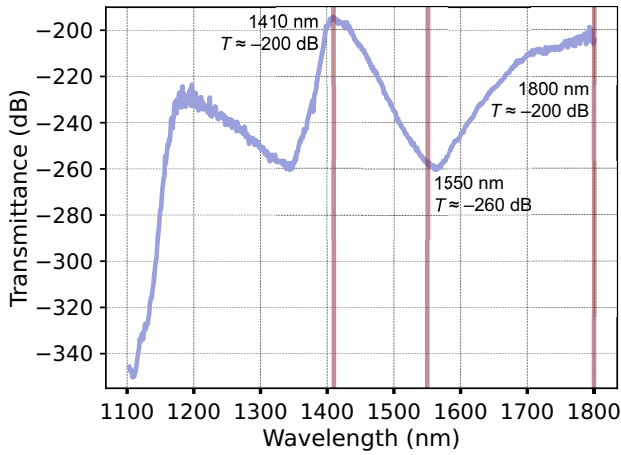


FIG. 7. Overall transmittance spectrum of the defense components inside Alice's setup.

(TF) was approximately 10 nm. The measurement results are shown in Fig. 6.

The obtained results confirm the unpredictability of the behavior of fiber-optic elements over a broad spectral range. As seen in Fig. 6, the transmission variations can reach 60 dB. It was found that the transmission spectra of the attenuators and the WDM are not dependent on the direction of radiation propagation; hence, only results from direct transmission measurements are presented here. The circulator and isolator significantly alter the transmittance level when deviating from the 1550-nm legitimate wavelength, similarly to in previously published results [27,28]. It is worth noting that low-transmittance regions in these studies were not accurately measured due to a limited dynamic range; in our research, however, the transmittance spectra of elements with losses reaching up to 100 dB were properly measured.

During OTDR measurements, we revealed that the maximum reflectivity corresponds to the connection between the ILP and the MZI; therefore, the VOA is not involved in THA mitigation. The overall transmittance spectrum is the sum of the forward and reverse isolator, circulator, fixed attenuator, and WDM transmittance spectra from Fig. 6. The result is depicted in Fig. 7.

IV. RESULTS

To estimate the value of μ_{Eve} , one needs to combine the data from Sec. III with the maximum input radiation power P_{max} . We will assume this value to be equal to the realistic LIDT value of 10 W (40 dBm), which agrees with the experimental data collected during LDA security analysis [13,14]. Applying consideration from Sec. III, we also set the attacking repetition rate $f_{\text{Eve}} = 1$ kHz, which is 4 orders of magnitude less than the communication repetition frequency $f = 10$ MHz. Thus, using (11) and (12), we obtain $\mu_{\text{Eve}} \approx 4 \times 10^{-16}$ for the legitimate wavelength

$\lambda = 1550$ nm. For such a small value, the fidelity root $\eta \approx 1$ with an accuracy of 15 decimal places, as any η bound is asymptotically linear (C5). In turn, the Holevo bound (5) is strongly dependent only on the QBER value.

However, it is not optimal for Eve to use the legitimate wavelength. As seen in Fig. 7, there are two windows of maximum transmittance near 1400 and 1800 nm, where $T \approx -200$ dB. According to (11), red-shifted light—i.e., less energetic photons—is more favorable for Eve; hence, Eve might prefer $\lambda = 1800$ nm. As we only revealed the reflectivity value for 1550 nm, we should make an assumption in favor of Eve and set $R = 0$ dB, which corresponds to total reflectance and in fact is unrealistic. As a result, we obtain $\mu_{\text{Eve}} \approx 9 \times 10^{-4}$. This is much more significant leakage than that seen at 1550 nm. For illustration, let us consider the zero-QBER case, using (5):

$$\bar{\chi}_{\text{Eve}} \approx h\left(\frac{1-\eta}{2}\right) \approx h(\mu_{\text{Eve}}) \approx 1\%. \quad (18)$$

This indicates that the impact of the THA on the key length should not be ignored; rather, one should eliminate the corresponding leakage during the privacy amplification stage. Alternatively, it is possible to add extra THA defense elements to suppress μ_{Eve} to a negligibly small level or to experimentally characterize the reflectivity outside the legitimate spectral region. The latter requires OTDR with ultrahigh resolution across a wide spectral range, which has not yet been demonstrated.

V. CONCLUSIONS

In this work, we have presented a comprehensive security analysis of Trojan-horse attacks on a real QKD system. We obtained the theoretical bounds on the information available to an adversary for both polarization- and phase-coded states. We also described the experimental setups for practical security analysis and applied them to a real QKD system. We obtained an OTDR picture for a legitimate wavelength $\lambda = 1550$ nm with centimeter resolution and an ultralow noise floor by using a single-photon avalanche photodiode as the receiver. Using a supercontinuum laser and single-photon detectors, we also measured the transmittance spectra for the THA defense elements with a record dynamic range of over 100 dB in the near-infrared spectral region. Combining our results with consideration of the maximum input power radiation, we obtained the values of the mean photon number available to Eve for the legitimate wavelength and the optimal attacking wavelength, which in our case appeared to be $\lambda = 1800$ nm. We calculated the leakage through the corresponding side channel that needs to be mitigated during the privacy amplification procedure. However, the results presented lie in the spectral region 1100–1800 nm, which is dictated by the sensitivity of our detector. Moreover, the OTDR was

carried out on the legitimate wavelength $\lambda = 1550$ nm. Unlike us, Eve is not limited in her wavelength choice. This opens up a wide area for further investigation, extending the spectral region until some physical constraints occur.

In addition, examining the side channel under varying conditions—such as thermal, magnetic, and acoustic influences on the system—could provide valuable insights. Finally, a comprehensive analysis should be conducted for other side channels, such as those associated with back-flash radiation from the detectors or radio-frequency emissions from the electronics inside the system. A complete evaluation of all side-channel information leaks, among others, will ensure the secrecy of quantum key distribution, reaffirming its status as the most secure and reliable method for transmitting confidential information.

ACKNOWLEDGMENTS

I.S.S. thanks S.N. Molotkov and A.N. Klimov for fruitful discussions.

APPENDIX A: EVE'S FIDELITY CALCULATION

The fidelity for Eve's auxiliary subsystem states can be determined as follows:

$$\mathcal{F}(\rho_Q^0, \rho_Q^1) = \left(\text{tr} \sqrt{\rho_Q^0 \rho_Q^1} \right). \quad (\text{A1})$$

Here, we used an equivalent definition of the Uhlmann fidelity [36] for commuting the density matrices. The product of the density matrices is

$$\begin{aligned} \rho_Q^0 \rho_Q^1 &= [(1-Q)|\Phi_0\rangle\langle\Phi_0| + Q|\Theta_0\rangle\langle\Theta_0|] \\ &\quad \times [(1-Q)|\Phi_1\rangle\langle\Phi_1| + Q|\Theta_1\rangle\langle\Theta_1|] \\ &= \varepsilon(1-Q)^2|\Phi_0\rangle\langle\Phi_1| + \varepsilon Q^2|\Theta_0\rangle\langle\Theta_1|. \end{aligned} \quad (\text{A2})$$

It is readily apparent that $|\Phi_0\rangle$ and $|\Theta_0\rangle$ are the eigenvectors of $\rho_Q^0 \rho_Q^1$:

$$\begin{aligned} \rho_Q^0 \rho_Q^1 |\Phi_0\rangle &= \varepsilon(1-Q)^2 |\Phi_0\rangle\langle\Phi_1| |\Phi_0\rangle \\ &= \varepsilon^2 (1-Q)^2 |\Phi_0\rangle, \end{aligned} \quad (\text{A3})$$

$$\rho_Q^0 \rho_Q^1 |\Theta_0\rangle = \varepsilon Q^2 |\Theta_0\rangle\langle\Theta_1| |\Theta_0\rangle = \varepsilon^2 Q^2 |\Theta_0\rangle. \quad (\text{A4})$$

The eigenvalues are $\varepsilon^2(1-Q)^2$ and $\varepsilon^2 Q^2$. The trace of $\sqrt{\rho_Q^0 \rho_Q^1}$ can be calculated as the sum of the square roots of these eigenvalues (the square root in the fidelity definition is assumed to be positive semidefinite, so the square roots of the eigenvalues are non-negative):

$$\begin{aligned} \sqrt{\mathcal{F}} &= \text{tr} \sqrt{\rho_Q^0 \rho_Q^1} = \sqrt{\varepsilon^2(1-Q)^2} + \sqrt{\varepsilon^2 Q^2} \\ &= \varepsilon(1-Q) + \varepsilon Q = \varepsilon. \end{aligned} \quad (\text{A5})$$

APPENDIX B: DISTINGUISHABILITY OF PURE POLARIZATION-CODED STATES

Let us construct the THA state vectors corresponding to different encoded bits in a fixed basis (we will use the Z basis, which consists of horizontal and vertical polarization states, but the calculations are equivalent for any choice of basis), considering perfect state preparation, which is the worst-case scenario for legitimate users:

$$\begin{aligned} |\xi_H\rangle &= (c_0|0\rangle_H + c_{10}|1\rangle_H + c_{20}|2\rangle_H + \dots) \otimes |0\rangle_V \\ &= c_0|\text{vac}\rangle + c_{10}|10\rangle + c_{20}|20\rangle + \dots, \end{aligned} \quad (\text{B1})$$

$$\begin{aligned} |\xi_V\rangle &= |0\rangle_H \otimes (c_0|0\rangle_V + c_{10}|1\rangle_V + c_{20}|2\rangle_V + \dots) \\ &= c_0|\text{vac}\rangle + c_{10}|01\rangle + c_{20}|02\rangle + \dots, \end{aligned} \quad (\text{B2})$$

where the vectors $|n\rangle_{H(V)}$ denote the horizontally (vertically) polarized state with n photons in a mode and $|\text{vac}\rangle$ stands for the vacuum state; c_{0k} , c_{k0} , and c_0 are the probability amplitudes for the corresponding Fock states. The scalar product between $|\xi_H\rangle$ and $|\xi_V\rangle$ can be expressed as follows:

$$\eta = \langle \xi_H | \xi_V \rangle = |c_0|^2 = p_0, \quad (\text{B3})$$

where p_0 is the probability of the vacuum component, which is equal for both states as we propose the same losses for different polarization states (without the loss of generality, as one may use μ_{Eve} as the maximum estimation for different polarization cases). This result highlights the fact that the vacuum component limits the scalar product (i.e., the indistinguishability of the states) as it does not correspond to any polarization. Using the results of the estimation of μ_{Eve} and (10), we get the lower bound for η :

$$\eta \geq 1 - \mu_{\text{Eve}}. \quad (\text{B4})$$

The example of optimal pure states that achieve the bound can be constructed as follows:

$$|\xi_H\rangle = \sqrt{1 - \mu_{\text{Eve}}}|\text{vac}\rangle + \sqrt{\mu_{\text{Eve}}}|10\rangle, \quad (\text{B5})$$

$$|\xi_V\rangle = \sqrt{1 - \mu_{\text{Eve}}}|\text{vac}\rangle + \sqrt{\mu_{\text{Eve}}}|01\rangle. \quad (\text{B6})$$

APPENDIX C: DISTINGUISHABILITY OF PURE PHASE-CODED STATES

Let us construct the THA state vectors corresponding to the phase-coding case. In this scenario, the phase modulator only twists the phases of the probability amplitudes in the Fock basis of the light state. The state vectors, in turn, are:

$$|\xi_0\rangle = c_0|\text{vac}\rangle + c_1|1\rangle + c_2|2\rangle + \dots, \quad (\text{C1})$$

$$|\xi_1\rangle = c_0|\text{vac}\rangle + c_1 e^{i\varphi_1}|1\rangle + c_2 e^{i\varphi_2}|2\rangle + \dots, \quad (\text{C2})$$

where c_k and $c_k e^{i\varphi_k}$ are the probability amplitudes of a Fock state $|k\rangle$ for the THA states corresponding to different bits.

Without loss of generality, we consider the phase of the vacuum component to be equal in both cases. For η , we have:

$$\eta = \langle \xi_0 | \xi_1 \rangle = |c_0|^2 + |c_1|^2 e^{i\varphi_1} + |c_2|^2 e^{i\varphi_2} + \dots, \quad (\text{C3})$$

which shows that η is a complex value. Once more without loss of generality, we will consider it as a real number, since it is the absolute value of the scalar product that affects distinguishability. The lower bound for η is then as follows:

$$\begin{aligned} \eta &\geq |c_0|^2 - |c_1|^2 - |c_2|^2 - \dots = p_0 - (1 - p_0) \\ &= 2p_0 - 1. \end{aligned} \quad (\text{C4})$$

The lower bound is reached when $\varphi_n = \pi$; as for the polarization case, it only depends on the probability of the vacuum component. Using (10), we obtain:

$$\eta \geq 1 - 2\mu_{\text{Eve}}. \quad (\text{C5})$$

This result makes sense for $\mu_{\text{Eve}} < 0.5$, which is suitable for our case.

For $\mu_{\text{Eve}} \ll 1$, our bound matches with the WCP case when $|\alpha\rangle$ and $|\alpha\rangle$ are applied:

$$\eta_\alpha = \langle \alpha | \alpha \rangle = e^{-2\mu_{\text{Eve}}} \approx 1 - 2\mu_{\text{Eve}} + \dots, \quad (\text{C6})$$

where $|\alpha|^2 = \mu_{\text{Eve}}$. Here, we used the famous result for the scalar product of coherent states and the Taylor-series expansion. Note that for any value of $\mu_{\text{Eve}} > 0$, our bound lies lower than η_α (see Fig. 9).

The example of optimal pure states reaching the bound can be constructed as follows:

$$|\xi_0\rangle = \sqrt{1 - \mu_{\text{Eve}}}|\text{vac}\rangle + \sqrt{\mu_{\text{Eve}}}|1\rangle, \quad (\text{C7})$$

$$|\xi_1\rangle = \sqrt{1 - \mu_{\text{Eve}}}|\text{vac}\rangle - \sqrt{\mu_{\text{Eve}}}|1\rangle. \quad (\text{C8})$$

APPENDIX D: DISTINGUISHABILITY OF MIXED STATES

To extend the results to a mixed-states case, one needs to calculate the fidelity between general THA states. This implies the calculation of Uhlmann fidelity, which seems challenging as it generally requires finding the eigenvalues in the Fock space, which is infinite-dimensional. However, the fidelity between the density matrices ρ and σ can be bounded by a simpler expression [44]:

$$\mathcal{F}(\rho, \sigma) \equiv \left(\text{tr} \sqrt{\sqrt{\sigma} \rho \sqrt{\sigma}} \right)^2 \geq \text{tr}(\rho \sigma). \quad (\text{D1})$$

We next estimate this bound for polarization and phase coding.

APPENDIX E: DISTINGUISHABILITY OF MIXED POLARIZATION-CODED STATES

The THA density matrices are constructed in a similar way to the pure-states case:

$$\rho^H = \sum_{m,n=0}^{\infty} \rho_{mn}^H |m\rangle_H \langle n| \otimes |0\rangle_V \langle 0|, \quad (\text{E1})$$

$$\rho^V = \sum_{m,n=0}^{\infty} \rho_{mn}^V |0\rangle_H \langle 0| \otimes |m\rangle_V \langle n|. \quad (\text{E2})$$

The matrix product is then

$$\rho^H \rho^V = \sum_{m,n=0}^{\infty} \rho_{m0}^H \rho_{0n}^V |m\rangle_H \langle 0| \otimes |0\rangle_V \langle n|. \quad (\text{E3})$$

Calculating the trace, we get

$$\begin{aligned} \text{tr}(\rho^H \rho^V) &= \sum_{i,j=0}^{\infty} \langle ij | \rho^H \rho^V | ij \rangle \\ &= \sum_{m,n,i,j=0}^{\infty} \rho_{m0}^H \rho_{0n}^V \langle i | m \rangle \langle 0 | i \rangle \langle j | 0 \rangle \langle n | j \rangle \\ &= \rho_{00}^H \rho_{00}^V = p_0^2. \end{aligned} \quad (\text{E4})$$

Here, we implement the fact that the probability of the vacuum component is identical for both states, as noted in this section. Finally, we obtain a lower bound on the fidelity:

$$\mathcal{F}(\rho^H, \rho^V) \geq p_0^2. \quad (\text{E5})$$

Thus, we can express the lower bound for η using μ_{Eve} :

$$\eta \geq p_0 \geq 1 - \mu_{\text{Eve}}. \quad (\text{E6})$$

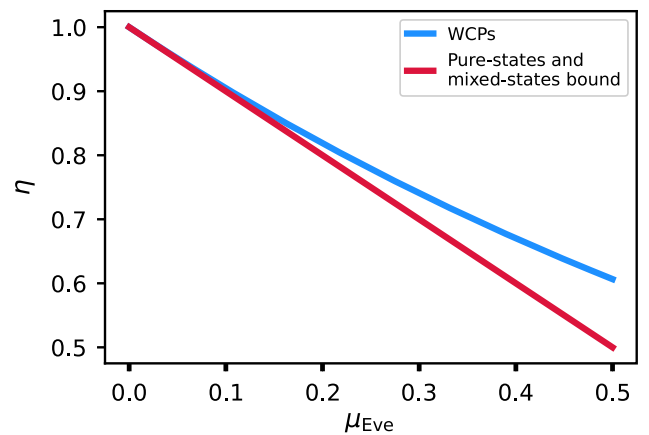


FIG. 8. Pure- and mixed-states lower bound (red) and coherent-states case (blue) for square root of fidelity between different bit polarization-coded states.

The bound is identical to the pure-states case, which is a feature of polarization coding. The dependencies for our case and the WCP case are depicted in Fig. 8.

APPENDIX F: DISTINGUISHABILITY OF MIXED PHASE-CODED STATES

For the phase coding, we have

$$\rho^0 = \sum_{m,n=0}^{\infty} \rho_{mn} |m\rangle \langle n|, \quad (\text{F1})$$

$$\begin{aligned} \rho^1 &= \sum_{m,n=0}^{\infty} \rho_{mn} e^{i(\varphi_m - \varphi_n)} |m\rangle \langle n| \\ &= \sum_{m,n=0}^{\infty} \rho_{mn} e^{i\Delta\varphi_{mn}} |m\rangle \langle n|, \end{aligned} \quad (\text{F2})$$

where $\Delta\varphi_{mn} = \varphi_m - \varphi_n$. The matrix product is then

$$\begin{aligned} \rho^0 \rho^1 &= \sum_{k,m,n=0}^{\infty} \rho_{mk} \rho_{kn} e^{i\Delta\varphi_{kn}} |m\rangle \langle n| \\ &= \sum_{k,m,n=0}^{\infty} \rho_{mk} \rho_{nk}^* e^{-i\Delta\varphi_{nk}} |m\rangle \langle n|. \end{aligned} \quad (\text{F3})$$

Calculating the trace, we get

$$\begin{aligned} \text{tr}(\rho^0 \rho^1) &= \sum_{j=0}^{\infty} \langle j | \rho^0 \rho^1 | j \rangle = \sum_{j,k=0}^{\infty} \rho_{jk} \rho_{jk}^* e^{-i\Delta\varphi_{jk}} \\ &= \frac{1}{2} \sum_{j,k=0}^{\infty} |\rho_{jk}|^2 e^{-i\Delta\varphi_{jk}} + \frac{1}{2} \sum_{j,k=0}^{\infty} |\rho_{jk}|^2 e^{i\Delta\varphi_{jk}} \\ &= \sum_{j,k=0}^{\infty} |\rho_{jk}|^2 \cos \Delta\varphi_{jk} \\ &= \sum_{k=0}^{\infty} |\rho_{kk}|^2 + \sum_{j \neq k}^{\infty} |\rho_{jk}|^2 \cos \Delta\varphi_{jk} \\ &\geq \sum_{k=0}^{\infty} |\rho_{kk}|^2 - \sum_{j \neq k}^{\infty} |\rho_{jk}|^2, \end{aligned} \quad (\text{F4})$$

where we set $\Delta\varphi_{mn} = \pi$ for nondiagonal elements in the bounding expression. The values of nondiagonal elements cannot be estimated by means of μ_{Eve} , as they contribute to the coherence of the radiation rather than its intensity; however, they appear in the purity value \mathcal{P} , which

characterizes how much a quantum state is mixed:

$$\mathcal{P} \equiv \text{tr} \rho^2 = \sum_{j,k=0}^{\infty} |\rho_{jk}|^2 = \sum_{k=0}^{\infty} |\rho_{kk}|^2 + \sum_{j \neq k}^{\infty} |\rho_{jk}|^2. \quad (\text{F5})$$

The value of the purity satisfies $0 \leq \mathcal{P} \leq 1$ [36]. Expressing the sum containing nondiagonal elements via \mathcal{P} , we obtain

$$\begin{aligned} \mathcal{F}(\rho^H, \rho^V) &\geq \text{tr}(\rho^0 \rho^1) \geq 2 \sum_{k=0}^{\infty} |\rho_{kk}|^2 - \\ &\geq 2\rho_{00}^2 - \mathcal{P} \geq 2p_o^2 - 1. \end{aligned} \quad (\text{F6})$$

We set $\mathcal{P} = 1$ for the bounding, which corresponds to the pure-states case. The bound, however, is strictly lower than that obtained for pure states when $\mu_{\text{Eve}} > 0$. This indicates that this bound seems to be unreachable, unlike the previous ones, for which the particular states were constructed. Finally, we can express the lower bound for η using μ_{Eve} :

$$\eta \geq \sqrt{2p_o^2 - 1} = \sqrt{2(1 - \mu_{\text{Eve}})^2 - 1}. \quad (\text{F7})$$

The bound (F7) also asymptotically coincides with the pure-state linear case (C5) when $\mu_{\text{Eve}} \ll 1$ (see Fig. 9). However, the difference is visible when μ_{Eve} approaches 0.3 photons per pulse.

Our bounds are conservative, unlike those presented in [24], as we consider the worst-case scenario when the bases are known to Eve and our only assumption regarding the state evolution is that the phase modulation preserves the absolute values of the probability amplitudes in the Fock basis.

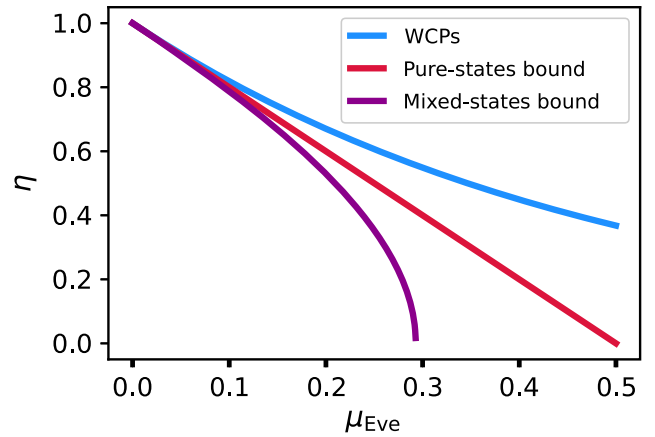


FIG. 9. Pure-states (red) and mixed-states (purple) lower bounds and coherent states case (blue) for square root of fidelity between different bit phase-coded states.

- [1] C. H. Bennett and G. Brassard, in *International Conference on Computers, Systems & Signal Processing* (IEEE, Bangalore, 1984), Vol. 175.
- [2] M. Tomamichel and R. Renner, Uncertainty relation for smooth entropies, *Phys. Rev. Lett.* **106**, 110506 (2011).
- [3] S. N. Molotkov, Entropic uncertainty relations and the extremely allowable critical error in quantum cryptography, *JETP Lett.* **94**, 820 (2012).
- [4] W.-Y. Hwang, Quantum key distribution with high loss: Toward global secure communication, *Phys. Rev. Lett.* **91**, 057901 (2003).
- [5] S. P. Kulik and S. N. Molotkov, Decoy state method for quantum cryptography based on phase coding into faint laser pulses, *Laser Phys. Lett.* **14**, 125205 (2017).
- [6] F.-Y. Lu, P. Ye, Z.-H. Wang, S. Wang, Z.-Q. Yin, R. Wang, X.-J. Huang, W. Chen, D.-Y. He, G.-J. Fan-Yuan, *et al.*, Hacking measurement-device-independent quantum key distribution, *Optica* **10**, 520 (2023).
- [7] F.-Y. Lu, Z.-H. Wang, Z.-Q. Yin, S. Wang, R. Wang, G.-J. Fan-Yuan, X.-J. Huang, D.-Y. He, W. Chen, Z. Zhou, *et al.*, Unbalanced-basis-misalignment-tolerant measurement-device-independent quantum key distribution, *Optica* **9**, 886 (2022).
- [8] S. Sauge, L. Lydersen, A. Anisimov, J. Skaar, and V. Makarov, Controlling an actively-quenched single photon detector with bright light, *Opt. Express* **19**, 23590 (2011).
- [9] D. S. Bulavkin, I. S. Sushchev, K. E. Bugai, S. A. Bogdanov, and D. A. Dvoretzky, in *Quantum and Nonlinear Optics IX*, edited by Q. He, C.-F. Li, and D.-S. Kim (SPIE, Beijing, 2023), p. 34.
- [10] C. Wiechers, L. Lydersen, C. Wittmann, D. Elser, J. Skaar, C. Marquardt, V. Makarov, and G. Leuchs, After-gate attack on a quantum cryptosystem, *New J. Phys.* **13**, 013043 (2011).
- [11] V. Makarov, A. Anisimov, and J. Skaar, Effects of detector efficiency mismatch on security of quantum cryptosystems, *Phys. Rev. A (College Park)* **74**, 022313 (2006).
- [12] A. N. Bugge, S. Sauge, A. M. M. Ghazali, J. Skaar, L. Lydersen, and V. Makarov, Laser damage helps the eavesdropper in quantum cryptography, *Phys. Rev. Lett.* **112**, 070503 (2014).
- [13] S. V. Alferov, K. E. Bugai, and I. A. Pargachev, Study of the vulnerability of neutral optical filters used in quantum key distribution systems against laser damage attack, *JETP Lett.* **116**, 123 (2022).
- [14] K. E. Bugai K, A. P. Zyzykin A, D. S. Bulavkin, S. A. Bogdanov, I. S. Sushchev, and D. A. Dvoretzky, in *2022 International Conference Laser Optics (ICLO)* (IEEE, Saint Petersburg, 2022), p. 1.
- [15] S. V. Alferov, K. E. Bugai, I. A. Pargachev, and Yu. V. Ivanova, Studying vulnerability in quantum-key-distribution systems to attacks with laser damage to optical components based on a collapsing mirror device, *Tech. Phys. Lett.* **49**, 11 (2023).
- [16] A. Huang, Á. Navarrete, S.-H. Sun, P. Chaiwongkhot, M. Curty, and V. Makarov, Laser-seeding attack in quantum key distribution, *Phys. Rev. Appl.* **12**, 064043 (2019).
- [17] S. N. Molotkov, Trojan horse attacks, decoy state method, and side channels of information leakage in quantum cryptography, *J. Exp. Theor. Phys.* **130**, 809 (2020).
- [18] S. Sajeed, C. Minshull, N. Jain, and V. Makarov, Invisible trojan-horse attack, *Sci. Rep.* **7**, 8403 (2017).
- [19] I. S. Sushchev, D. M. Guzairova, A. N. Klimov, D. A. Dvoretzky, S. A. Bogdanov, K. D. Bondar, and A. P. Naumenko, in *Emerging Imaging and Sensing Technologies for Security and Defence VI*, edited by R. C. Hollins, G. S. Buller, R. A. Lamb, and M. Laurenzis (SPIE, Madrid, 2021), p. 15.
- [20] S. A. Bogdanov, I. S. Sushchev, A. N. Klimov, K. E. Bugay, D. S. Bulavkin, and D. A. Dvoretzky, in *Quantum Technologies 2022*, edited by S. Ducci, E. Diamanti, N. Treps, and S. Whitlock (SPIE, Strasbourg, 2022), p. 57.
- [21] A. Lamas-Linares and C. Kurtsiefer, Breaking a quantum key distribution system through a timing side channel, *Opt. Express* **15**, 9388 (2007).
- [22] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, Trojan-horse attacks on quantum-key-distribution systems, *Phys. Rev. A (College Park)* **73**, 022320 (2006).
- [23] N. Jain, E. Anisimova, I. Khan, V. Makarov, C. Marquardt, and G. Leuchs, Trojan-horse attacks threaten the security of practical quantum cryptography, *New J. Phys.* **16**, 123030 (2014).
- [24] S. E. Vinay and P. Kok, Extended analysis of the trojan-horse attack in quantum key distribution, *Phys. Rev. A (College Park)* **97**, 042335 (2018).
- [25] M. Lucamarini, I. Choi, M. B. Ward, J. F. Dynes, Z. L. Yuan, and A. J. Shields, Practical security bounds against the trojan-horse attack in quantum key distribution, *Phys. Rev. X* **5**, 031030 (2015).
- [26] B. A. Nasedkin, I. M. Filipov, A. O. Ismagilov, V. V. Chistiakov, F. D. Kiselev, A. N. Tsyppkin, and V. I. Egorov, Analyzing transmission spectra of fiber-optic elements in the near IR range to improve the security of quantum key distribution systems, *Bull. Russ. Acad. Sci.: Phys.* **86**, 1164 (2022).
- [27] 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).
- [28] A. V. Borisova, B. D. Garmaev, I. B. Bobrov, S. S. Negodyaev, and I. V. Sinil'shchikov, Risk analysis of countermeasures against the trojan-horse attacks on quantum key distribution systems in 1260–1650 nm spectral range, *Opt. Spectrosc.* **128**, 1892 (2020).
- [29] B. Nasedkin, F. Kiselev, I. Filipov, D. Tolochko, A. Ismagilov, V. Chistiakov, A. Gaidash, A. Tsyppkin, A. Kozubov, and V. Egorov, Loopholes in the 1500–2100-nm range for quantum-key-distribution components: Prospects for trojan-horse attacks, *Phys. Rev. Appl.* **20**, 014038 (2023).
- [30] S. Sajeed, I. Radchenko, S. Kaiser, J.-P. Bourgoin, L. Monat, M. Legré, and V. Makarov, Securing two-way quantum communication: The monitoring detector and its flaws, in *4th International Conference on Quantum Cryptography (QCrypt)*, Paris, France, 2014 (unpublished), <http://www.vad1.com/publications/sajeed2014.QCryp2014.pdf>
- [31] F.-Y. Lu, Z.-H. Wang, V. Zapatero, J.-L. Chen, S. Wang, Z.-Q. Yin, M. Curty, D.-Y. He, R. Wang, W. Chen, *et al.*, Experimental demonstration of fully passive quantum key distribution, *Phys. Rev. Lett.* **131**, 110802 (2023).

- [32] C. H. Bennett, G. Brassard, C. Crepeau, and U. M. Maurer, Generalized privacy amplification, *IEEE Trans. Inf. Theory* **41**, 1915 (1995).
- [33] A. S. Holevo, Bounds for the quantity of information transmitted by a quantum communication channel, *Probl. Peredachi Inf.* **9**, 3 (1973).
- [34] A. Uhlmann, The “transition probability” in the state space of a $*$ -algebra, *Rep. Math. Phys.* **9**, 273 (1976).
- [35] W. Roga, M. Fannes, and K. Życzkowski, Universal bounds for the Holevo quantity, coherent information, and the Jensen-Shannon divergence, *Phys. Rev. Lett.* **105**, 040505 (2010).
- [36] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, 2016).
- [37] A. N. Klimov, K. A. Balygin, and S. N. Molotkov, Two-parameter single-pass plug and play quantum cryptography without adjustment of states in the quantum channel, *Laser Phys. Lett.* **15**, 075207 (2018).
- [38] M. Kihara, S. Nagasawa, and T. Tanifuji, Return loss characteristics of optical fiber connectors, *J. Lightwave Technol.* **14**, 1986 (1996).
- [39] A. Muller, T. Herzog, B. Huttner, W. Tittel, H. Zbinden, and N. Gisin, “Plug and play” systems for quantum cryptography, *Appl. Phys. Lett.* **70**, 793 (1997).
- [40] J. Wang, X. Qin, Y. Jiang, X. Wang, L. Chen, F. Zhao, Z. Wei, and Z. Zhang, Experimental demonstration of polarization encoding quantum key distribution system based on intrinsically stable polarization-modulated units, *Opt. Express* **24**, 8302 (2016).
- [41] I. Lucio-Martinez, P. Chan, X. Mo, S. Hosier, and W. Tittel, Proof-of-concept of real-world quantum key distribution with quantum frames, *New J. Phys.* **11**, 095001 (2009).
- [42] M. K. Barnoski, M. D. Rourke, S. M. Jensen, and R. T. Melville, Optical time domain reflectometer, *Appl. Opt.* **16**, 2375 (1977).
- [43] P. Healey and P. Hensel, Optical time domain reflectometry by photon counting, *Electron. Lett.* **16**, 631 (1980).
- [44] J. L. A. Miszczak, Z. P. La, P. L. Horodecki, A. Uhlmann, and K. Życzkowski, Sub- and super-fidelity as bounds for quantum fidelity, *Quantum Inf. Comput.* **9**, 0103 (2009).