FRONTIERS OF PHYSICS

RESEARCH ARTICLE
Volume 21 / Issue 12 / 123202 / 2026

# Security analysis of Trojan horse attacks on single-photon based quantum secure direct communication systems

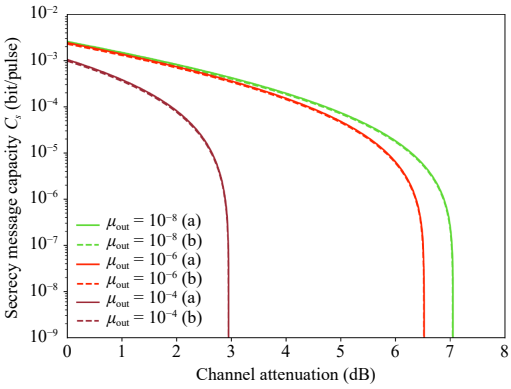
Jin-Yu Wang[1], Jia-Wei Ying[1], Qi Zhang[2], Shi-Pu Gu[1], Xing-Fu Wang[2], Lan Zhou[3], Yu-Bo Sheng[1,3,†]

1 College of Electronic and Optical Engineering, & College of Flexible Electronics (Future Technology), Nanjing University of Posts and Telecommunications, Nanjing 210023, China
2 College of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
3 School of Physics, Hangzhou Normal University, Hangzhou 311121, China
Corresponding author. E-mail: †shengyb@njupt.edu.cn
Received March 3, 2026; accepted March 31, 2026



## ABSTRACT

Quantum secure direct communication (QSDC) enables the direct transmission of confidential information without pre-shared keys. However, device imperfections in practical systems can lead to information leakage. The Trojan-horse attack (THA), a stealthy side-channel threat, steals information by injecting photons into encoding devices and analyzing the reflected light. This paper focuses on the single-photon-based (DL04) protocol, establishing for the first time a comprehensive THA model against it. By combining weak coherent pulse (WCP) sources with the decoy-state method, we systematically analyze the security under THA during both the first and second rounds of transmission, derive analytical expressions for the secrecy message capacity, and quantitatively evaluate the impact of attack strength on system performance through numerical simulations. Our results demonstrate that the system's security is highly sensitive to the average number of reflected Trojan photons, $\mu_{\text{out}}$. For a first-round attack, the secrecy message capacity remains nearly identical to the attack-free case when $\mu_{\text{out}} \leq 10^{-8}$, but drops to zero when $\mu_{\text{out}}$ increases to $10^{-3}$. In contrast, the second round of attacks had a weaker impact. Even at the strongest attack intensity ($\mu_{\text{out}} = 10^{-2}$), the system could still maintain a certain level of secrecy message capacity. When THA simultaneously targets both the first and second rounds, the overall security of the system is predominantly governed by the more sensitive first-round attack, whose security boundary remains largely comparable to that observed in a single first-round attack scenario. Furthermore, We perform parameter optimization to achieve the optimal secrecy message capacity for each channel attenuation. Our study provides the first quantitative security bound for THA on DL04 protocol, offering practical guidance for encoder isolation requirements in future implementations.




**Keywords** quantum secure direct communication, Trojan horse attack, decoy-state

Higher Education Press

https://doi.org/10.15302/frontphys.2026.123202

## 1 Introduction

Quantum secure direct communication (QSDC) constitutes a vital branch of quantum communication, enabling the direct transmission of confidential information without relying on pre-shared keys or complex key management infrastructures. The concept was originally introduced by Long *et al.* [1], and its foundational framework was further refined by Deng *et al.* [2]. Over the subsequent decades, QSDC has witnessed substantial theoretical progress, as evidenced by a series of influential works [3–20]. In addition, considerable developments have been made across several subfields, such as continuous-variable protocols [21–23], quantum error-correcting codes [24, 25], quantum networks [26–29], and quantum-memory-free protocols [30]. Experimentally, the field has also achieved significant milestones, including the realization of single-photon-based [31–34], entanglement-based [35, 36], and continuous-variable QSDC systems [37–39], along with advances in quantum network implementations [40, 41].

Although QSDC is theoretically absolutely secure, information leaks may occur in practical applications due to imperfect devices. Imperfect light sources and detectors, in particular, introduce vulnerabilities that eavesdroppers can exploit. In response, a range of robust protocols has been developed. Notable among these are measurement-device-independent (MDI) QSDC [42–50] and device-independent (DI) QSDC [51–55]. The former removes detector-side vulnerabilities by relaying measurements to an untrusted third party, while the latter certifies security through the violation of Bell's inequality, thus ensuring robustness against general eavesdropping strategies. In theory, DI protocols are capable of resisting all side-channel attacks targeting implementation devices. However, such protocols currently face severe limitations in photon transmission loss and transmission distance [56, 57]. Moreover, experimental realizations of device-independent quantum key distribution (DI-QKD) were not achieved until 2022 [58–60], and experimental DI-QSDC has yet to be realized. Therefore, considering the practical complexity involved, the QSDC protocols that are currently approaching practical application are predominantly grounded in single-photon-based (DL04) protocol.

Currently, the DL04 protocol has been proven secure against photon-number-splitting (PNS) attacks and collective attacks [61–63]. However, to date, no quantitative analysis has been conducted on the security of this protocol against optical injection attacks, such as Trojan-horse attack (THA) triggered by device reflections. During the THA process, an eavesdropper injects concealed photons into the transmitting device. These photons interact with internal optical components—such as modulators or phase shifters—and are subsequently reflected out of the system. By collecting and analyzing these reflected photons, the eavesdropper can extract confidential information without triggering conventional security monitoring. Such attacks often evade detection by mimicking legitimate signal patterns, employing techniques such as collecting back-reflected light, manipulating source characteristics, or exploiting timing vulnerabilities [64–69]. In addition, THA also poses a serious threat to multi-party quantum communication, including measurement-device-independent quantum key distribution (MDI-QKD) [70], quantum secure sharing (QSS) [71], quantum conference key agreement (QCKA) [72] and quantum cryptographic conferencing (QCC) [73].

To investigate the impact of THA on QSDC systems, this paper conducted a comprehensive analysis, covering attack principles, security analysis using a weak coherent pulse (WCP) source combined with the decoy-state method, and numerical simulations. Our work aims to: (i) establish a THA model targeting the DL04 QSDC protocol; (ii) derive the secrecy message capacity under THA and extend the formulation to incorporate decoy-state analysis; and (iii) quantitatively evaluate the relationship between attack intensity and system performance through numerical simulations.

This paper is structured as follows. In Section 2, we begin with a concise overview of the DL04 protocol steps, along with Eve's attack strategy. Subsequently, a security analysis is conducted for the first and second rounds of transmission under THA, leading to the derivation of secrecy message capacity expressions for both rounds under such attacks. Furthermore, the analysis is extended by incorporating decoy-state techniques, establishing the estimated single-photon error rate and single-photon yield. In Section 3, we perform numerical simulations to examine how THA affects the secrecy message capacity and the maximum communication distance. In Section 4, we provide a discussion of the results and summarize the conclusions.

## 2 Trojan horse attack on DL04 protocol

### 2.1 DL04 protocol

This paper presents an analysis of a THA targeting the DL04 protocol, and the corresponding system structure is illustrated in Fig. 1. We begin by outlining the DL04 protocol, which involves two parties — the sender Bob and the receiver Alice — and consists of five steps in total.

(i) Bob performs random phase encoding on the initial single-photon state $|+\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)$, where $|H\rangle$ and $|V\rangle$ denote the horizontal and vertical polarization states, respectively. Subsequently, this state is processed by the encoding device, which generates four single-photon states $\{|+\rangle, |-\rangle, |R\rangle, |L\rangle\}$, corresponding to $\frac{1}{\sqrt{2}}(|H\rangle + e^{i\alpha_B}|V\rangle)$ with phase values $\alpha_B = \{0, \pi,$

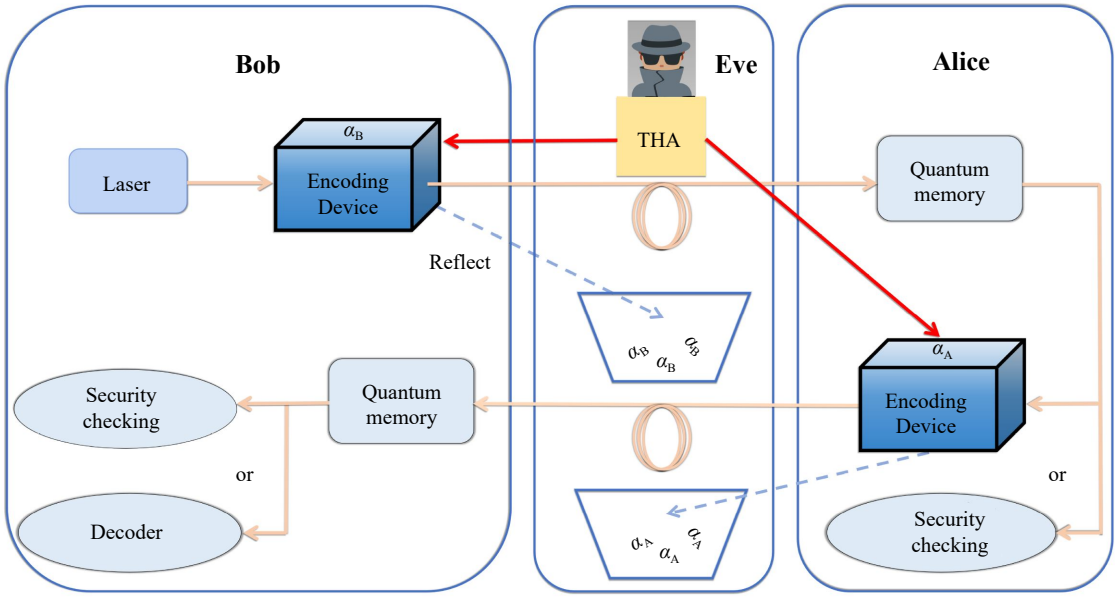



**Fig. 1** Schematic of a THA during the first and second round transmission phase of a QSDC system. The label THA refers both to the attack strategy and to Eve's independent Trojan-horse light source. The rectangles labeled $\alpha_B$ and $\alpha_A$ represent the encoding modules at Bob's and Alice's ends, respectively. Eve injects coherent optical pulses into these modules; in the first and second rounds, portions of the pulses are encoded by Bob and Alice, then reflected and collected by Eve. Red arrows: Trojan photon pulses injected by Eve. Blue dashed arrows: Reflected pulses carrying modulation information. Light-colored lines: Quantum and classical channels among Alice, Bob, and Eve.

$\pi/2, 3\pi/2\}$. The $X$ basis comprises the states $\left\{|+\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle), |-\rangle = \frac{1}{\sqrt{2}}(|H\rangle - |V\rangle)\right\}$; the $Y$ basis is defined as $\left\{|R\rangle = \frac{1}{\sqrt{2}}(|H\rangle + \mathrm{i}|V\rangle), |L\rangle = \frac{1}{\sqrt{2}}(|H\rangle - \mathrm{i}|V\rangle)\right\}$. Then, Bob sends these encoded single photons to Alice.

(ii) Upon receiving the photons, Alice randomly selects a subset of them as samples for an eavesdropping check. For each sampled photon, she randomly selects a measurement basis ($X$ or $Y$), performs the measurement, and then announces the photon's position, her chosen basis, and the measurement outcome. Using an authenticated classical channel, Alice and Bob compare their results to estimate the detection bit error rate (DBER).

(iii) If the DBER remains below a pre-agreed threshold, Alice proceeds with the following steps. First, she randomly selects a subset of the received photons for random encoding, which is a process used for error estimation. Subsequently, she performs information encoding on the remaining photons. Specifically, Alice applies a operation $U = |H\rangle\langle H| + \mathrm{e}^{\mathrm{i}\alpha_A}|V\rangle\langle V|$ to $\{|+\rangle, |-\rangle, |R\rangle, |L\rangle\}$, where the phase value $\alpha_A$ is chosen from the set $\{0, \pi\}$ to encode a secret information bit: $\alpha_A = 0$ corresponds to bit 0, and $\alpha_A = \pi$ corresponds to bit 1. Then, Alice returns the encoded sequence of photons to Bob.

(iv) Bob receives and stores the photon sequence. Subsequently, Alice announces the position of the randomly encoded photon within the sequence. Bob then randomly selects either the $X$ basis or the $Y$ basis for measurement. After the measurement succeeds, Alice announces the specific details of the random encoding performed in Step 3. Then, Bob extracts these designated photons to perform the second round of security testing and compute the quantum bit error rate (QBER).

(v) Upon successful completion of the second security check, Bob infers Alice's encoded information based on his initial state information and the decoded information.

## 2.2 Attack strategy

As shown in Fig. 1, for Eve, she has two opportunities to execute THA, including attacking on both the first-round transmission and the second-round transmission.

During the first-round transmission, Eve can execute the THA. Specifically, during Step 1 of the protocol, she injects a bright pulse of Trojan photons ($\mu_{\mathrm{in}}$, prepared as coherent states $\left|\sqrt{\mu_{\mathrm{in}}}\right\rangle$) into Bob's encoding device to probe the secret phase $\alpha_B$. These externally injected photons co-propagate with Bob's own single-photon signal and experience the same phase modulation. Due to the reflectivity of the encoding device, a fraction of the modulated Trojan photons ($\left|\mathrm{e}^{\mathrm{i}\alpha_B}\sqrt{\mu_{\mathrm{out}}}\right\rangle$, where $\mu_{\mathrm{out}} = \beta\mu_{\mathrm{in}}$ and $\beta \ll 1$ quantifies the optical isolation of the encoding device) is reflected out of the encoder. By collecting and analyzing the phase of these reflected photons, Eve can influence the results of basis selection measurements, causing Alice's basis selection measure-

ments to deviate, thereby affecting the error rate and leading Alice to underestimate the error rate. By combining collective attacks, Eve can obtain more information.

During the second-round transmission, Eve executes THA by sending a large number of Trojan photons to Alice's encoding device in the third step. After undergoing encoding operations, some Trojan photons are similarly reflected outward due to the reflection mechanism. Eve gathers these reflected photons $e^{i\alpha_A}$ (when $\alpha_A$ is 0 or $\pi$, it corresponds to the values +1 and –1, respectively). Therefore, Eve can use appropriate measurements to distinguish between the resulting coherent states $|+\sqrt{\mu_{out}}\rangle$ and $|-\sqrt{\mu_{out}}\rangle$. This allows her to gain access to the encoded information.

### 2.3 Security analysis

In this section, we analyze the secrecy message capacity of the protocol under Eve's attacks during the first and second rounds of transmission.

#### 2.3.1 First-round transmission

In the first round, Eve inputs Trojan photons $|+\sqrt{\mu_{in}}\rangle$ into Bob's encoding device. Then, these photons undergo the same phase encoding as the initial state $|+\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)$. After performing the encoding operation, the encoded Trojan photons become coupled with the four single-photon states generated by Bob's encoding, the portion of reflected photons collected by Eve can be expressed as tensor products:

$$\begin{aligned}
|\psi_+\rangle_{BE} &= |+\rangle_B \otimes |+\sqrt{\mu_{out}}\rangle_E, \\
|\psi_-\rangle_{BE} &= |-\rangle_B \otimes |-\sqrt{\mu_{out}}\rangle_E, \\
|\psi_R\rangle_{BE} &= |R\rangle_B \otimes |+i\sqrt{\mu_{out}}\rangle_E, \\
|\psi_L\rangle_{BE} &= |L\rangle_B \otimes |-i\sqrt{\mu_{out}}\rangle_E,
\end{aligned} \tag{1}$$

where $\mu_{out}$ is the average number of reflected photons from Bob's encoding device. The subscript BE indicates that the quantum state belongs to Bob and Eve.

First, the original DL04 protocol secrecy message capacity formula can be expressed as [8]

$$\begin{aligned}
C_s = {} & Q_\mu^{BAB}\left[1 - h\left(E_\mu^{BAB}\right)\right] - Q_{\mu,n=1}^{BAE} h\left(e_{x,1}^{BA} + e_{y,1}^{BA}\right) \\
& - Q_{\mu,n=2}^{BAE}\left[\frac{1}{2}h\left(e_{x,2}^{BA} + e_{y,2}^{BA}\right) + \frac{1}{2}\right] - Q_{\mu,n\geq 3}^{BAE} \times 1,
\end{aligned} \tag{2}$$

where $Q_\mu^{BAB}$ is the overall signal gain of Bob after a round-trip BAB, and $E_\mu^{BAB}$ is the QBER. $Q_{\mu,n}^{BAE}$ denotes the gain of $n$-photon events from Eve. $e_{x,1}^{BA}$ and $e_{y,1}^{BA}$ are the single-photon error rates under the $X$ basis and $Y$ basis, respectively. Similarly, $e_{x,2}^{BA}$ and $e_{y,2}^{BA}$ are the two photon error rates under the $X$ basis and $Y$ basis, respectively. $h(x)$ is the binary entropy function, which is defined as $h(x) = -x\log_2 x - (1-x)\log_2(1-x)$.

Additionally, since the two-photon component contributes less to the protocol, we cite the simplified formula in Ref. [9] to analyze the secrecy message capacity of the protocol, which can be represented as follows:

$$\begin{aligned}
C_s = {} & Q_\mu^{BAB}\left[1 - h\left(E_\mu^{BAB}\right)\right] - Q_{\mu,n=1}^{BAE} h\left(e_{x,1}^{BA} + e_{y,1}^{BA}\right) \\
& - Q_{\mu,n\geq 2}^{BAE} \times 1.
\end{aligned} \tag{3}$$

Since THA causes measurement inaccuracies in the protocol, it leads to deviations in Alice's basis selection, thereby affecting the single-photon error rate. Therefore, the secrecy message capacity formula for the DL04 protocol after suffering a THA is

$$\begin{aligned}
C_s = {} & Q_\mu^{BAB}\left[1 - h\left(E_\mu^{BAB}\right)\right] - Q_{\mu,n=1}^{BAE} h\left(e_{x,1}^{BA'} + e_{y,1}^{BA'}\right) \\
& - Q_{\mu,n\geq 2}^{BAE} \times 1.
\end{aligned} \tag{4}$$

The term $e_{x,1}^{BA'}$ ($e_{y,1}^{BA'}$) is the single-photon error rate estimated in a virtual protocol, where Alice measures in the $Y(X)$ basis and Bob announces the $X(Y)$ basis, defined as [76]

$$\begin{aligned}
e_{x,1(y,1)}^{BA'} = {} & e_{y,1(x,1)}^{BA} + 4\Delta'(1-\Delta')(1-2e_{y,1(x,1)}^{BA}) \\
& + 4(1-2\Delta')\sqrt{\Delta'(1-\Delta')e_{y,1(x,1)}^{BA}(1-e_{y,1(x,1)}^{BA})}, \\
\Delta' = {} & \frac{\Delta}{y},
\end{aligned} \tag{5}$$

where $y = \min[y_x, y_y]$, $y_x$ and $y_y$ are the single-photon yields in the $X$ and $Y$ basis, respectively. In practice, Bob selects either the $X$ basis or $Y$ basis with equal probability when encoding single photons. When subjected to a THA at Bob's end, this equilibrium is disrupted. Therefore, we utilize the "quantum coin" $\Delta$ [76] to quantify the severity of the attack. Since quantum coins $\Delta$ are affected by the actual yield obtained from the $X$ or $Y$ basis, the communicating parties must accordingly renormalize $\Delta$ to $\Delta'$, referred to as the effective coin imbalance (see Appendix B for details).

#### 2.3.2 Second-round transmission

In the second transmission round, Eve inputs Trojan photons $|+\sqrt{\mu_{in}}\rangle$ into Alice's encoding device. Then, Alice performs encoding operations, which involve applying an operation $U = |H\rangle\langle H| + e^{i\alpha_A}|V\rangle\langle V|$ to $\{|+\rangle, |-\rangle, |R\rangle, |L\rangle\}$. When $\alpha_A$ is 0, the phase remains unchanged, equivalent to an $I = |0\rangle\langle 0| + |1\rangle\langle 1|$ operation; when $\alpha_A$ is $\pi$, the phase is inverted, equivalent to a $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$ operation. Therefore, the output from Alice's encoding device can be represented as

$$
\begin{aligned}
&I\otimes|+\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}=|+\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\quad Z\otimes|+\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}=|-\rangle_{\mathrm{A}}|-\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\\
&I\otimes|-\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}=|-\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\quad Z\otimes|-\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}=|+\rangle_{\mathrm{A}}|-\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\\
&I\otimes|R\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}=|R\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\quad Z\otimes|R\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}=|L\rangle_{\mathrm{A}}|-\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\\
&I\otimes|L\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}=|L\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\quad Z\otimes|L\rangle_{\mathrm{A}}|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}=|R\rangle_{\mathrm{A}}|-\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}.
\end{aligned}
\tag{6}
$$

The subscripts A and E indicate that the reflected photons belong to Alice and Eve, respectively. After Eve performs the THA operation in the second round, since Alice's encoding operation is state-independent, Eve's system becomes independent from both Alice's and Bob's systems. This means Eve's attack will not cause any errors. In this scenario, Eve's eavesdropping remains undetected by Alice and Bob, and she merely needs to discriminate between the two coherent states $|+\sqrt{\mu_{\mathrm{out}}}\rangle$ and $|-\sqrt{\mu_{\mathrm{out}}}\rangle$, to obtain the encoded bit information from Alice's end.

The information entropy accessible to Eve can be determined from the eigenvalues of the corresponding Gram matrix [74]. For the states under consideration, the Gram matrix is given as follows:

$$
\begin{aligned}
G&=\begin{pmatrix}\langle+\sqrt{\mu_{\mathrm{out}}}|+\sqrt{\mu_{\mathrm{out}}}\rangle & \langle+\sqrt{\mu_{\mathrm{out}}}|-\sqrt{\mu_{\mathrm{out}}}\rangle\\ \langle-\sqrt{\mu_{\mathrm{out}}}|+\sqrt{\mu_{\mathrm{out}}}\rangle & \langle-\sqrt{\mu_{\mathrm{out}}}|-\sqrt{\mu_{\mathrm{out}}}\rangle\end{pmatrix}\\
&=\begin{pmatrix}1 & \mathrm{e}^{-2\mu_{\mathrm{out}}}\\ \mathrm{e}^{-2\mu_{\mathrm{out}}} & 1\end{pmatrix}.
\end{aligned}
\tag{7}
$$

The eigenvalues of the Gram matrix are $\lambda_1=\frac{1}{2}+\frac{1}{2}\mathrm{e}^{-2\mu_{\mathrm{out}}}$ and $\lambda_2=\frac{1}{2}-\frac{1}{2}\mathrm{e}^{-2\mu_{\mathrm{out}}}$. Then, the amount of information leakage caused by Eve executing THA in the second round is expressed as $I(\mathrm{A}:\mathrm{E})=h(\lambda_1)=h(\lambda_2)$.

Therefore, the secrecy message capacity formula for Eve after performing THA only on the second round of transmission can be expressed as

$$
\begin{aligned}
C_s=&Q_{\mu}^{\mathrm{BAB}}\left[1-h\left(E_{\mu}^{\mathrm{BAB}}\right)\right]-Q_{\mu,n=1}^{\mathrm{BAE}}h\left(e_{x,1}^{\mathrm{BA}}+e_{y,1}^{\mathrm{BA}}\right)\\
&-Q_{\mu,n\geq 2}^{\mathrm{BAE}}\times 1-Q_{\mu,\mathrm{encoder}}^{\mathrm{BAE}}h(\lambda_1),
\end{aligned}
\tag{8}
$$

where $Q_{\mu,\mathrm{encoder}}^{\mathrm{BAE}}$ denotes the gain when Eve intercepts the photon and Alice completes the encoding operation.

#### 2.3.3 Two-round transmission

Considering the worst-case scenario, we assume that the information obtained from the first round of attacks and the second round of attacks is non-overlapping. Attacking with both rounds can be regarded as the cumulative effect of the first and second rounds of attacks. Therefore, when Eve performs THA during both the first and second rounds, the secrecy message capacity for the QSDC system is derived from Eqs. (4) and (7) as follows:

$$
\begin{aligned}
C_s=&Q_{\mu}^{\mathrm{BAB}}\left[1-h\left(E_{\mu}^{\mathrm{BAB}}\right)\right]-Q_{\mu,n=1}^{\mathrm{BAE}}h(e_{x,1}^{\mathrm{BA}'}+e_{y,1}^{\mathrm{BA}'})\\
&-Q_{\mu,n\geq 2}^{\mathrm{BAE}}\times 1-Q_{\mu,\mathrm{encoder}}^{\mathrm{BAE}}h(\lambda_1).
\end{aligned}
\tag{9}
$$

### 2.4 Decoy-state method

In this section, we estimated the single-photon error rate for the first round and the second round using the decoy-state method.

In this paper, we assume that the single-photon error rates under the $X$ basis and $Y$ basis are equal, i.e., $e_{x,1}^{\mathrm{BA}}=e_{y,1}^{\mathrm{BA}}=e_1^{\mathrm{BA}}$. The use of the decoy-state method in QSDC can effectively defend against PNS attacks and collective attacks. In here, we employ four decoy states with intensities $\mu$, $v_1$, $v_2$, and $v_3$ (where $\mu$ denotes the signal-state intensity). The upper bounds of single photon DBER $e_1^{\mathrm{BA}}$ are given by [8]

$$
e_1^{\mathrm{BA}}=\frac{E_{\nu_3}^{\mathrm{BA}}Q_{\nu_3}^{\mathrm{BA}}e^{\nu_3}-e_0Y_0^{\mathrm{A}}}{Y_1^{A,L}\nu_3},
\tag{10}
$$

where

$$
Y_1^{\mathrm{A},L}=\frac{\mu^2\left(Q_{\nu_2}^{\mathrm{BA}}e^{\nu_2}-Q_{\nu_3}^{\mathrm{BA}}e^{\nu_3}\right)-\left(\nu_2^2-\nu_3^2\right)\left(Q_{\mu}^{\mathrm{BA}}e^{\mu}-Y_0^{\mathrm{A}}\right)}{\mu\left(\nu_2-\nu_3\right)\left(\mu-\nu_2-\nu_3\right)}.
\tag{11}
$$

In addition, the signal state and decoy state intensity we utilize must also satisfy the following conditions:

$$
\begin{aligned}
0<\nu_3<\nu_2\leq\frac{2}{3}\mu<\nu_1\leq\frac{3}{4}\mu,\\
\nu_1+\nu_2>\mu,\\
\nu_2+\nu_3<\mu,\\
\nu_1-\nu_2-\frac{\nu_1^3-\nu_2^3}{\mu^2}=0.
\end{aligned}
\tag{12}
$$

## 3 Numerical simulation

In numerical simulations, based on the inequality conditions in Eq. (12), we set the strength of the decoy state as: $v_1=0.7\mu$, $v_2=0.445\mu$, and $v_3=0.3\mu$. The parameters in Table 1 are primarily used for numerical simulations [8].

Figure 2(a) illustrates the secrecy message capacity during the first round of the THA attack. It is worth noting that the secrecy message capacity corresponding

**Table 1** Parameters adopted in numerical simulations for the THA model, selected to align with practical QSDC system specifications [8].

| $Y_0^{\mathrm{A(B)}}$ | $\eta_d^{\mathrm{A(B)}}$ | $\eta_{\mathrm{opt}}^{\mathrm{BA}}$ | $\eta_{\mathrm{opt}}^{\mathrm{BAB}}$ | $e_d^{\mathrm{A}}$ | $e_d^{\mathrm{B}}$ |
|---|---|---|---|---|---|
| $8\times10^{-8}$ | 0.7 | 0.21 | 0.088 | 0.0131 | 0.0026 |

to the value of $\mu_{\mathrm{out}}=10^{-8}$ remains indistinguishable from the no attack case ($\mu_{\mathrm{out}}=0$) over almost the entire range of transmission. In addition, we note that when $\mu_{\mathrm{out}}$ increases to $10^{-3}$, the secrecy message is still all stolen by Eve, i.e., the secrecy message capacity vanishes entirely. Finally, when the attack intensity $\mu_{\mathrm{out}}=10^{-4}$ is applied, the maximum secrecy message capacity achievable is $1.89\times10^{-4}$ bit/pulse, with a maximum secure communication distance of 6.97 km.

Figure 2(b) illustrates the secrecy message capacity during the second round of the THA attack. As shown in Fig. 2(b), we observe that for an average reflected photon number $\mu_{\mathrm{out}}$ reaches the magnitude of $10^{-4}$, the secrecy message capacity remains nearly indistinguishable from the unattacked case over almost the entire transmission distance. Compared with Fig. 2(a), when $\mu_{\mathrm{out}}$ is uniformly set to $10^{-4}$, the system can still achieve a maximum secrecy message capacity of $3.23\times10^{-4}$ bit/pulse and a maximum secure communication distance of 16.68 km. Unlike the scenario where the THA targets only the first round, selecting the maximum attack strength $\mu_{\mathrm{out}}=10^{-2}$ allows the system to attain a maximum secrecy message capacity of $2.07\times10^{-4}$ bit/pulse and a maximum secure communication distance of 9.18 km. These results collectively indicate that confining the THA to the second transmission round substantially reduces its effectiveness relative to an attack launched solely in the first round. From a physical standpoint, this is because Eve's attack in the first round affects Bob's quantum state, and by combining it with a coherent attack, she can obtain a greater amount of information. But when Eve conducts a THA during the second transmission round, although her activities remain undetectable to the legitimate parties, she can only extract information by distinguishing among a constrained number of reflected photons. Consequently, the amount of information she acquires is substantially reduced compared to an attack mounted in the first round.

Figure 3(a) illustrates the secrecy message capacity of the first round of the THA attack after numerical optimization. There is a trade-off between the total gain $Q$ and the error rate $E$. A larger $\mu$ increases the probability of photon reception, thereby improving the response rate; however, it also increases the multiphoton component, leading to more information leakage to Eve. These two influences contribute positively and negatively to the secrecy message capacity, and our so-called numerical optimization is to find an appropriate value of $\mu$ in the channel attenuation such that the equilibrium point between the two effects can be achieved at a specific $\mu$. As shown in Fig. 3(a), when the attack is relatively weak, the impact is minimal. For instance, when $\mu_{\mathrm{out}}=10^{-7}$ and $10^{-8}$, at a channel attenuation of 6 dB, the message capacities are $1.41\times10^{-5}$ bit/pulse and $1.73\times10^{-5}$ bit/pulse, respectively, with maximum communication distances of 17.80 km and 17.46 km. For comparison, when $\mu_{\mathrm{out}}=0$, the secrecy message capacity is $4.13\times10^{-5}$ bit/pulse, and the maximum communication distance is 17.96 km. As attack intensity gradually increases, both secrecy message capacity and maximum communication distance decrease accordingly. For example, at an attack intensity of $\mu_{\mathrm{out}}=10^{-5}$ and $10^{-6}$, the secrecy message capacities at a channel attenuation of 5 dB are $6.68\times10^{-6}$ bit/pulse and $4.56\times10^{-5}$ bit/pulse, respectively, with maximum communication distances of 13.71 km and 16.46 km. Notably, when the attack

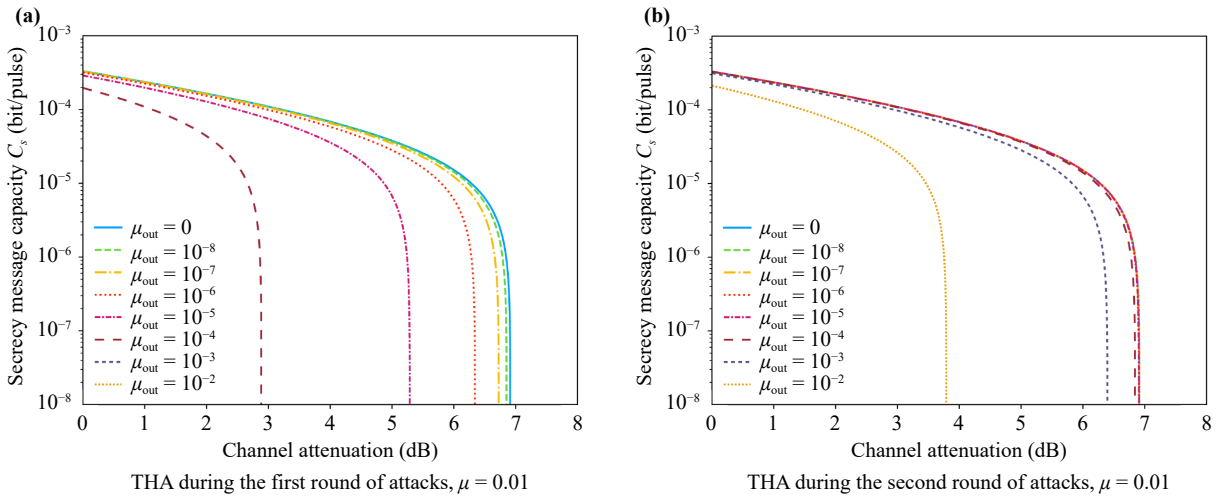



**Fig. 2** Secrecy message capacity $C_s$ as a function of channel attenuation for the decoy-state DL04 protocol under THA.

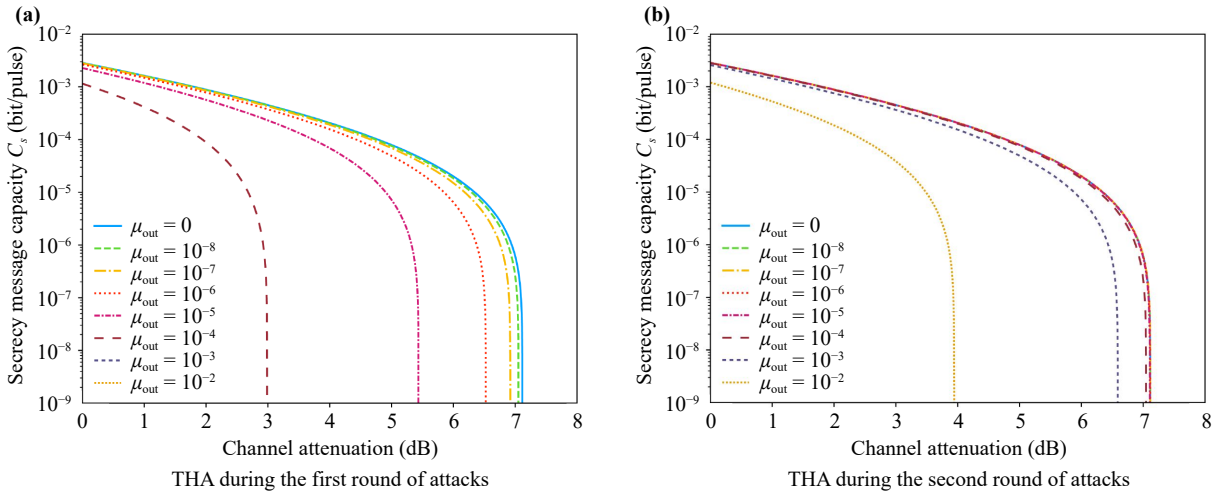



**Fig. 3** Secrecy message capacity $C_s$ versus channel attenuation for the numerically optimized decoy-state DL04 QSDC protocol in the first and second transmission rounds under THA. Numerical optimization involves optimizing the input average photon number $\mu$ based on channel attenuation to maximize $C_s$ within the range $[0, 1]$.

strength increases again to the $10^{-3}$ magnitude, the secrecy message is still all stolen by Eve, i.e., the secrecy message capacity is always zero.

Figure 3(b) illustrates the secrecy message capacity of the second round of the THA attack after numerical optimization. Numerical simulation results indicate that when $\mu_{out}$ is set to $10^{-8}$ and $10^{-4}$ under a 7 dB channel attenuation, the secrecy message capacities are $5.54 \times 10^{-7}$ and $1.79 \times 10^{-7}$ bit/pulse, respectively. With maximum communication distances of 17.80 and 17.63 km, respectively. Therefore, we conclude that when attack intensity is low, the impact is minimal. However, as attack intensity increases, the resulting impact grows rapidly. When attack intensity reaches $\mu_{out} = 10^{-2}$, the maximum secrecy message capacity is $1.07 \times 10^{-3}$ bit/pulse, and the maximum secure communication distance reaches 9.85 km.

Figure 4 illustrates the secrecy message capacity of the THA attack over two rounds under finite decoy states and infinite decoy states following numerical optimization. Here, we employ Eqs. (A3) and (A7) to calculate the error rate and yield under the theoretical model, respectively. The values calculated by this theoretical model are equivalent to those computed under conditions of infinite decoy states. Provided that the stolen secrecy message is non-repeating, Fig. 4 can be regarded as a simple accumulation of Figs. 3(a) and (b). For example, when $\mu_{out} = 10^{-8}$, at a channel attenuation of 6 dB, the secrecy message capacity under finite and infinite decoy-state conditions is $1.41 \times 10^{-5}$ bit/pulse and $1.84 \times 10^{-5}$ bit/pulse, respectively, with maximum communication distances of 17.32 km. As attack intensity gradually increases, both secrecy message capacity and maximum communication distance decrease accordingly. For example, at an attack intensity of $\mu_{out} = 10^{-6}$, the secrecy message capacities at a channel attenuation of 5 dB are $4.55 \times 10^{-5}$ bit/pulse and $4.88 \times 10^{-5}$ bit/pulse, respectively, with maximum communication distances of 16.32 km. Similar to Fig. 3(a), when the attack intensity increases again to $\mu_{out} = 10^{-3}$, the secrecy message capacity is zero. In summary, under infinite decoy states, the secrecy message capacity is slightly higher than under finite decoy states, while the maximum secure communication distance remains consistent.

## 4 Discussion and conclusion

In this work, we have presented a comprehensive security analysis of THA on the DL04 protocol. By developing a detailed threat model and combining WCP source with the decoy-state method, we derived analytical expressions for the secrecy message capacity under attacks targeting the first round, the second round. Our numerical simulations are based on practical experimental parameters and are visualized in Figs. (2), (3) and (4). They yield several key, quantifiable insights into the system's security.

First, the system's security is highly sensitive to the average number of reflected Trojan photons, $\mu_{out}$. We identify a critical boundary: In the first round, for $\mu_{out} \leq 10^{-8}$, the secrecy message capacity and maximum communication distance remain nearly identical to the attack-free case ($\mu_{out} = 0$). In contrast, when $\mu_{out}$ increases to $10^{-3}$, the secrecy message capacity drops to zero, indicating compromised security. This threshold underscores the necessity of stringent optical isolation in encoder design to suppress $\mu_{out}$ below $10^{-8}$. In the second round, when $\mu_{out} < 10^{-4}$, the secrecy message capacity and maximum communication distance are nearly identical to those under no attack conditions. Therefore, this threshold range slightly reduces the

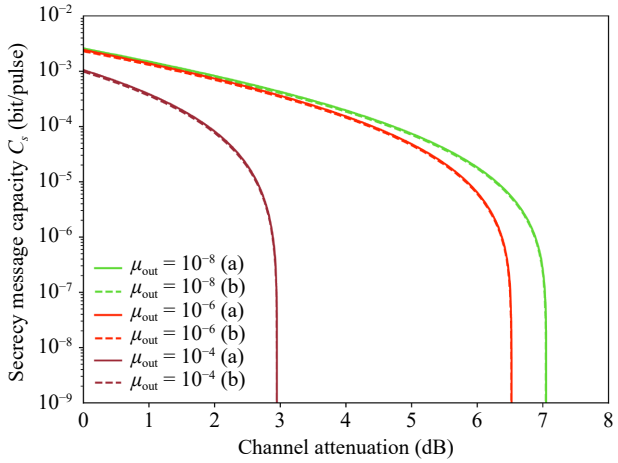



**Fig. 4** Secrecy message capacity $C_s$ versus channel attenuation for the numerically optimized decoy-state DL04 QSDC protocol in the two rounds under THA. Solid lines (a) and dashed lines (b) represent attack strengths under finite and infinite decoy states, respectively. Numerical optimization involves optimizing the input average photon number $\mu$ based on channel attenuation to maximize $C_s$ within the range $[0, 1]$.

requirements for optical isolators.

Second, we reveal a pronounced and operationally significant asymmetry in vulnerability between the two transmission rounds. Figure 2 provides a direct comparison: a first-round attack at strongest ($\mu_{\text{out}} = 10^{-2}$) nullifies the secrecy message capacity [Fig. 2(a)], whereas a second-round attack at the same intensity still supports a maximum secrecy message capacity of $2.07 \times 10^{-4}$ bit/pulse and extends the maximum secure distance to 9.18 km [Fig. 2(b)]. This stark contrast arises from a fundamental difference in Eve's information-gain mechanism. During a first-round attack, she influences the quantum state and matches collective attacks to obtain more information. Consequently, this significantly compromises security. In the second round, she could theoretically steal more information undetected. However, in practice, she can only obtain information by distinguishing a limited number of reflected photons, resulting in substantially less information compared to the first round.

Finally, Fig. 3 demonstrates the practical value of numerical optimization for the system. Figure 4 illustrates the secrecy message capacity of the THA attack over two rounds under finite decoy states and infinite decoy states following numerical optimization. By strategically adjusting the signal strength $\mu$ according to channel attenuation, we can balance the trade-off between higher photon reception probability and increased multi-photon leakage. Figure 3(a) shows that optimization recovers performance even under moderate attack strength (e.g., achieving $6.68 \times 10^{-6}$ bit/pulse at 5 dB for $\mu_{\text{out}} = 10^{-5}$). By comparison, Fig. 3(b) shows that when attack strength is $\mu_{\text{out}} = 10^{-4}$, a performance of $1.79 \times 10^{-7}$ bit/pulse can still be achieved at 7 dB. Therefore, the second round of attacks has a relatively minor impact on the protocol's secrecy message capacity. The values in Fig. 4 can be regarded as a simple summation of those in Figs. 3(a) and (b). Numerical simulations show that when $\mu_{\text{out}} = 10^{-4}$, the secrecy message capacity becomes zero under 3 dB channel attenuation conditions. This result further demonstrates that when THA attacks both the first and second transmission rounds, the first round of attacks is dominant.

Our results bridge a critical gap in QSDC security analysis, as prior work on THA has focused primarily on QKD [64–67]. Unlike QKD, where THA targets detector vulnerabilities, QSDC's direct encoding of the secrecy message into quantum states makes it uniquely susceptible to phase leakage via reflected photons. This distinction underscores the necessity of tailored security frameworks for QSDC, such as the one presented here. Our research findings provide guidance on safety boundaries for practical system design.

As captured by Eq. (1) and Eq. (6), the overall effect of the THA is fully characterized by the parameter $\mu_{\text{out}}$. Therefore, the derived secrecy message capacity holds for any QSDC system that enforces a strict upper bound on the number of Trojan photons reflected back to Eve at the transmitter. Traditionally, to prevent THA attacks at their source, typical countermeasures include the use of optical isolators or optical power limiters [77, 78]. These components significantly attenuate externally injected optical signals, thereby reducing the effectiveness of such attacks. However, recent studies indicate that under high power optical illumination, the attenuation of optical attenuators and the isolation of optical isolators degrade, potentially allowing Eve to resume light injection attacks [79]. When combined with THA, this compromises the security of QSDC systems. Fortunately, significant advances have been made in THA defense strategies. For example, in 2022, Chen *et al.* [80] employed reference technology to enhance the security of four-phase MDI-QKD protocols under imperfect real-world light sources. By constructing a reference state representation, this method systematically characterizes and quantifies potential vulnerabilities such as light source preparation defects, mode-dependent side channels, pulse correlations, and THA, correlating these parameters with experimentally measurable data. In 2026, Chen *et al.* [81] proposed and experimentally verified an improved coherent one-way quantum key distribution (COW-QKD) protocol. By employing quantum state modulation using only vacuum and coherent states, this method avoids potential side-channel leaks caused by imperfections in light sources, a common issue in traditional decoy-state protocols. In addition, Wei *et al.* [82] proposed a numerical method to enhance the security boundary of

decoy-state QKD systems under THA. Lastly, fully passive QSDC protocols [83] — which avoid active modulation of quantum states — offer a promising approach to mitigate side-channel risks introduced by THA. In terms of experiments, Avesani *et al.* [84] developed a polarization encoder capable of producing predefined polarization states with a fixed reference frame in free space, achieving long-term stability without requiring calibration at either the transmitter or receiver. Building on this work, Toni *et al.* [85] introduced an adaptive countermeasure that utilizes such a polarization encoder to effectively suppress THA. In practice, these methods intrinsically suppress the value of $\mu_{\text{out}}$, and when combined with decoy-state techniques, they collectively improve the practical security of the system. This direction aligns with recent progress in passive quantum communication protocols [86–92], reflecting a growing research focus on achieving robust, device-agnostic security.

It should be noted that our analytical framework is not directly applicable to other types of quantum communication protocols without modification. For example, the MDI-QSDC protocol introduces additional complexity due to its reliance on untrusted third parties and a measurement-device-agnostic architecture. In such a setting, THA may target not only the encoding devices of the legitimate users but also the measurement stations. Although the current model does not extend straightforwardly to the measurement side, the core methodology — bounding information leakage via reflected photons from the encoding module — remains transferable. Specifically, by examining the encoding operations of both Alice and Bob, as well as the characteristics of the corresponding reflected optical signals, analogous adaptation strategies can be developed. Hence, future work could aim to generalize this framework to accommodate MDI-QSDC systems.

Notably, our security analysis is based on asymptotic assumptions, whereas practical systems must account for finite-key effects. Therefore, our results should be regarded as upper bounds on performance under ideal conditions. A more comprehensive finite-key analysis will be addressed in future research. We plan to address this limitation in future work by incorporating finite-key effects into the security model [93, 94]. This will enable us to derive more precise estimates of the secrecy message capacity and security thresholds, thereby providing more practical guidance for selecting experimental parameters.

**Declarations** The authors declare that they have no competing interests and there are no conflicts.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China under Grant Nos. 12574393, 92365110, and 12175106.

## Appendix A: System model

To analyze the security of QSDC systems under THA, we establish a comprehensive system model encompassing the light source, quantum channel, and detection components, with key parameters defined as follows.

The transmission efficiency of the quantum channel is characterized by signal attenuation, expressed as

$$t^{\text{path}} = 10^{-\frac{\alpha^{\text{path}}}{10}}, \tag{A1}$$

where $\alpha^{\text{path}}$ denotes the quantum channel losses and $\text{path} \in \{\text{BA}, \text{BAB}\}$, where BA denotes the initial transmission path and BAB denotes the two-way path (Bob $\rightarrow$ Alice $\rightarrow$ Bob) involving round-trip photon transmission.

The total transmission efficiency, accounting for both channel loss and device imperfections, is

$$\eta^{\text{path}} = t^{\text{path}} \eta_{\text{opt}}^{\text{path}} \eta_d^{\text{par}}, \tag{A2}$$

where $\eta_{\text{opt}}^{\text{path}}$ are the specific device's inherent optical losses and $\eta_d^{\text{par}}$ denotes the detection efficiency associated with either Alice or Bob, i.e., $\text{par} \in \{\text{Alice}, \text{Bob}\}$.

The channel's yield and gain are determined to ascertain the secrecy message capacity. The yields of $n$-photon signals at the locations of Alice and Bob, respectively, can be represented as $Y_n^{\text{A}}$ and $Y_n^{\text{B}}$, which can be expressed as

$$Y_n^{\text{par}} = 1 - (1 - Y_0^{\text{par}})(1 - \eta^{\text{path}})^n, \tag{A3}$$

where $Y_0^{\text{par}}$ is the zero-photon yield. $(1 - Y_0^{\text{par}})(1 - \eta^{\text{path}})^n$ means that all photons are lost during quantum channel transmission and no dark count occurs.

We can ascertain the overall signal gain and error rate by developing the aforementioned channel model. The overall signal gain can be expressed as follows:

$$Q_\mu^{\text{path}} = \sum_{n=0}^{\infty} Q_{\mu,n}^{\text{path}} = \sum_{n=0}^{\infty} P(n,\mu) Y_n^{\text{par}}, \tag{A4}$$

where $P(n,\mu) = \frac{\mathrm{e}^{-\mu}}{n!}\mu^n$ is the Poisson distribution of photon numbers with average $\mu$, and $Q_{\mu,n}^{\text{path}}$ are the $n$-photon signal gain at Alice or Bob.

According to Ref. [8], the total signal gain of Eve can be computed as follows:

$$\begin{aligned} Q_\mu^{\text{BAE}} &= \sum_{n=0}^{\infty} Q_{\mu,n}^{\text{BAE}} \leq \sum_{n=0}^{\infty} [Q_{\mu,n}^{\text{BA}} - P(n,\mu) Y_0^{\text{A}}] \\ &\quad \times \max\left\{1, \frac{\eta^{\text{BAE}}}{\eta^{\text{BA}}}\right\}, \\ Q_{\mu,n=1}^{\text{BAE}} &= P(n,\mu)\left(Y_1 - Y_0^{\text{A}}\right) \times \max\left\{1, \frac{\eta^{\text{BAE}}}{\eta^{\text{BA}}}\right\}, \\ Q_{\mu,n\geq 2}^{\text{BAE}} &= \left(Q_\mu^{\text{BA}} - Y_0^{\text{A}} - Q_{\mu,n=1}^{\text{BAE}}\right) \times \max\left\{1, \frac{\eta^{\text{BAE}}}{\eta^{\text{BA}}}\right\}, \end{aligned} \tag{A5}$$

where $\eta^{\text{BAE}}$ represents the total transmission efficiency of

the information that Eve acquires after Alice receives and encodes the photon, and $\eta^{\mathrm{BA}}$ symbolizes the total transmission efficiency of the photon that Alice receives and measures. Similarly, we can calculate the error rate of our QSDC protocol as follows:

$$E_{\mu}^{\mathrm{path}}=\frac{\sum_{n=0}^{\infty}p(n,\mu)e_{n}Y_{n}^{\mathrm{par}}}{Q_{\mu}^{\mathrm{path}}}. \tag{A6}$$

Here, we construct the error model $e_nY_n^{\mathrm{par}}$ as

$$e_{n}Y_{n}^{\mathrm{par}}=e_{0}^{\mathrm{par}}Y_{0}^{\mathrm{par}}+e_{d}^{\mathrm{par}}[1-(1-\eta^{\mathrm{path}})^{n}], \tag{A7}$$

where $e_0^{\mathrm{par}}$ is the error rate resulting from dark counts, and $e_d^{\mathrm{par}}$ is the detector error response rate. $e_0^{\mathrm{par}}$ is assumed to be 0.5, which means when no photons arrive, one of the two detectors may produce an erroneous reading due to dark counting, with a probability of 0.5.

## Appendix B: Rate equations for Trojan horse attack on QSDC

Here, we employ entanglement-based quantum state preparation [76]. The states to be prepared are given in Eq. (1). We rewrite them here for convenience:

$$\begin{aligned}
|\psi_{+}\rangle_{\mathrm{BE}}&=|+\rangle_{\mathrm{B}}\otimes|+\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\\
|\psi_{-}\rangle_{\mathrm{BE}}&=|-\rangle_{\mathrm{B}}\otimes|-\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\\
|\psi_{R}\rangle_{\mathrm{BE}}&=|R\rangle_{\mathrm{B}}\otimes|+\mathrm{i}\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}},\\
|\psi_{L}\rangle_{\mathrm{BE}}&=|L\rangle_{\mathrm{B}}\otimes|-\mathrm{i}\sqrt{\mu_{\mathrm{out}}}\rangle_{\mathrm{E}}.
\end{aligned} \tag{B1}$$

The $X$ basis states can be prepared by Bob by measuring in the basis $\{|+\rangle,|-\rangle\}$ the following entangled state:

$$|\Psi_{X}\rangle=\frac{|+\rangle_{\mathrm{B}}|\psi_{+}\rangle_{\mathrm{BE}}+|-\rangle_{\mathrm{B}}|\psi_{-}\rangle_{\mathrm{BE}}}{\sqrt{2}}. \tag{B2}$$

The $Y$ basis states of Eq. (B1) can be prepared by Bob by measuring in the basis $\{|R\rangle,|L\rangle\}$ the following entangled state:

$$|\Psi_{Y}\rangle=\frac{|L\rangle_{\mathrm{B}}|\psi_{R}\rangle_{\mathrm{BE}}+|R\rangle_{\mathrm{B}}|\psi_{L}\rangle_{\mathrm{BE}}}{\sqrt{2}}, \tag{B3}$$

where the states $|\Psi_X\rangle$ and $|\Psi_Y\rangle$ are hard to distinguish when $\mu_{\mathrm{out}}\to 0$. Using

$$\begin{aligned}
\langle+\mathrm{i}\sqrt{\mu_{\mathrm{out}}}\mid+\sqrt{\mu_{\mathrm{out}}}\rangle&=\mathrm{e}^{-|\sqrt{\mu_{\mathrm{out}}}|^{2}}\mathrm{e}^{-\mathrm{i}|\sqrt{\mu_{\mathrm{out}}}|^{2}}\\
&=\langle-\mathrm{i}\sqrt{\mu_{\mathrm{out}}}\mid-\sqrt{\mu_{\mathrm{out}}}\rangle,\\
\langle-\mathrm{i}\sqrt{\mu_{\mathrm{out}}}\mid+\sqrt{\mu_{\mathrm{out}}}\rangle&=\mathrm{e}^{-|\sqrt{\mu_{\mathrm{out}}}|^{2}}\mathrm{e}^{+\mathrm{i}|\sqrt{\mu_{\mathrm{out}}}|^{2}}\\
&=\langle+\mathrm{i}\sqrt{\mu_{\mathrm{out}}}\mid-\sqrt{\mu_{\mathrm{out}}}\rangle,
\end{aligned} \tag{B4}$$

we find

$$\begin{aligned}
\langle\Psi_{X}|\Psi_{Y}\rangle&=\frac{1}{2}\left(\langle+|\otimes\langle\psi_{+}|_{\mathrm{BE}}+\langle-|\otimes\langle\psi_{-}|_{\mathrm{BE}}\right)\\
&\quad\otimes\left(|L\rangle\otimes|\psi_{R}\rangle_{\mathrm{BE}}+|R\rangle\otimes|\psi_{L}\rangle_{\mathrm{BE}}\right)\\
&=\frac{1}{2}(\langle+|L\rangle\langle\psi_{+}|\psi_{R}\rangle_{\mathrm{BE}}+\langle+|R\rangle\langle\psi_{+}|\psi_{L}\rangle_{\mathrm{BE}}\\
&\quad+\langle-|L\rangle\langle\psi_{-}|\psi_{R}\rangle_{\mathrm{BE}}+\langle-|R\rangle\langle\psi_{-}|\psi_{L}\rangle_{\mathrm{BE}})\\
&=\frac{1}{8}[2\mathrm{e}^{+\mathrm{i}\mu_{\mathrm{out}}}\mathrm{e}^{-\mu_{\mathrm{out}}}+2\mathrm{e}^{-\mathrm{i}\mu_{\mathrm{out}}}\mathrm{e}^{-\mu_{\mathrm{out}}}\\
&\quad+2\mathrm{e}^{-\mathrm{i}\mu_{\mathrm{out}}}\mathrm{e}^{-\mu_{\mathrm{out}}}+2\mathrm{e}^{+\mathrm{i}\mu_{\mathrm{out}}}\mathrm{e}^{-\mu_{\mathrm{out}}}]\\
&=\frac{1}{2}\mathrm{e}^{-\mu_{\mathrm{out}}}\left(\mathrm{e}^{+\mathrm{i}\mu_{\mathrm{out}}}+\mathrm{e}^{-\mathrm{i}\mu_{\mathrm{out}}}\right)\\
&=\mathrm{e}^{-\mu_{\mathrm{out}}}\cos\mu_{\mathrm{out}}.
\end{aligned} \tag{B5}$$

When $\mu_{\mathrm{out}}\to 0$, the inner product of $|\Psi_X\rangle$ and $|\Psi_Y\rangle$ approaches 1. This implies that the states $|\Psi_X\rangle$ and $|\Psi_Y\rangle$ largely overlap. Consequently, it becomes difficult to distinguish whether the emitted state originated from the choice of the $X$ basis or the $Y$ basis. Meanwhile, since the attack analysis considers the single-photon scenario, we consider the worst-case scenario, i.e., two-photon and above information leakage is taken as 1. Therefore, the secrecy message capacity for the DL04 protocol using $X$ or $Y$ basis random preparation would be [9]

$$\begin{aligned}
C_{s}=&Q_{\mu}^{\mathrm{BAB}}\left[1-h\left(E_{\mu}^{\mathrm{BAB}}\right)\right]-Q_{\mu,n=1}^{\mathrm{BAE}}h(e_{x,1}^{\mathrm{BA}}+e_{y,1}^{\mathrm{BA}})\\
&-Q_{\mu,n\geq 2}^{\mathrm{BAE}}\times 1,
\end{aligned} \tag{B6}$$

where $Q_{\mu}^{\mathrm{BAB}}$ is the overall signal gain of Bob after a round-trip BAB. $E_{\mu}^{\mathrm{BAB}}$ is the QBER. $Q_{\mu,n}^{\mathrm{BAE}}$ denotes the gain of $n$-photon events from Eve. The terms $e_{x,1}^{\mathrm{BA}}$ and $e_{y,1}^{\mathrm{BA}}$ denote the single-photon error rates under the $X$ basis and $Y$ basis, respectively.

In our worst-case scenario analysis, we attribute full information leakage $(h(e)=1)$ to multi-photon events $(n\geq 2)$. This is a standard and conservative approach in quantum cryptography, justified by the PNS attack, where Eve can perfectly split off and store one photon from a multi-photon pulse without introducing detectable loss. While this might overestimate Eve's actual information gain, it ensures the unconditional security of the remaining secrecy message capacity.

When the preparation is not perfect, the states $|\Psi_X\rangle$ and $|\Psi_Y\rangle$ will become basically distinguishable, and the above secrecy message capacity formula needs to be replaced by the following:

$$\begin{aligned}
C_{s}=&Q_{\mu}^{\mathrm{BAB}}\left[1-h\left(E_{\mu}^{\mathrm{BAB}}\right)\right]-Q_{\mu,n=1}^{\mathrm{BAE}}h(e_{x,1}^{\mathrm{BA}'}+e_{y,1}^{\mathrm{BA}'})\\
&-Q_{\mu,n\geq 2}^{\mathrm{BAE}}\times 1,
\end{aligned} \tag{B7}$$

where the term $e_{x,1}^{\mathrm{BA}'}$ $(e_{y,1}^{\mathrm{BA}'})$ is the single-photon error

rate, where Alice chooses $Y$-based measurements while Bob declares $X$ basis (or Alice chooses $X$ basis measurements while Bob declares $Y$ basis). In other words, the terms $e_{x,1}^{\mathrm{BA}'}$ and $e_{y,1}^{\mathrm{BA}'}$ are upper bounds on the single-photon error rate.

To find the relationship between the upper limit of the single-photon error rate $e_{x,1}^{\mathrm{BA}'}$ and $e_{y,1}^{\mathrm{BA}'}$, we can assume that Bob owns a private bidimensional quantum system, the "quantum coin" [76], and prepares the following state:

$$|\Phi\rangle = \frac{|H\rangle_C|\Psi_X\rangle + |V\rangle_C|\Psi_Y\rangle}{\sqrt{2}}, \tag{B8}$$

where the subscript $C$ denotes the quantum coin. $|H\rangle$ and $|V\rangle$ denote the two quantum states prepared in the $Z$ basis, respectively. Bob can then prepare the states in Eq. (B1) by first measuring the quantum coins in the basis of $\{|H\rangle, |V\rangle\}$, and then based on the results of the measurements of the states $|\Psi_X\rangle$ or $|\Psi_Y\rangle$ obtained by measuring them based on $\{|+\rangle, |-\rangle\}$ or $\{|R\rangle, |L\rangle\}$, respectively.

To quantify how Bob's emitted state depends on the basis choice, we examine the imbalance of the quantum coin. First, we rewrite Eq. (B9) in the $Z$ basis:

$$\begin{aligned}|\Phi\rangle &= \frac{|H\rangle_C|\Psi_X\rangle + |V\rangle_C|\Psi_Y\rangle}{\sqrt{2}} \\ &= \frac{\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)|\Psi_X\rangle + \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)|\Psi_Y\rangle}{\sqrt{2}} \\ &= \frac{|+\rangle(|\Psi_X\rangle + |\Psi_Y\rangle) + |-\rangle(|\Psi_X\rangle - |\Psi_Y\rangle)}{2}.\end{aligned} \tag{B9}$$

In addition, we also need to evaluate the probability that two states $|\Psi_X\rangle$ and $|\Psi_Y\rangle$ are different. A useful way to do this is to consider what would happen if Bob measured each of his coins based on the $X$ eigenstates rather than the $Y$ eigenstates; the result $X = -1$ (associated with the state $|-\rangle$) would then occur with probability $\Delta$ that

$$\Delta(prob, X = -1) = \left|\frac{(|\Psi_X\rangle - |\Psi_Y\rangle)}{2}\right|^2 = \frac{1 - Re\langle\Psi_X|\Psi_Y\rangle}{2}. \tag{B10}$$

It should be emphasized that $\Delta$ quantifies the degree of leakage of the "basis correlation" at the source end, and Eve exploits this correlation to enhance the efficiency of her attacks.

By combining Eq. (B5), we can further derive the probability of $\Delta$ [75]

$$\Delta = \frac{1}{2}(1 - \mathrm{e}^{-\mu_{\mathrm{out}}}\cos\mu_{\mathrm{out}}). \tag{B11}$$

We can find that when $\mu_{\mathrm{out}} = 0$, $\Delta = 0$, the quantum state emitted by Bob is independent of the basis. However, when $\mu_{\mathrm{out}} > 0$, $\Delta$ is positive, and the quantum state carries some basis information out of Bob's device. Eve can utilize some basic information to enhance her attack strategy. Specifically, since not all signals from the source are necessarily detected, Eve can replace the actual channel with a lossless one. Then, it will selectively drop some states that are unfavorable until the loss rates measured by both communicating parties match. To account for this possibility, we must consider the worst-case scenario in which all undetected events come from the $X = 1$ eigenstates of the quantum coin, and adjust $\Delta$ accordingly:

$$\Delta' = \frac{\Delta}{y}, \tag{B12}$$

where $y = \min[y_x, y_y]$, $y_x$ and $y_y$ denote the single-photon yields measured under the $X$ basis and $Y$ basis, respectively. The relationship between the estimated upper limit of the single-photon error rate, $e_{x,1(y,1)}^{\mathrm{BA}'}$, and the theoretical single-photon error rate, $e_{x,1(y,1)}^{\mathrm{BA}}$, can be obtained by utilizing the Bloch sphere bound [95] and the effective coin imbalance $\Delta'$.

We now introduce the relevant mathematical objects. Let $\rho_x$ and $\rho_y$ denote the density matrices of the $X$ and $Y$ basis states, respectively, when preparation errors are taken into account. Let $\{E_n\}$ represent a POVM set [96]. The fidelity between these two density operators, which relates to the statistical overlap of measurements, satisfies the following inequality:

$$\sqrt{F(\rho_x, \rho_y)} \equiv \left\|\sqrt{\rho_x}\sqrt{\rho_y}\right\|_{tr} \le \sum_n \sqrt{\mathrm{tr}(\rho_x E_n)\,\mathrm{tr}(\rho_y E_n)}. \tag{B13}$$

We make the following assumptions for our security analysis. For each signal state that Bob sends: If Bob announces the X basis, the joint state of his qubit and the bosonic mode is $\rho_x$. If Bob announces the $Y$ basis, the joint state is $\rho_y$. Alice's detection efficiency is independent of the basis choice. Under these conditions, let $y_x$ and $y_y$ denote the measurement probabilities corresponding to the joint states $\rho_x$ and $\rho_y$, respectively. In addition, there are three kinds of results in the POVM set: one is to output "Uncertain" when Alice's measurement result is uncertain; the other is to output "Consistent" when Alice utilizes $X$ basis to make a measurement and the result is certain and consistent with Bob's $X$ basis result; the other is to output "Consistent" when Alice utilizes $Y$ basis to make a measurement and the result is certain and consistent with Bob's $Y$ basis result. Third, when Alice's $X$ basis measurement result is certain, but the result is consistent with Bob's $X$ basis measurement result, the output is "consistent". When Alice's $X$ basis measurement result is certain but inconsistent with Bob's $X$ basis measurement result, the output is "inconsistent". Then inequality (B14) is further written as

$$\sqrt{F(\rho_x,\rho_y)} \le \sqrt{(1-y_x)(1-y_y)} + \sqrt{y_x y_y}\left[\sqrt{e^{\mathrm{BA}}_{x,1}e^{\mathrm{BA}}_{y,1}} + \sqrt{\left(1-e^{\mathrm{BA}}_{x,1}\right)\left(1-e^{\mathrm{BA}}_{y,1}\right)}\right]. \tag{B14}$$

For fixed values of $y_x + y_y$, the right-hand side of the inequality obtains its maximum value when $y_x = y_y$. Thus, defining $y = (y_x + y_y)/2$ further yields

$$\sqrt{F(\rho_x,\rho_y)} \le 1 - y + y\left[\sqrt{e^{\mathrm{BA}}_{x,1}e^{\mathrm{BA}}_{y,1}} + \sqrt{\left(1-e^{\mathrm{BA}}_{x,1}\right)\left(1-e^{\mathrm{BA}}_{y,1}\right)}\right]. \tag{B15}$$

Note that Eve's attack does not increase the distinguishability of $\rho_x$ and $\rho_y$, so we have

$$1 - 2\Delta \le |\langle\Psi_X|\Psi_Y\rangle| \le \sqrt{F(\rho_x,\rho_y)}, \tag{B16}$$

where $\Delta$ quantifies the dependence of Bob's sending signaling state on the basis. By combining Eq. (B13), Eq. (B16) and Eq. (B17) can be obtained:

$$1 - 2\Delta' \le \sqrt{e^{\mathrm{BA}}_{x,1}e^{\mathrm{BA}}_{y,1}} + \sqrt{\left(1-e^{\mathrm{BA}}_{x,1}\right)\left(1-e^{\mathrm{BA}}_{y,1}\right)}. \tag{B17}$$

If further using the Cauchy–Schwartz inequality [97], the upper bound for $e^{\mathrm{BA}'}_{x,1(y,1)}$ can be estimated as

$$e^{\mathrm{BA}'}_{x,1(y,1)} = e^{\mathrm{BA}}_{y,1(x,1)} + 4\Delta'(1-\Delta')(1-2e^{\mathrm{BA}}_{y,1(x,1)}) + 4(1-2\Delta')\sqrt{\Delta'(1-\Delta')e^{\mathrm{BA}}_{y,1(x,1)}(1-e^{\mathrm{BA}}_{y,1(x,1)})}. \tag{B18}$$

## References


1. G. L. Long and X. S. Liu, Theoretically efficient high-capacity quantum-key-distribution scheme, *Phys. Rev. A* 65(3), 032302 (2002)
2. F. G. Deng, G. L. Long, and X. S. Liu, Two-step quantum direct communication protocol using the Einstein–Podolsky–Rosen pair block, *Phys. Rev. A* 68(4), 042317 (2003)
3. F. G. Deng and G. L. Long, Secure direct communication with a quantum one-time pad, *Phys. Rev. A* 69(5), 052319 (2004)
4. C. Wang, F. G. Deng, Y. S. Li, X. S. Liu, and G. L. Long, Quantum secure direct communication with high-dimension quantum superdense coding, *Phys. Rev. A* 71(4), 044305 (2005)
5. C. Wang, F. G. Deng, and G. L. Long, Multi-step quantum secure direct communication using multi-particle Green–Horne–Zeilinger state, *Opt. Commun.* 253(1–3), 15 (2005)
6. X. F. Zou and D. W. Qiu, Three-step semiquantum secure direct communication protocol, *Sci. China Phys. Mech. Astron.* 57(9), 1696 (2014)
7. T. Li and G. L. Long, Quantum secure direct communication based on single-photon Bell-state measurement, *New J. Phys.* 22(6), 063017 (2020)
8. D. Pan, Z. S. Lin, J. W. Wu, H. R. Zhang, Z. Sun, D. Ruan, L. G. Yin, and G. L. Long, Experimental free-space quantum secure direct communication and its security analysis, *Photon. Res.* 8(9), 1522 (2020)
9. X. Liu, Z. J. Li, D. Luo, C. F. Huang, D. Ma, M. M. Geng, J. W. Wang, Z. R. Zhang, and K. J. Wei, Practical decoy-state quantum secure direct communication, *Sci. China Phys. Mech. Astron.* 64(12), 120311 (2021)
10. G. L. Long and H. Zhang, Drastic increase of channel capacity in quantum secure direct communication using masking, *Sci. Bull. (Beijing)* 66(13), 1267 (2021)
11. Y. B. Sheng, L. Zhou, and G. L. Long, One-step quantum secure direct communication, *Sci. Bull. (Beijing)* 67(4), 367 (2022)
12. Y. X. Xiao, L. Zhou, W. Zhong, M. M. Du, and Y. B. Sheng, The hyperentanglement-based quantum secure direct communication protocol with single-photon measurement, *Quantum Inform. Process.* 22(9), 339 (2023)
13. J. F. Liu, X. F. Zou, X. Wang, Y. Chen, Z. B. Rong, Z. M. Huang, S. G. Zheng, X. Y. Liang, and J. X. Wu, Discussion on the initial states of controlled bidirectional quantum secure direct communication, *Quantum Inform. Process.* 22(12), 426 (2023)
14. K. X. Liang, Z. W. Cao, X. L. Chen, L. Wang, G. Chai, and J. Y. Peng, A quantum secure direct communication scheme based on intermediate-basis, *Front. Phys. (Beijing)* 18(5), 51301 (2023)
15. D. Pan, G. L. Long, L. G. Yin, Y. B. Sheng, D. Ruan, S. X. Ng, J. H. Lu, and L. Hanzo, The evolution of quantum secure direct communication: On the road to the qinternet, *IEEE Commun. Surv. Tutor.* 26(3), 1898 (2024)
16. Q. Zhang, M. M. Du, W. Zhong, Y. B. Sheng, and L. Zhou, Single-photon based three-party quantum secure direct communication with identity authentication, *Ann. Phys. (Berlin)* 536, 2300407 (2024)
17. X. F. Zou, X. Wang, S. G. Zheng, Z. B. Rong, Z. M. Huang, Y. Chen, J. F. Liu, X. Y. Liang, and J. X. Wu, Problems of a quantum secure direct communication scheme based on intermediate-basis, *Quantum Inform. Process.* 23(6), 218 (2024)
18. P. Zhao, W. Zhong, M. M. Du, X. Y. Li, L. Zhou, and Y. B. Sheng, Quantum secure direct communication with hybrid entanglement, *Front. Phys. (Beijing)* 19(5), 51201 (2024)
19. C. Liu, C. Zhang, S. P. Gu, X. F. Wang, L. Zhou, and Y. B. Sheng, Receiver-device-independent quantum secure direct communication, *Sci. China Phys. Mech. Astron.* 68(5), 250311 (2025)
20. L. Wang, G. Chai, Z. W. Cao, and X. L. Chen, Highly reliable quantum secure direct communication based on concatenated GKP- QLDPC codes, *Sci. China Inf. Sci.* (2026)
21. G. Chai, Z. W. Cao, W. Q. Liu, M. H. Zhang, K. X. Liang, and J. Y. Peng, Novel continuous-variable quantum secure direct communication and its security analysis,

*Laser Phys. Lett.* 16(9), 095207 (2019)
22. Z. W. Cao, L. Wang, K. X. Liang, G. Chai, and J. Y. Peng, Continuous-variable quantum secure direct communication based on Gaussian mapping, *Phys. Rev. Appl.* 16(2), 024012 (2021)
23. Z. Y. Zuo, Z. T. Liang, N. Y. Mao, Y. J. Wang, and Y. Guo, Continuous-variable quantum secure direct communication against dual-sequence Gaussian attacks with quantum memory, *Sci. China Phys. Mech. Astron.* 69(4), 240314 (2026)
24. K. Wen and G. L. Long, One-party quantum-error-correcting codes for unbalanced errors: principles and application to quantum dense coding and quantum secure direct communication, *Int. J. Quant. Inf.* 8(4), 697 (2010)
25. C. W. Ding, W. Y. Wang, W. D. Zhang, L. Zhou, and Y. B. Sheng, Quantum secure direct communication based on quantum error correction code, *Appl. Phys. Lett.* 126(2), 024002 (2025)
26. F. G. Deng, X. H. Li, C. Y. Li, P. Zhou, and H. Y. Zhou, Quantum secure direct communication network with Einstein–Podolsky–Rosen pairs, *Phys. Lett. A* 359(5), 359 (2006)
27. F. G. Deng, X. H. Li, C. Y. Li, P. Zhou, and H. Y. Zhou, Quantum secure direct communication network with superdense coding and decoy photons, *Phys. Scr.* 76(1), 25 (2007)
28. L. H. Gong, Y. Liu, and N. R. Zhou, Novel quantum virtual private network scheme for PON via quantum secure direct communication, *Int. J. Theor. Phys.* 52(9), 3260 (2013)
29. C. Shukla, J. ur Rehman, and S. Chatzinotas, Quantum network applications in 6G paradigm, *AAPPS Bull.* 35(1), 28 (2025)
30. Z. Sun, L. Y. Song, Q. Huang, L. G. Yin, G. L. Long, J. H. Lu, and L. Hanzo, Toward practical quantum secure direct communication: A quantum-memory-free protocol and code design, *IEEE Trans. Commun.* 68(9), 5778 (2020)
31. J. Y. Hu, B. Yu, M. Y. Jing, L. T. Xiao, S. T. Jia, G. Q. Qin, and G. L. Long, Experimental quantum secure direct communication with single photons, *Light Sci. Appl.* 5(9), e16144 (2016)
32. H. R. Zhang, Z. Sun, R. Y. Qi, L. G. Yin, G. L. Long, and J. H. Lu, Realization of quantum secure direct communication over 100 km fiber with time-bin and phase quantum states, *Light Sci. Appl.* 11(1), 83 (2022)
33. X. Liu, D. Luo, G. S. Lin, Z. H. Chen, C. F. Huang, S. Z. Li, C. X. Zhang, Z. R. Zhang, and K. J. Wei, Fiber-based quantum secure direct communication without active polarization compensation, *Sci. China Phys. Mech. Astron.* 65(12), 120311 (2022)
34. D. Pan, Y. C. Liu, P. H. Niu, H. R. Zhang, F. H. Zhang, M. Wang, X. T. Song, X. W. Chen, C. Zheng, and G. L. Long, Simultaneous transmission of information and key exchange using the same photonic quantum states, *Sci. Adv.* 11(8), eadt4627 (2025)
35. F. Zhu, W. Zhang, Y. B. Sheng, and Y. D. Huang, Experimental long distance quantum secure direct communication, *Sci. Bull. (Beijing)* 62(22), 1519 (2017)
36. W. Zhang, D. S. Ding, Y. B. Sheng, L. Zhou, B. S. Shi, and G. C. Guo, Quantum secure direct communication with quantum memory, *Phys. Rev. Lett.* 118(22), 220501 (2017)
37. Z. W. Cao, Y. Lu, G. Chai, H. Yu, and K. X. Liang, Realization of quantum secure direct communication with continuous variable, *Research* 6, 0193 (2023)
38. I. Paparelle, F. Mousavi, F. Scazza, A. Bassi, M. Paris, and A. Zavatta, Experimental direct quantum communication with squeezed states, *Opt. Express* 33(14), 28917 (2025)
39. L. Wang, G. Chai, Z. W. Cao, X. L. Chen, K. X. Liang, and J. Y. Peng, Quantum purification for coherent states and its application, *Sci. China Phys. Mech. Astron.* 68(2), 220313 (2025)
40. Z. T. Qi, Y. H. Li, W. Y. Huang, J. Feng, Y. L. Zheng, and X. F. Chen, A 15-user quantum secure direct communication network, *Light Sci. Appl.* 10(1), 183 (2021)
41. Y. L. Yang, Y. H. Li, H. Li, C. N. Wu, Y. L. Zheng, and X. F. Chen, A 300-km fully-connected quantum secure direct communication network, *Sci. Bull. (Beijing)* 70(9), 1445 (2025)
42. P. H. Niu, Z. R. Zhou, Z. S. Lin, Y. B. Sheng, L. G. Yin, and G. L. Long, Measurement-device-independent quantum communication without encryption, *Sci. Bull. (Beijing)* 63(20), 1345 (2018)
43. Z. Gao, T. Li, and Z. Li, Long-distance measurement-device-independent quantum secure direct communication, *Europhys. Lett.* 125(4), 40004 (2019)
44. Z. R. Zhou, Y. B. Sheng, P. H. Niu, L. G. Yin, G. L. Long, and L. Hanzo, Measurement-device-independent quantum secure direct communication, *Sci. China Phys. Mech. Astron.* 63(3), 230362 (2020)
45. Z. K. Zou, L. Zhou, W. Zhong, and Y. B. Sheng, Measurement-device-independent quantum secure direct communication of multiple degrees of freedom of a single photon, *Europhys. Lett.* 131(4), 40005 (2020)
46. X. D. Wu, L. Zhou, W. Zhong, and Y. B. Sheng, High-capacity measurement-device-independent quantum secure direct communication, *Quantum Inform. Process.* 19(10), 354 (2020)
47. J. W. Ying, L. Zhou, W. Zhong, and Y. B. Sheng, Measurement-device-independent one-step quantum secure direct communication, *Chin. Phys. B* 31(12), 120303 (2022)
48. Y. P. Hong, L. Zhou, W. Zhong, and Y. B. Sheng, Measurement-device-independent three-party quantum secure direct communication, *Quantum Inform. Process.* 22(2), 111 (2023)
49. J. F. Liu, X. F. Zou, X. Wang, Y. Chen, Z. B. Rong, Z. M. Huang, S. G. Zheng, X. Y. Liang, and J. X. Wu, Applying a class of general maximally entangled states in measurement-device-independent quantum secure direct communication, *Phys. Rev. Appl.* 21(4), 044010 (2024)
50. C. Zhang, L. Zhou, W. Zhong, M. M. Du, and Y. B. Sheng, Measurement-device-independent quantum dialogue based on entanglement swapping and phase encoding, *Quantum Inform. Process.* 23(2), 52 (2024)
51. L. Zhou, Y. B. Sheng, and G. L. Long, Device-independent quantum secure direct communication against collective

attacks, *Sci. Bull. (Beijing)* 65(1), 12 (2020)
52. L. Zhou and Y. B. Sheng, One-step device-independent quantum secure direct communication, *Sci. China Phys. Mech. Astron.* 65(5), 250311 (2022)
53. L. Zhou, B. W. Xu, W. Zhong, and Y. B. Sheng, Device-independent quantum secure direct communication with single-photon sources, *Phys. Rev. Appl.* 19(1), 014036 (2023)
54. P. Roy, S. Bera, S. Gupta, and A. S. Majumdar, Device-independent quantum secure direct communication under non-Markovian quantum channels, *Quantum Inform. Process.* 23(5), 170 (2024)
55. H. Zeng, M. M. Du, W. Zhong, L. Zhou, and Y. B. Sheng, High-capacity device-independent quantum secure direct communication based on hyper-encoding, *Fundam. Res. (Beijing)* 4(4), 851 (2024)
56. 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(23), 230501 (2007)
57. S. Pironio, A. Acín, N. Brunner, N. Gisin, S. Massar, and V. Scarani, Device independent quantum key distribution secure against collective attacks, *New J. Phys.* 11(4), 045021 (2009)
58. D. P. Nadlinger, P. Drmota, B. C. Nichol, G. Araneda, D. Main, R. Srinivas, D. M. Lucas, C. J. Ballance, K. Ivanov, E. Y. Z. Tan, P. Sekatski, R. L. Urbanke, R. Renner, N. Sangouard, and J. D. Bancal, Experimental quantum key distribution certified by Bell's theorem, *Nature* 607(7920), 682 (2022)
59. W. Zhang, T. van Leent, K. Redeker, R. Garthoff, R. Schwonnek, F. Fertig, S. Eppelt, W. Rosenfeld, V. Scarani, C. C. W. Lim, and H. Weinfurter, A device-independent quantum key distribution system for distant users, *Nature* 607(7920), 68 (2022)
60. W. Z. Liu, Y. Z. Zhang, Y. Z. Zhen, M. H. Li, Y. Liu, J. Y. Fan, F. H. Xu, Q. Zhang, and J. W. Pan, Toward a photonic demonstration of device-independent quantum key distribution, *Phys. Rev. Lett.* 129(5), 050502 (2022)
61. J. W. Wu, Z. S. Lin, L. G. Yin, and G. L. Long, Security of quantum secure direct communication based on Wyner's wiretap channel theory, *Quantum Eng.* 1(4), e26 (2019)
62. R. Y. Qi, Z. Sun, Z. S. Lin, P. H. Niu, W. T. Hao, L. Y. Song, Q. Huang, J. C. Gao, L. G. Yin, and G. L. Long, Implementation and security analysis of practical quantum secure direct communication, *Light Sci. Appl.* 8(1), 22 (2019)
63. D. Pan, Z. Lin, J. W. Wu, H. R. Zhang, Z. Sun, D. Ruan, L. G. Yin, and G. L. Long, Experimental free-space quantum secure direct communication and its security analysis, *Photon. Res.* 8(9), 1522 (2020)
64. 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(13), 2023 (2001)
65. N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, Trojan-horse attacks on quantum-key-distribution systems, *Phys. Rev. A* 73(2), 022320 (2006)
66. 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(3), 168 (2015)
67. 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(3), 031030 (2015)
68. S. E. Vinay and P. Kok, Extended analysis of the Trojan-horse attack in quantum key distribution, *Phys. Rev. A* 97(4), 042335 (2018)
69. I. S. Sushchev, D. S. Bulavkin, K. E. Bugai, A. S. Sidelnikova, and D. A. Dvoretskiy, Trojan-horse attack on a real-world quantum key distribution system: Theoretical and experimental security analysis, *Phys. Rev. Appl.* 22(3), 034032 (2024)
70. Y. Fu, H. L. Yin, T. Y. Chen, and Z. B. Chen, Long-distance measurement-device-independent multiparty quantum communication, *Phys. Rev. Lett.* 114(9), 090501 (2015)
71. Y. R. Xiao, H. L. Yin, W. J. Hua, X. Y. Cao, and Z. B. Chen, Experimental efficient source-independent quantum secret sharing against coherent attacks, *Phys. Rev. Lett.* 135(15), 150801 (2025)
72. Y. S. Lu, H. L. Yin, Y. M. Xie, Y. Fu, and Z. B. Chen, Repeater-like asynchronous measurement-device-independent quantum conference key agreement, *Rep. Prog. Phys.* 88(6), 067901 (2025)
73. Y. F. Du, Y. F. Liu, C. D. Yang, X. D. Zheng, S. N. Zhu, and X. S. Ma, Experimental measurement-device-independent quantum cryptographic conferencing, *Phys. Rev. Lett.* 134(4), 040802 (2025)
74. R. Jozsa and J. Schlienz, Distinguishability of states and von Neumann entropy, *Phys. Rev. A* 62(1), 012301 (2000)
75. D. Gottesman, H. K. Lo, N. Lütkenhaus, and J. Preskill, Security of quantum key distribution with imperfect devices, *Quantum Inf. Comput.* 4(5), 325 (2004)
76. M. Koashi, Simple security proof of quantum key distribution based on complementarity, *New J. Phys.* 11(4), 045018 (2009)
77. H. Tan, W. Li, L. K. Zhang, K. J. Wei, and F. H. Xu, Chip-based quantum key distribution against trojan-horse attack, *Phys. Rev. Appl.* 15(6), 064038 (2021)
78. A. Ponosova, D. Ruzhitskaya, P. Chaiwongkhot, V. Egorov, V. Makarov, and A. Q. Huang, Protecting fiber-optic quantum key distribution sources against light-injection attacks, *PRX Quantum* 3(4), 040307 (2022)
79. A. Q. Huang, R. P. Li, V. Egorov, S. Tchouragoulov, K. Kumar, and V. Makarov, Laser-damage attack against optical attenuators in quantum key distribution, *Phys. Rev. Appl.* 13(3), 034017 (2020)
80. J. Gu, X. Y. Cao, Y. Fu, Z. W. He, Z. J. Yin, H. L. Yin, and Z. B. Chen, Experimental measurement-device-independent type quantum key distribution with flawed and correlated sources, *Sci. Bull. (Beijing)* 67(21), 2167 (2022)
81. X. Y. Cao, X. R. Sun, M. Y. Li, Y. S. Lu, H. L. Yin, and Z. B. Chen, Experimental coherent one-way quantum key distribution with simplicity and practical security, *Sci. Adv.* 12(1), eaec2776 (2026)
82. Z. J. Li, B. B. Zheng, C. X. Zhang, Z. R. Zhang, H. B.

Xie, and K. J. Wei, Improved security bounds against the Trojan-horse attack in decoy-state quantum key distribution, *Quantum Inform. Process.* 23(2), 40 (2024)
83. J. W. Ying, Q. Zhang, S. P. Gu, X. F. Wang, L. Zhou, and Y. B. Sheng, Fully passive quantum secure direct communication. arXiv: 2502.12652 (2025)
84. M. Avesani, C. Agnesi, A. Stanco, G. Vallone, and P. Villoresi, Stable, low-error, and calibration-free polarization encoder for free-space quantum communication, *Opt. Lett.* 45(17), 4706 (2020)
85. A. De Toni, A. C. Aka, C. Agnesi, D. G. Marangon, G. Vallone, and P. Villoresi, Countermeasures for Trojan-horse attacks on self-compensating all-fiber polarization modulator, arXiv: 2510.16868 (2025)
86. W. Y. Wang, R. Wang, C. Q. Hu, V. Zapatero, L. Qian, B. Qi, M. Curty, and H. K. Lo, Fully passive quantum key distribution, *Phys. Rev. Lett.* 130(22), 220801 (2023)
87. C. Q. Hu, W. Y. Wang, K. S. Chan, Z. H. Yuan, and H. K. Lo, Proof-of-principle demonstration of fully passive quantum key distribution, *Phys. Rev. Lett.* 131(11), 110801 (2023)
88. X. Wang, F. Y. Lu, Z. H. Wang, Z. Q. Yin, S. Wang, J. Q. Geng, W. Chen, D. Y. He, G. C. Guo, and Z. F. Han, Fully passive measurement-device-independent quantum key distribution, *Phys. Rev. Appl.* 21(6), 064067 (2024)
89. J. J. Li, W. Y. Wang, and H. K. Lo, Fully passive measurement-device-independent quantum key distribution, *Phys. Rev. Appl.* 21(6), 064056 (2024)
90. J. W. Ying, Q. Zhang, S. P. Gu, X. F. Wang, L. Zhou, and Y. B. Sheng, Fully passive quantum key distribution with parametric down-conversion source, arXiv: 2502.12651 (2025)
91. J. W. Ying, S. P. Gu, X. F. Wang, L. Zhou, and Y. B. Sheng, Fully passive reference frame independent quantum key distribution, *Sci. China Inf. Sci.*, doi: 10.1007/s11432-025-4759-x (2025)
92. J. W. Ying, J. Y. Wang, Y. X. Xiao, S. P. Gu, X. F. Wang, L. Zhou, and Y. B. Sheng, Passive decoy state quantum secure direct communication, *Fundam. Res. (Beijing)*, doi: 10.1016/j.fmre.2025.12.008 (2025)
93. J. W. Wu, G. L. Long, and M. Hayashi, Quantum secure direct communication with private dense coding using a general preshared quantum state, *Phys. Rev. Appl.* 17(6), 064011 (2022)
94. Z. Z. Sun, D. Pan, Y. B. Cheng, Y. C. Liu, D. Ruan, and G. L. Long, Multi-intensity quantum secure direct communication relying on finite block-length, *IEEE Trans. Commun.* 72(8), 4633 (2024)
95. K. Tamaki, M. Koashi, and N. Imoto, Unconditionally secure key distribution based on two nonorthogonal states, *Phys. Rev. Lett.* 90(16), 167904 (2003)
96. M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, Simulating positive-operator-valued measures with projective measurements, *Phys. Rev. Lett.* 119(19), 190501 (2017)
97. M. Pereira, G. Kato, A. Mizutani, M. Curty, and K. Tamaki, Quantum key distribution with correlated sources, *Sci. Adv.* 6(37), eaaz4487 (2020)