Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Finite-Key Security Analysis of MP-QKD under Trojan-horse Attacks

Fangze Ma
Kaiyi Shi
Chen Shen
Tianyu Xu
Haiqiang Ma
hqma@bupt.edu.cn

Research Article

Keywords:

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10440122/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Finite-Key Security Analysis of MP-QKD under Trojan–horse Attacks

Fangze Ma[1,2], Kaiyi Shi[1,2], Chen Shen[1,2], Tianyu Xu[1,2], Haiqiang Ma[1,2*]

[1*]School of Physical Science and Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, People's Republic of China.
[2*]State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing, 100876, People's Republic of China.

*Corresponding author(s). E-mail(s): hqma@bupt.edu.cn;
Contributing authors: 15647605165@163.com; 2023010488@bupt.cn ;
shenchen1411@126.com ; 906406518@qq.com ;

Abstract

Trojan-horse attack (THA) pose a serious threat to the practical security of quantum key distribution by leaking source-encoding information through back-reflected probe light. Although mode-pairing QKD (MP-QKD) can overcome the conventional rate–distance limit, its finite-key security under THA has not been systematically investigated. In this work, we develop a unified framework for analyzing the finite-key security of decoy-state MP-QKD with source-side leakage. Based on the reference-state method and the Cauchy–Schwarz inequality, we introduce a leakage parameter to quantify the deviation between practical emitted states and ideal reference states, and derive statistical constraints that enable valid decoy-state parameter estimation. We further obtain an explicit and computable finite-key secret-key-rate formula by incorporating statistical fluctuation analysis. Numerical simulations show that both side-channel leakage and finite-size effects reduce the achievable secret-key rate and maximum secure transmission distance, while the degradation becomes less significant for weak leakage or a sufficiently large number of transmitted pulses. These results provide a rigorous basis for evaluating and improving the practical security of MP-QKD systems against Trojan-horse attacks.

Keywords: Mode-pairing quantum key distribution; Trojan-horse attack; finite-key security; side-channel leakage; decoy-state method

1 Introduction

Quantum key distribution (QKD) enables information-theoretically secure key sharing based on the fundamental principles of quantum mechanics [1–6], and is regarded as a key technology for future information security. In recent years, with the development of experimental techniques, various QKD protocols have been proposed and extensively studied. Among them, mode-pairing quantum key distribution (MP-QKD) has attracted considerable attention due to its ability to overcome transmission loss limitations and achieve improved key rate performance [7–14].

Most existing theoretical analyses assume ideal devices and the absence of information leakage. However, in practical implementations, various side-channel attacks pose serious threats to the security of QKD systems [15]. Among these, the Trojan–horse attack (THA) [16–20] is a typical active side-channel attack, in which an eavesdropper injects probing light into the transmitter and extracts encoding information by analyzing the back-reflected signals. For conventional QKD protocols such as BB84, the security under THA has been studied [21], and countermeasures based on the reference-state method have been proposed. In contrast, for MP-QKD, systematic security analysis in the presence of THA remains relatively limited, particularly in the finite-key regime where computable expressions for the secure key rate are still lacking [22–26].

To characterize the deviation between practical emitted states and ideal states, the reference-state method constructs a virtual protocol corresponding to the actual implementation and establishes statistical relations between them. This method provides an effective tool for security analysis under side-channel conditions. Furthermore, the Cauchy–Schwarz inequality can be employed to derive constraints between the observed statistics of the actual protocol and those of the reference protocol, enabling the estimation of key parameters even in the presence of information leakage. However, the systematic application of these techniques to MP-QKD, together with the incorporation of the decoy-state method to obtain computable key rate expressions, remains to be further explored.

To address these issues, this work investigates the security of the MP-QKD protocol under Trojan–horse attacks within a unified analytical framework based on the reference-state method and the Cauchy–Schwarz inequality. A leakage parameter is introduced to characterize the deviation between practical emitted states and ideal reference states caused by information leakage. By establishing statistical constraints between the actual and reference protocols, computable bounds on key parameters are derived, enabling the application of the decoy-state method under side-channel conditions. Furthermore, the analysis is extended to the finite-key regime by incorporating statistical fluctuation analysis, leading to an explicit expression for the secure key rate. Numerical simulations are finally performed to evaluate the effects of leakage strength and finite-key size on system performance.

The remainder of this paper is organized as follows. Section 2 introduces the Trojan–horse attack model in MP-QKD. Section 3 presents the security analysis framework and derives the finite-key secret key rate under Trojan–horse attacks. Section 4 provides numerical simulations and discussion. Finally, Section 5 concludes the paper and discusses future directions.

2 Mechanism of Trojan–horse Attacks in the MP-QKD Protocol

In the standard security analysis of the MP-QKD protocol, the transmitters are typically assumed to be perfectly isolated and free from information leakage. That is, the quantum states prepared by Alice and Bob are assumed to be transmitted only through the quantum channel to the untrusted relay Charlie, without leaking any information about the encoding choices to a potential eavesdropper. Under this ideal assumption, the emitted states are independent of the adversary's system, and the decoy-state method can be directly applied for parameter estimation.

However, in practical implementations, this assumption is generally not strictly satisfied due to imperfections of physical devices such as modulators and optical isolators. An eavesdropper (Eve) can perform a Trojan–horse attack (THA) by injecting probing light into the transmitters and analyzing the back-reflected signals, thereby obtaining partial information about the encoding choices, such as the intensity or phase settings. As a result, correlations are introduced between the emitted states and Eve's side-channel system, which invalidates the independence assumption used in conventional security analyses [27, 28].

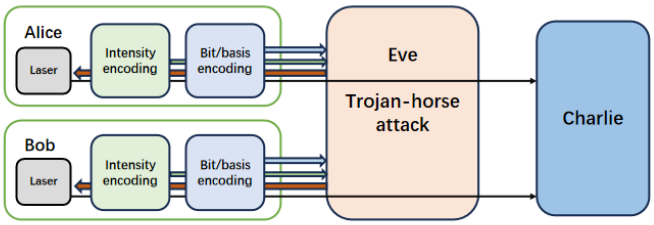


Fig. 1 Schematic illustration of the Trojan–horse attack considered in the MP-QKD protocol. Eve injects probing light into Alice's and Bob's transmitters and collects the back-reflected signals after they pass through the intensity and bit/basis encoding modules. The reflected light may carry information about the users' modulation settings, while the legitimate quantum signals are sent to the untrusted relay Charlie. The blue, green, orange, and gray lines schematically represent optical signals corresponding to different combinations of intensity and bit/basis settings. The colors are used only for visual distinction and do not indicate different wavelengths or physical channels.

In an experimental implementation, the existence of back-reflected light alone does not necessarily imply encoding-information leakage. Leakage occurs when the state of the back-reflected light depends on Alice's or Bob's modulation settings. For example, different intensity or phase settings may produce distinguishable changes in the reflected optical power, phase, temporal waveform, spectrum, or polarization. If the reflected states are identical for all encoding settings, Eve cannot infer the selected setting from the returned light, corresponding to the ideal no-leakage case.

In the ideal case without information leakage, the quantum state prepared by Alice and Bob in the i-th round can be expressed as an independent product state

$$|\Phi_i\rangle_{AB} = |\alpha_i\rangle_A \otimes |\beta_i\rangle_B, \tag{1}$$

where $|\alpha_i\rangle_A$ and $|\beta_i\rangle_B$ denote the weak coherent states prepared by Alice and Bob, respectively.

In the presence of a Trojan–horse attack, Eve injects probing light into the transmitters and obtains partial information from the back-reflected signals. The emitted state is then no longer restricted to the signal systems, but should be described as a joint state of the signal systems and Eve's side-channel systems:

$$|\Psi_i\rangle_{ABE} = |\alpha_i\rangle_A|\beta_i\rangle_B \otimes |\chi^A_{\mu_i,\phi_i}\rangle_{E_A} \otimes |\chi^B_{\mu_i,\phi_i}\rangle_{E_B}, \tag{2}$$

where $|\chi^A_{\mu_i,\phi_i}\rangle$ and $|\chi^B_{\mu_i,\phi_i}\rangle$ represent the back-reflected states obtained by Eve from Alice's and Bob's transmitters, respectively, which generally depend on the intensity and phase modulation settings.

In the MP-QKD protocol, key generation is based on the pairing of two successful rounds. Therefore, for a paired event (i, j), the emitted state is given by

$$|\Psi_{ij}\rangle = |\Psi_i\rangle \otimes |\Psi_j\rangle. \tag{3}$$

This shows that, under Trojan–horse attacks, the emitted states are extended from independent signal states to joint states including side-channel systems, thereby modifying the fundamental assumptions of the security analysis [29].

Equations (2) and (3) show that, under Trojan–horse attacks, the emitted states are no longer independent signal states, but joint states including Eve's side-channel systems. Since the back-reflected light may be correlated with the intensity and phase settings, Eve can partially distinguish different encoding choices, which invalidates the ideal-source assumption used in conventional decoy-state analysis. As a result, the estimation of key parameters, such as the lower bound on the number of single-photon effective Z-pairs and the upper bound on the corresponding phase error rate, may be biased. Moreover, because the raw key in MP-QKD is generated after the mode-pairing step, the security analysis should be formulated at the level of valid paired events rather than individual transmission rounds. Therefore, a leakage parameter is introduced in the next section to quantify the deviation between practical emitted states and ideal reference states [30, 31].

3 Security Analysis under Trojan–Horse Attacks

In this section, we prove the finite-key security of the MP-QKD protocol in the presence of Trojan–horse attacks. Different from conventional MDI-QKD, the basic object used for parameter estimation in MP-QKD is not a single transmission round, but a valid mode pair formed by two effective detection events. Therefore, all statistical quantities below are defined at the level of paired events. The proof follows the

reference-state method: we first introduce a virtual reference protocol without side-channel information leakage, then relate the practical protocol to the reference protocol by the Cauchy–Schwarz constraint, and finally combine the decoy-state method with finite-key statistical bounds to estimate the single-photon contribution and the phase error rate.

3.1 Cauchy–Schwarz constraint

In the practical MP-QKD protocol, Alice and Bob send phase-randomized coherent states to the untrusted measurement node Charlie. After Charlie announces the effective detection events, two such events, say the ith and jth rounds, may be grouped into a valid mode pair $P^{i,j}$. For Alice, the corresponding settings of this pair are denoted by

$$\mathbf{k}_A^{i,j} = \left(k_A^i, k_A^j, \theta_A^i, \theta_A^j\right), \tag{4}$$

and similarly for Bob,

$$\mathbf{k}_B^{i,j} = \left(k_B^i, k_B^j, \theta_B^i, \theta_B^j\right), \tag{5}$$

where $k_A^i, k_A^j \in \{\mu_A, \nu_A, 0\}$ and $k_B^i, k_B^j \in \{\mu_B, \nu_B, 0\}$ are the intensity settings, while $\theta_A^i, \theta_A^j, \theta_B^i, \theta_B^j$ are the random phases.

When Eve performs a Trojan–horse attack, the back-reflected light may carry information about Alice's and Bob's intensity and phase settings. Therefore, the actual state associated with the pair $P^{i,j}$ should be written as

$$\left|\Psi^{i,j}\right\rangle = \sum_{\mathbf{k}_A^{i,j},\mathbf{k}_B^{i,j}} \sqrt{p_{\mathbf{k}_A^{i,j}} p_{\mathbf{k}_B^{i,j}}} \left|R_{\mathbf{k}_A^{i,j}}\right\rangle_A \left|R_{\mathbf{k}_B^{i,j}}\right\rangle_B \left|\psi^{i,j}_{\mathbf{k}_A^{i,j},\mathbf{k}_B^{i,j}}\right\rangle_{CPE}, \tag{6}$$

where C denotes the optical modes sent to Charlie, P denotes the purifying systems, and E denotes Eve's side-channel systems generated by the back-reflected light.

To analyze the security, we define a reference state

$$\left|\Phi^{i,j}\right\rangle = \sum_{\mathbf{k}_A^{i,j},\mathbf{k}_B^{i,j}} \sqrt{p_{\mathbf{k}_A^{i,j}} p_{\mathbf{k}_B^{i,j}}} \left|R_{\mathbf{k}_A^{i,j}}\right\rangle_A \left|R_{\mathbf{k}_B^{i,j}}\right\rangle_B \left|\phi^{i,j}_{\mathbf{k}_A^{i,j},\mathbf{k}_B^{i,j}}\right\rangle_{CP} |\tau\rangle_E . \tag{7}$$

Here $|\tau\rangle_E$ is a fixed state independent of all intensity and phase settings. Thus, the reference protocol has no side-channel information leakage, and the standard finite-key decoy-state analysis of MP-QKD can be applied to it.

The deviation between the practical state and the reference state is characterized by

$$\delta_{i,j} = \left|\left\langle \Phi^{i,j} \middle| \Psi^{i,j} \right\rangle\right|. \tag{8}$$

For a conservative analysis, we define

$$\delta = \min_{(i,j)\in\mathcal{P}} \delta_{i,j}, \tag{9}$$

where $\mathcal{P}$ is the set of valid paired events. Then, for any valid pair,

$$\left|\left\langle \Phi^{i,j}\middle|\Psi^{i,j}\right\rangle\right| \geq \delta. \tag{10}$$

The parameter δ quantifies the overlap between the practical state containing the back-reflected side-channel system and the ideal reference state without setting-dependent leakage. Therefore, δ should be understood as an indistinguishability parameter rather than the leaked information itself. When $\delta = 1$, the practical and reference states coincide, corresponding to the ideal case in which the reflected light is independent of the intensity and phase settings. A smaller value of δ indicates that the reflected states associated with different modulation settings become more distinguishable, and thus that Eve may obtain more side-channel information. In practice, a conservative lower bound on δ can be obtained by characterizing setting-dependent variations in the back-reflected optical power, phase, temporal waveform, spectrum, or polarization. These variations are affected by the injected probe intensity, transmitter isolation, internal optical loss, and the accuracy of monitoring devices.

Let $\hat{M}$ be an arbitrary POVM element corresponding to a post-selected event in MP-QKD, such as a Z-pair event, an X-pair event, a vacuum-related event, or an error event. We define

$$p = \left\langle \Psi^{i,j}\middle|\, \hat{M}\, \middle|\Psi^{i,j}\right\rangle, \qquad p^{\mathrm{ref}} = \left\langle \Phi^{i,j}\middle|\, \hat{M}\, \middle|\Phi^{i,j}\right\rangle. \tag{11}$$

Using the Cauchy–Schwarz inequality, one obtains

$$\sqrt{pp^{\mathrm{ref}}} + \sqrt{(1-p)(1-p^{\mathrm{ref}})} \geq \delta. \tag{12}$$

Solving this inequality gives

$$G_\delta^L(p) \leq p^{\mathrm{ref}} \leq G_\delta^U(p), \tag{13}$$

where

$$G_\delta^L(p) = \begin{cases} g_\delta^-(p), & p > 1-\delta^2, \\ 0, & \text{otherwise}, \end{cases} \tag{14}$$

and

$$G_\delta^U(p) = \begin{cases} g_\delta^+(p), & p < \delta^2, \\ 1, & \text{otherwise}. \end{cases} \tag{15}$$

Here

$$g_\delta^\pm(p) = p + (1-\delta^2)(1-2p) \pm 2\delta\sqrt{(1-\delta^2)p(1-p)}. \tag{16}$$

When $\delta = 1$, the practical and reference protocols coincide. When $\delta < 1$, the interval between G_δ^L and G_δ^U quantifies the additional uncertainty caused by the Trojan–horse side-channel leakage.

3.2 Decoy-state method

We now estimate the two quantities required in the finite-key formula of MP-QKD: the lower bound n_{Z1}^{L} on the number of single-photon effective Z-pairs in the raw key, and the upper bound $e_{Z1}^{\mathrm{ph,U}}$ on their phase error rate.

For a valid pair $P^{i,j}$, define the total intensity labels

$$s_A = k_A^i + k_A^j, \qquad s_B = k_B^i + k_B^j, \tag{17}$$

and write $\mathbf{s} = (s_A, s_B)$. After basis sifting, Alice and Bob obtain three relevant sets of pairs, denoted by Z, X, and V, corresponding to Z-pairs, X-pairs, and vacuum-related pairs, respectively. Let $n_{\Omega,\mathbf{s}}$ be the observed number of pairs in the set $\Omega \in \{Z, X, V\}$ with intensity label $\mathbf{s}$, and let $m_{X,\mathbf{s}}$ be the number of bit-error events among the corresponding X-pairs.

For the practical protocol, define

$$Q_{\Omega,\mathbf{s}} = \frac{n_{\Omega,\mathbf{s}}}{N_{\mathrm{pair}}}, \qquad \Omega \in \{Z, X, V\}, \tag{18}$$

and

$$T_{X,\mathbf{s}} = \frac{m_{X,\mathbf{s}}}{N_{\mathrm{pair}}}, \tag{19}$$

where N_{pair} is the total number of valid paired events. The corresponding reference quantities are denoted by $Q_{\Omega,\mathbf{s}}^{\mathrm{ref}}$ and $T_{X,\mathbf{s}}^{\mathrm{ref}}$. According to the Cauchy–Schwarz constraint,

$$G_\delta^L\left(Q_{\Omega,\mathbf{s}}\right) \le Q_{\Omega,\mathbf{s}}^{\mathrm{ref}} \le G_\delta^U\left(Q_{\Omega,\mathbf{s}}\right), \tag{20}$$

and

$$G_\delta^L\left(T_{X,\mathbf{s}}\right) \le T_{X,\mathbf{s}}^{\mathrm{ref}} \le G_\delta^U\left(T_{X,\mathbf{s}}\right). \tag{21}$$

In the finite-key regime, statistical fluctuations should be included. Let $\mathcal{K}_{N,\epsilon}^{L}(\cdot)$ and $\mathcal{K}_{N,\epsilon}^{U}(\cdot)$ denote the lower and upper concentration bounds with failure probability ϵ. Then the reference statistics used in the decoy-state estimation are bounded by

$$Q_{\Omega,\mathbf{s}}^{\mathrm{ref,L}} = G_\delta^L\left[\frac{\mathcal{K}_{N_{\mathrm{pair}},\epsilon}^{L}\left(n_{\Omega,\mathbf{s}}\right)}{N_{\mathrm{pair}}}\right], \tag{22}$$

and

$$Q_{\Omega,\mathbf{s}}^{\mathrm{ref,U}} = G_\delta^U\left[\frac{\mathcal{K}_{N_{\mathrm{pair}},\epsilon}^{U}\left(n_{\Omega,\mathbf{s}}\right)}{N_{\mathrm{pair}}}\right], \tag{23}$$

where $\Omega \in \{Z, X, V\}$. Therefore,

$$Q_{\Omega,\mathbf{s}}^{\mathrm{ref,L}} \le Q_{\Omega,\mathbf{s}}^{\mathrm{ref}} \le Q_{\Omega,\mathbf{s}}^{\mathrm{ref,U}}. \tag{24}$$

Similarly, for the X-basis error statistics, we define

$$T_{X,\mathbf{s}}^{\mathrm{ref,L}} = G_\delta^L \left[\frac{\mathcal{K}_{N_{\mathrm{pair}},\epsilon}^L (m_{X,\mathbf{s}})}{N_{\mathrm{pair}}} \right], \tag{25}$$

and

$$T_{X,\mathbf{s}}^{\mathrm{ref,U}} = G_\delta^U \left[\frac{\mathcal{K}_{N_{\mathrm{pair}},\epsilon}^U (m_{X,\mathbf{s}})}{N_{\mathrm{pair}}} \right]. \tag{26}$$

Thus,

$$T_{X,\mathbf{s}}^{\mathrm{ref,L}} \leq T_{X,\mathbf{s}}^{\mathrm{ref}} \leq T_{X,\mathbf{s}}^{\mathrm{ref,U}}. \tag{27}$$

Since the reference protocol contains no side-channel leakage, the finite-key decoy-state method of MP-QKD can be applied to the reference statistics. In the following, F_Z^L, F_X^L, and F_e^U denote the standard finite-key decoy-state estimators of MP-QKD. Specifically, F_Z^L gives the lower bound on the number of single-photon effective Z-pairs, F_X^L gives the lower bound on the number of single-photon effective X-pairs, and F_e^U gives the upper bound on the single-photon bit-error rate of the X-pairs. Their explicit implementation can be realized by the Gaussian-elimination formulas or equivalently by the linear-programming constraints given in Appendix B.

Thus, the lower bound on the number of single-photon effective Z-pairs in the reference protocol is written as

$$n_{Z1}^{\mathrm{ref,L}} = F_Z^L \left(\left\{ Q_{Z,\mathbf{s}}^{\mathrm{ref,L}}, Q_{Z,\mathbf{s}}^{\mathrm{ref,U}}, Q_{V,\mathbf{s}}^{\mathrm{ref,L}}, Q_{V,\mathbf{s}}^{\mathrm{ref,U}} \right\}_{\mathbf{s}} \right), \tag{28}$$

where the input intervals are obtained from the Cauchy–Schwarz constraints above. Defining

$$q_{Z1}^{\mathrm{ref,L}} = \frac{n_{Z1}^{\mathrm{ref,L}}}{N_{\mathrm{pair}}}, \tag{29}$$

the Cauchy–Schwarz constraint implies

$$q_{Z1} \geq G_\delta^L \left(q_{Z1}^{\mathrm{ref,L}} \right). \tag{30}$$

Therefore, the lower bound on the practical single-photon contribution is

$$n_{Z1}^L = \mathcal{K}_{N_{\mathrm{pair}},\epsilon}^L \left[N_{\mathrm{pair}} G_\delta^L \left(\frac{n_{Z1}^{\mathrm{ref,L}}}{N_{\mathrm{pair}}} \right) \right]. \tag{31}$$

Next, we estimate the phase error rate. In MP-QKD, the phase error rate of single-photon Z-pairs is estimated from the bit-error statistics of single-photon X-pairs. The lower bound on the number of single-photon effective X-pairs in the reference protocol is

$$n_{X1}^{\mathrm{ref,L}} = F_X^L \left(\left\{ Q_{X,\mathbf{s}}^{\mathrm{ref,L}}, Q_{X,\mathbf{s}}^{\mathrm{ref,U}}, Q_{V,\mathbf{s}}^{\mathrm{ref,L}}, Q_{V,\mathbf{s}}^{\mathrm{ref,U}} \right\}_{\mathbf{s}} \right). \tag{32}$$

Similarly, the upper bound on the single-photon bit-error rate of reference X-pairs is

$$e_{X1}^{\mathrm{bit,ref,U}} = F_e^U\left(\left\{T_{X,\mathbf{s}}^{\mathrm{ref,L}}, T_{X,\mathbf{s}}^{\mathrm{ref,U}}, Q_{X,\mathbf{s}}^{\mathrm{ref,L}}, Q_{X,\mathbf{s}}^{\mathrm{ref,U}}, Q_{V,\mathbf{s}}^{\mathrm{ref,L}}, Q_{V,\mathbf{s}}^{\mathrm{ref,U}}\right\}_{\mathbf{s}}\right), \tag{33}$$

where $F_X^L(\cdot)$ and $F_e^U(\cdot)$ are the corresponding finite-key decoy-state estimators.

Using the random-sampling relation between the single-photon Z-pairs and X-pairs in MP-QKD, the phase error rate in the reference protocol satisfies

$$e_{Z1}^{\mathrm{ph,ref,U}} = e_{X1}^{\mathrm{bit,ref,U}} + \Gamma\left(\epsilon_{\mathrm{rs}}, e_{X1}^{\mathrm{bit,ref,U}}, n_{X1}^{\mathrm{ref,L}}, n_{Z1}^{\mathrm{ref,L}}\right), \tag{34}$$

where ϵ_{rs} is the failure probability of the random-sampling step, and

$$\Gamma(a,b,c,d) = \sqrt{\frac{(c+d)(1-b)b}{cd}\ln\left[\frac{c+d}{2\pi cd(1-b)ba^2}\right]}. \tag{35}$$

Finally, the practical phase-error probability is upper bounded through the Cauchy–Schwarz constraint. Let

$$q_{\mathrm{ph}}^{\mathrm{ref,U}} = \frac{n_{Z1}^{\mathrm{ref,L}}}{N_{\mathrm{pair}}} e_{Z1}^{\mathrm{ph,ref,U}}. \tag{36}$$

Then

$$q_{\mathrm{ph}} \le G_\delta^U\left(q_{\mathrm{ph}}^{\mathrm{ref,U}}\right), \tag{37}$$

and the number of phase errors in the practical protocol is bounded by

$$m_{\mathrm{ph}}^U = \mathcal{K}_{N_{\mathrm{pair}},\epsilon}^U\left[N_{\mathrm{pair}} G_\delta^U\left(q_{\mathrm{ph}}^{\mathrm{ref,U}}\right)\right]. \tag{38}$$

Hence, the upper bound on the practical phase error rate is

$$e_{Z1}^{\mathrm{ph,U}} = \min\left\{\frac{m_{\mathrm{ph}}^U}{n_{Z1}^L}, \frac{1}{2}\right\}. \tag{39}$$

3.3 Secret-key rate

After obtaining the lower bound n_{Z1}^L and the upper bound $e_{Z1}^{\mathrm{ph,U}}$, the finite-key secret key length can be derived under the universally composable security framework. Let n_Z be the length of the raw key generated from the retained Z-pairs, and let E_Z be the corresponding bit-error rate. The information revealed in error correction is

$$\lambda_{\mathrm{EC}} = f n_Z h(E_Z), \tag{40}$$

where f is the error-correction efficiency and

$$h(x) = -x\log_2 x - (1-x)\log_2(1-x) \tag{41}$$

is the binary entropy function.

If the error correction step is passed, the final secret key length satisfies

$$\ell_{\mathrm{THA}} \leq n_{Z1}^{L}\left[1-h\left(e_{Z1}^{\mathrm{ph,U}}\right)\right]-\lambda_{\mathrm{EC}}-\log_2\frac{2}{\epsilon_{\mathrm{cor}}}-2\log_2\frac{1}{\sqrt{2}\hat{\epsilon}\epsilon_{\mathrm{PA}}}. \tag{42}$$

Here ϵ_{cor} is the correctness parameter, $\hat{\epsilon}$ is the smoothing parameter, and ϵ_{PA} is the failure probability of privacy amplification. The finite-key secret key rate under THA is therefore

$$R_{\mathrm{fin}}^{\mathrm{THA}}=\frac{\ell_{\mathrm{THA}}}{N}, \tag{43}$$

where N is the total number of transmitted optical pulses.

The total security parameter is

$$\epsilon_{\mathrm{tol}}=\epsilon_{\mathrm{cor}}+\epsilon_{\mathrm{sec}}, \tag{44}$$

with

$$\epsilon_{\mathrm{sec}}=2\left(\hat{\epsilon}+2\sqrt{\epsilon_e+\epsilon_1}\right)+\epsilon_{\mathrm{PA}}. \tag{45}$$

Here ϵ_1 and ϵ_e are the failure probabilities associated with the estimation of n_{Z1}^{L} and $e_{Z1}^{\mathrm{ph,U}}$, respectively.

When $\delta = 1$, the functions G_{δ}^{L} and G_{δ}^{U} reduce to the identity map, and the above result recovers the standard finite-key formula of MP-QKD without side-channel leakage. When $\delta < 1$, the Cauchy–Schwarz constraints reduce the lower bound on the single-photon contribution and increase the upper bound on the phase error rate, which leads to a smaller secret key rate.

4 Numerical Results and Discussion

In this section, numerical simulations are performed to evaluate the finite-key performance of the MP-QKD protocol under Trojan–horse attacks. The simulations are based on a symmetric quantum channel model, and the system parameters are chosen according to typical experimental settings reported in previous studies.

In the numerical analysis, a symmetric configuration is considered, where the untrusted relay is located at the midpoint between Alice and Bob. The channel loss is modeled using the standard optical fiber loss model with a loss coefficient of 0.2 dB/km. The detector parameters include the dark count rate p_d, detection efficiency η_d, and intrinsic misalignment error rate e_d.

To reflect realistic experimental statistics, the observed quantities are generated at the level of valid mode pairs, consistent with the security analysis in Sec. III. For each pair-intensity label $\mathbf{s} = (s_A, s_B)$, the expected number of retained pairs in the set $\Omega \in \{Z, X, V\}$ is written as

$$n_{\Omega,\mathbf{s}}=N_{\mathrm{pair}}Q_{\Omega,\mathbf{s}}, \tag{46}$$

where $Q_{\Omega,\mathbf{s}}$ denotes the pair-level gain of the corresponding set. Similarly, the expected number of bit-error events among the X-pairs is given by

$$m_{X,\mathbf{s}}=N_{\mathrm{pair}}T_{X,\mathbf{s}}, \tag{47}$$

where $T_{X,\mathbf{s}}$ denotes the pair-level error gain. These quantities are then mapped to the corresponding reference statistics through the Cauchy–Schwarz constraints derived in Sec. III.

In the finite-key analysis, the security parameters are set as ϵ_{cor} and ϵ_{sec}, and are appropriately allocated among parameter estimation and privacy amplification. The error-correction leakage is evaluated according to Eq. (40). For each transmission distance, the intensity parameters and their probability distributions are optimized numerically to maximize the secret key rate. The channel transmittance is modeled as $\eta = 10^{-\alpha L/10}$, where α is the fiber loss coefficient and L is the transmission distance.

Table 1 Simulation parameters used in the numerical analysis. Here e_d denotes the intrinsic misalignment error rate, p_d denotes the dark count rate, η_d denotes the detection efficiency, and N denotes the number of transmitted pulses.

Set	e_d	p_d	η_d	N
i	1%	7.2×10^{-8}	65%	10^{12}

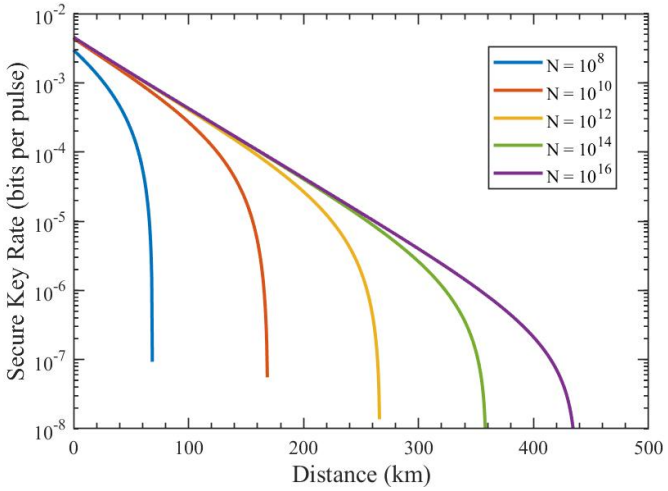


Fig. 2 Finite-key secret key rate as a function of transmission distance for different total numbers of transmitted pulses N. Smaller data sizes lead to stronger statistical fluctuations, resulting in lower key rates and shorter secure transmission distances.

Figure 2 shows the finite-key secret key rate as a function of the transmission distance for different values of the total number of pulses N. It can be observed that, when N is small, statistical fluctuations are significant, leading to a lower bound on the single-photon contribution and a higher upper bound on the phase error rate. As a result, the secret key rate is substantially reduced, and the maximum secure transmission distance is shortened.

As N increases, the statistical fluctuations gradually diminish, and the finite-key secret key rate approaches the large-data-size regime. This behavior is consistent with the theoretical analysis and indicates that a sufficiently large data size is required to suppress finite-size effects in practical systems.

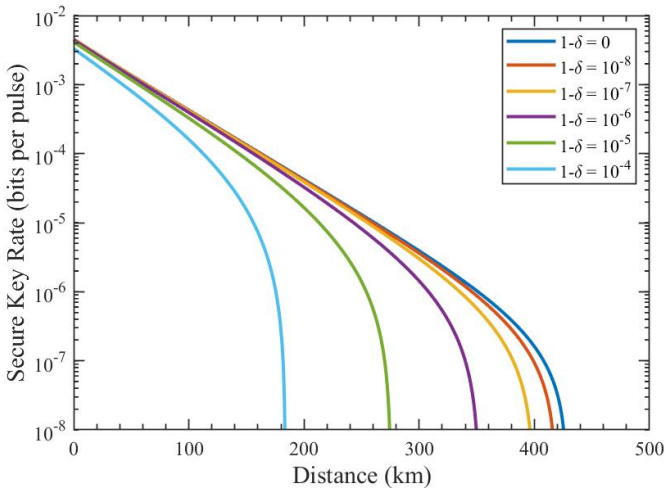


Fig. 3 Secret key rate as a function of transmission distance for different Trojan–horse attack strengths, characterized by the leakage parameter δ. As δ decreases (i.e., information leakage increases), the key rate significantly decreases, and the maximum secure transmission distance is markedly reduced, indicating that side-channel leakage has a substantial impact on system performance.

Figure 3 illustrates the secret key rate as a function of the transmission distance for different values of the leakage parameter δ, which characterizes the strength of the Trojan–horse attack. As δ decreases (i.e., the information leakage becomes stronger), the secret key rate decreases significantly, and the maximum secure transmission distance is reduced.

A smaller δ enlarges the Cauchy–Schwarz intervals between the practical and reference statistics. Consequently, the lower bound on the single-photon contribution n_{Z1}^{L} decreases, while the upper bound on the phase error rate $e_{Z1}^{\mathrm{ph,U}}$ increases. Both effects reduce the final secret key rate.

In contrast, when the leakage is weak (i.e., δ is close to unity), the system performance approaches that of the ideal case without leakage.

Figure 4 shows the finite-key secret key rate as a function of the transmission distance for different detector dark-count rates p_d. The results show that the influence of p_d is weak in the short-distance regime, where the signal-induced detection events dominate the total detection statistics. However, in the long-distance regime, the channel loss greatly reduces the signal detection probability, and dark counts become a major noise source. This increases the bit-error rate and the estimated phase error rate, thereby reducing the final secret key rate.

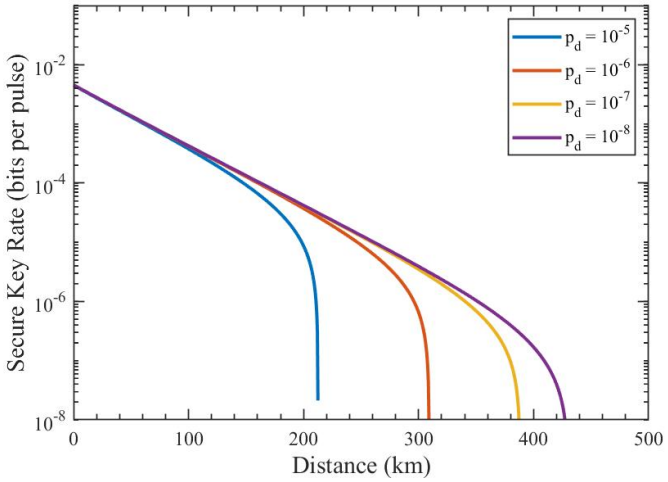


Fig. 4 Finite-key secret key rate as a function of the transmission distance for different detector dark-count rates p_d. The other simulation parameters are kept fixed.

It can also be seen that decreasing p_d significantly extends the maximum secure transmission distance. When p_d is relatively large, such as $p_d = 10^{-5}$, the key rate decreases rapidly and vanishes at a shorter distance. By contrast, for smaller dark-count rates, such as $p_d = 10^{-7}$ and $p_d = 10^{-8}$, positive key rates can be maintained over longer transmission distances. This demonstrates that detector dark counts are one of the dominant factors limiting the long-distance performance of finite-key MP-QKD.

5 Conclusion

In this work, we studied the finite-key security of the MP-QKD protocol under Trojan–horse attacks. Different from the ideal-source model used in conventional security analysis, we considered the practical case where Eve may inject probing light into the transmitters and obtain partial information from the back-reflected signals. To describe this source-side information leakage, we introduced a leakage parameter and modeled the emitted states at the level of valid mode-paired events.

Based on the reference-state method and the Cauchy–Schwarz inequality [32], we established a connection between the observable statistics of the practical protocol and the virtual statistics of the reference protocol. With this relation, the decoy-state parameter estimation can still be performed in the presence of Trojan–horse attacks. Furthermore, by considering finite-size statistical fluctuations, we derived a computable finite-key secret key rate formula for MP-QKD under side-channel leakage.

The numerical results show that the proposed framework can evaluate the secure transmission performance of MP-QKD in realistic conditions. Under weak leakage and sufficiently large data size, the protocol can still maintain a positive secret key rate over a distance of about 400 km. However, when the leakage becomes stronger, the total

number of transmitted pulses decreases, or the detector dark-count rate increases, the achievable key rate is reduced and the maximum secure transmission distance becomes shorter.

Overall, this work provides a practical finite-key security evaluation method for MP-QKD against Trojan–horse attacks. The results show that source-side leakage must be carefully considered in real MP-QKD systems, and that improving transmitter isolation, monitoring back-reflected light, increasing the data size, and reducing detector noise are important for achieving long-distance secure quantum key distribution.

Appendix A Composable Finite-Key Security Analysis

In this appendix, we present the derivation of the finite-key secret key length formula for the MP-QKD protocol within the composable security framework [33, 34]. The goal is to derive the final secure key length under finite-size effects based on smooth entropy uncertainty relations and privacy amplification [35].

Assume that Alice and Bob send a total of N optical pulses in the MP-QKD protocol. After Charlie's measurement announcement, mode pairing, basis sifting, and parameter estimation, Alice and Bob obtain raw keys of length n_Z in the Z basis, denoted by Z and Z'.

The protocol is required to satisfy composable security. If the protocol is ϵ_{cor}-correct and ϵ_{sec}-secret, the total security parameter is

$$\epsilon_{\mathrm{tot}} = \epsilon_{\mathrm{cor}} + \epsilon_{\mathrm{sec}}. \tag{A1}$$

According to the quantum leftover hash lemma, when a two-universal hash function is used for privacy amplification, the security of the final key of length ℓ is determined by the smooth min-entropy:

$$2\epsilon + \frac{1}{2}\sqrt{2^{\ell - H_{\min}^{\epsilon}(Z|E')}} \leq \epsilon_{\mathrm{sec}}. \tag{A2}$$

During error correction, λ_{EC} bits of information are leaked, together with an additional verification cost $\log_2(2/\epsilon_{\mathrm{cor}})$. Therefore,

$$H_{\min}^{\epsilon}(Z|E') \geq H_{\min}^{\epsilon}(Z|E) - \lambda_{\mathrm{EC}} - \log_2 \frac{2}{\epsilon_{\mathrm{cor}}}. \tag{A3}$$

The secure key in MP-QKD mainly originates from single-photon events in the Z basis. The raw key can be decomposed as

$$Z = Z_1 Z_{\mathrm{rest}}, \tag{A4}$$

where Z_1 corresponds to the single-photon contribution.

Using the chain rule for smooth entropy, one obtains

$$H_{\min}^{\epsilon}(Z|E) \geq H_{\min}^{\epsilon}(Z_1|Z_{\text{rest}}E) - 2\log_2 \frac{1}{\sqrt{2}\hat{\epsilon}}. \tag{A5}$$

According to the uncertainty relation for smooth entropy, the secrecy of the single-photon contribution is related to the phase error rate:

$$H_{\min}^{\epsilon}(Z_1|Z_{\text{rest}}E) \geq n_{Z1}^{L}\left[1 - h(e_{Z1}^{\text{ph,U}})\right], \tag{A6}$$

where n_{Z1}^{L} denotes the lower bound on the number of single-photon effective Z-pairs and $e_{Z1}^{\text{ph,U}}$ is the corresponding upper bound on the phase error rate [34].

Combining the above relations yields

$$H_{\min}^{\epsilon}(Z|E') \geq n_{Z1}^{L}\left[1 - h(e_{Z1}^{\text{ph},U})\right] - \lambda_{\text{EC}} - \log_2 \frac{2}{\epsilon_{\text{cor}}} - 2\log_2 \frac{1}{\sqrt{2}\hat{\epsilon}}. \tag{A7}$$

Finally, incorporating the privacy amplification parameter ϵ_{PA}, the secure key length satisfies

$$\ell \leq n_{Z1}^{L}\left[1 - h(e_{Z1}^{\text{ph,U}})\right] - \lambda_{\text{EC}} - \log_2 \frac{2}{\epsilon_{\text{cor}}} - 2\log_2 \frac{1}{\sqrt{2}\hat{\epsilon}\epsilon_{\text{PA}}}. \tag{A8}$$

The finite-key secret key rate is therefore

$$R_{\text{fin}} = \frac{\ell}{N}. \tag{A9}$$

A.1 Simulation of Observed Values

To evaluate the secret key rate without performing physical experiments, a theoretical model is employed to simulate the observed detection statistics for different intensity [9].

Let

$$k_a^i \in \{\mu_a, \nu_a, o_a\}, \qquad k_b^i \in \{\mu_b, \nu_b, o_b\} \tag{A10}$$

denote Alice's and Bob's intensity settings in the i-th round, respectively.

Define

$$k_a = k_a^i + k_a^j, \qquad k_b = k_b^i + k_b^j. \tag{A11}$$

The response probabilities corresponding to the left and right detectors are

$$q_{k_a^i k_b^i}^{L,\theta_i} = (1-p_d)e^{-\frac{k_a^i\eta_a + k_b^i\eta_b}{2}}\left[e^{\sqrt{k_a^i\eta_a k_b^i\eta_b}\cos\theta_i} - (1-p_d)e^{-\frac{k_a^i\eta_a + k_b^i\eta_b}{2}}\right], \tag{A12}$$

$$q_{k_a^i k_b^i}^{R,\theta_i} = (1-p_d)e^{-\frac{k_a^i\eta_a + k_b^i\eta_b}{2}}\left[e^{-\sqrt{k_a^i\eta_a k_b^i\eta_b}\cos\theta_i} - (1-p_d)e^{-\frac{k_a^i\eta_a + k_b^i\eta_b}{2}}\right]. \tag{A13}$$

Here,

$$\theta_i = \theta_a^i - \theta_b^i, \tag{A14}$$

and

$$\eta_a = \eta_d 10^{-\alpha L_a/10}, \qquad \eta_b = \eta_d 10^{-\alpha L_b/10}. \tag{A15}$$

The average response probability is given by

$$p = \sum_{k_a^i, k_b^i} p_{k_a^i} p_{k_b^i} \Pr(C_i = 1 | k_a^i k_b^i), \tag{A16}$$

where

$$\Pr(C_i = 1 | k_a^i k_b^i) = \int_0^{2\pi} \frac{d\theta_i}{2\pi} \left(q_{k_a^i k_b^i}^{L,\theta_i} + q_{k_a^i k_b^i}^{R,\theta_i} \right). \tag{A17}$$

Appendix B Detailed Derivation of the Finite-Key Bounds under Trojan–Horse Attacks

In this appendix, we provide the detailed derivation of the statistical bounds used in Sec. III. The purpose is to clarify how the reference-state method and the Cauchy–Schwarz constraint are combined with the finite-key decoy-state analysis of MP-QKD. Since the raw key of MP-QKD is generated after the mode-pairing step, the elementary object in the following derivation is a valid paired event $P^{i,j}$ rather than a single optical pulse [36].

For a paired event $P^{i,j}$, Alice's settings are denoted by

$$\mathbf{k}_A^{i,j} = \left(k_A^i, k_A^j, \theta_A^i, \theta_A^j \right), \tag{B18}$$

and Bob's settings are denoted by

$$\mathbf{k}_B^{i,j} = \left(k_B^i, k_B^j, \theta_B^i, \theta_B^j \right). \tag{B19}$$

Here $k_A^i, k_A^j \in \{\mu_A, \nu_A, 0\}$ and $k_B^i, k_B^j \in \{\mu_B, \nu_B, 0\}$ are the intensity settings, while $\theta_A^i, \theta_A^j, \theta_B^i, \theta_B^j$ are the phase settings. In the presence of a Trojan–horse attack, Eve's back-reflected light may depend on these settings. Thus, the practical state associated with $P^{i,j}$ can be written as

$$\left| \Psi^{i,j} \right\rangle = \sum_{\mathbf{k}_A^{i,j}, \mathbf{k}_B^{i,j}} \sqrt{p_{\mathbf{k}_A^{i,j}} p_{\mathbf{k}_B^{i,j}}} \left| R_{\mathbf{k}_A^{i,j}} \right\rangle_A \left| R_{\mathbf{k}_B^{i,j}} \right\rangle_B \left| \psi_{\mathbf{k}_A^{i,j}, \mathbf{k}_B^{i,j}}^{i,j} \right\rangle_{CPE}. \tag{B20}$$

The system C denotes the optical modes sent to Charlie, P denotes the purifying systems, and E denotes Eve's side-channel system. The corresponding reference state is chosen as

$$\left| \Phi^{i,j} \right\rangle = \sum_{\mathbf{k}_A^{i,j}, \mathbf{k}_B^{i,j}} \sqrt{p_{\mathbf{k}_A^{i,j}} p_{\mathbf{k}_B^{i,j}}} \left| R_{\mathbf{k}_A^{i,j}} \right\rangle_A \left| R_{\mathbf{k}_B^{i,j}} \right\rangle_B \left| \phi_{\mathbf{k}_A^{i,j}, \mathbf{k}_B^{i,j}}^{i,j} \right\rangle_{CP} |\tau\rangle_E, \tag{B21}$$

where $|\tau\rangle_E$ is independent of Alice's and Bob's intensity and phase choices. Therefore, the reference protocol does not leak side-channel information to Eve.

The indistinguishability between the practical and reference states is measured by

$$\delta_{i,j} = \left|\left\langle \Phi^{i,j} \middle| \Psi^{i,j} \right\rangle\right|. \tag{B22}$$

For a conservative analysis, we define the worst-case leakage parameter

$$\delta = \min_{(i,j)\in\mathcal{P}} \delta_{i,j}, \tag{B23}$$

where $\mathcal{P}$ is the set of valid paired events. Hence,

$$\left|\left\langle \Phi^{i,j} \middle| \Psi^{i,j} \right\rangle\right| \geq \delta. \tag{B24}$$

for all valid pairs. The parameter δ summarizes the amount of side-channel leakage: $\delta = 1$ corresponds to the ideal no-leakage case, whereas smaller δ indicates stronger information leakage.

Let $\hat{M}$ be a POVM element corresponding to an arbitrary post-selected event in the MP-QKD protocol. It may represent, for example, the event that a valid pair belongs to the Z set, the X set, the vacuum-related set V, or the event that an X-pair contains a bit error. Define

$$p = \left\langle \Psi^{i,j} \middle| \hat{M} \middle| \Psi^{i,j} \right\rangle, \qquad p^{\mathrm{ref}} = \left\langle \Phi^{i,j} \middle| \hat{M} \middle| \Phi^{i,j} \right\rangle. \tag{B25}$$

Since $0 \leq \hat{M} \leq I$, we have

$$\left\langle \Phi^{i,j} \middle| \Psi^{i,j} \right\rangle = \left\langle \Phi^{i,j} \middle| \hat{M} \middle| \Psi^{i,j} \right\rangle + \left\langle \Phi^{i,j} \middle| (I - \hat{M}) \middle| \Psi^{i,j} \right\rangle. \tag{B26}$$

Using the triangle inequality and the Cauchy–Schwarz inequality, one obtains

$$\begin{aligned}
\left|\left\langle \Phi^{i,j} \middle| \Psi^{i,j} \right\rangle\right| &\leq \left|\left\langle \Phi^{i,j} \middle| \hat{M} \middle| \Psi^{i,j} \right\rangle\right| + \left|\left\langle \Phi^{i,j} \middle| (I - \hat{M}) \middle| \Psi^{i,j} \right\rangle\right| \\
&\leq \sqrt{\langle \Phi^{i,j}| \hat{M} |\Phi^{i,j}\rangle \langle \Psi^{i,j}| \hat{M} |\Psi^{i,j}\rangle} + \sqrt{\langle \Phi^{i,j}| (I - \hat{M}) |\Phi^{i,j}\rangle \langle \Psi^{i,j}| (I - \hat{M}) |\Psi^{i,j}\rangle} \\
&= \sqrt{p^{\mathrm{ref}} p} + \sqrt{(1 - p^{\mathrm{ref}})(1 - p)}.
\end{aligned} \tag{B27}$$

Together with $\left|\left\langle \Phi^{i,j} \middle| \Psi^{i,j} \right\rangle\right| \geq \delta$, this gives

$$\sqrt{p p^{\mathrm{ref}}} + \sqrt{(1 - p)(1 - p^{\mathrm{ref}})} \geq \delta. \tag{B28}$$

Solving this inequality for p^{ref} yields

$$G_\delta^L(p) \leq p^{\mathrm{ref}} \leq G_\delta^U(p), \tag{B29}$$

where

$$G_\delta^L(p) = \begin{cases} g_\delta^-(p), & p > 1-\delta^2, \\ 0, & \text{otherwise}, \end{cases} \tag{B30}$$

and

$$G_\delta^U(p) = \begin{cases} g_\delta^+(p), & p < \delta^2, \\ 1, & \text{otherwise}. \end{cases} \tag{B31}$$

Here

$$g_\delta^\pm(p) = p + (1-\delta^2)(1-2p) \pm 2\delta\sqrt{(1-\delta^2)p(1-p)}. \tag{B32}$$

This is the Cauchy–Schwarz constraint used in Sec. III. It converts the actually observed statistics into conservative bounds on the corresponding reference statistics.

We now apply this constraint to the pair-level statistics of MP-QKD. Let N_{pair} be the number of valid paired events. For $\Omega \in \{Z, X, V\}$, let $n_{\Omega,\mathbf{s}}$ be the observed number of pairs in the set Ω with pair-intensity label $\mathbf{s} = (s_A, s_B)$, and let $m_{X,\mathbf{s}}$ be the number of bit-error events among the X-pairs with label $\mathbf{s}$. The corresponding practical probabilities are

$$Q_{\Omega,\mathbf{s}} = \frac{n_{\Omega,\mathbf{s}}}{N_{\text{pair}}}, \qquad T_{X,\mathbf{s}} = \frac{m_{X,\mathbf{s}}}{N_{\text{pair}}}. \tag{B33}$$

In the finite-key regime, the observed counts are affected by statistical fluctuations. Let $\mathcal{K}_{N,\epsilon}^L(\cdot)$ and $\mathcal{K}_{N,\epsilon}^U(\cdot)$ be lower and upper concentration bounds with failure probability ϵ. Then the reference probabilities are bounded as

$$Q_{\Omega,\mathbf{s}}^{\text{ref,L}} = G_\delta^L\left[\frac{\mathcal{K}_{N_{\text{pair}},\epsilon}^L(n_{\Omega,\mathbf{s}})}{N_{\text{pair}}}\right], \tag{B34}$$

and

$$Q_{\Omega,\mathbf{s}}^{\text{ref,U}} = G_\delta^U\left[\frac{\mathcal{K}_{N_{\text{pair}},\epsilon}^U(n_{\Omega,\mathbf{s}})}{N_{\text{pair}}}\right]. \tag{B35}$$

Therefore,

$$Q_{\Omega,\mathbf{s}}^{\text{ref,L}} \le Q_{\Omega,\mathbf{s}}^{\text{ref}} \le Q_{\Omega,\mathbf{s}}^{\text{ref,U}}. \tag{B36}$$

Similarly, the reference error-gain statistics satisfy

$$T_{X,\mathbf{s}}^{\text{ref,L}} = G_\delta^L\left[\frac{\mathcal{K}_{N_{\text{pair}},\epsilon}^L(m_{X,\mathbf{s}})}{N_{\text{pair}}}\right], \tag{B37}$$

and

$$T_{X,\mathbf{s}}^{\text{ref,U}} = G_\delta^U\left[\frac{\mathcal{K}_{N_{\text{pair}},\epsilon}^U(m_{X,\mathbf{s}})}{N_{\text{pair}}}\right], \tag{B38}$$

so that

$$T_{X,\mathbf{s}}^{\text{ref,L}} \le T_{X,\mathbf{s}}^{\text{ref}} \le T_{X,\mathbf{s}}^{\text{ref,U}}. \tag{B39}$$

The concentration bounds can be chosen according to the finite-key statistical model. For the standard finite-key MP-QKD analysis, they may be taken as Chernoff-type

bounds. For a more general setting involving dependent statistics, they may instead be chosen as Kato-type bounds [37]. The following derivation only requires valid lower and upper confidence bounds.

Since the reference protocol contains no side-channel leakage, the decoy-state method can be applied to the reference statistics [38–41]. For a paired event, define

$$s_A = k_A^i + k_A^j, \qquad s_B = k_B^i + k_B^j, \qquad \mathbf{s} = (s_A, s_B). \tag{B40}$$

Let $u = (u_A, u_B)$ denote the photon-number label associated with the two mode-paired systems. In the reference protocol, the expected number of effective Z-pairs with intensity label $\mathbf{s}$ can be written as

$$\langle n_{Z,\mathbf{s}}^{\mathrm{ref}} \rangle = \sum_u P_Z(\mathbf{s}|u) n_{Z,u}^{\mathrm{ref}}, \tag{B41}$$

where $P_Z(\mathbf{s}|u)$ is the conditional probability that the pair-intensity label is $\mathbf{s}$ given the photon-number label u, and $n_{Z,u}^{\mathrm{ref}}$ is the number of effective Z-pairs with photon-number label u in the reference protocol. The single-photon contribution in the raw key corresponds to $u = (1, 1)$, and we denote

$$n_{Z1}^{\mathrm{ref}} \equiv n_{Z,(1,1)}^{\mathrm{ref}}. \tag{B42}$$

Using the finite-key intervals $Q_{Z,\mathbf{s}}^{\mathrm{ref,L/U}}$ and $Q_{V,\mathbf{s}}^{\mathrm{ref,L/U}}$ as inputs, the lower bound on the single-photon contribution can be written compactly as

$$n_{Z1}^{\mathrm{ref,L}} = F_Z^L \left(\left\{ Q_{Z,\mathbf{s}}^{\mathrm{ref,L}}, Q_{Z,\mathbf{s}}^{\mathrm{ref,U}}, Q_{V,\mathbf{s}}^{\mathrm{ref,L}}, Q_{V,\mathbf{s}}^{\mathrm{ref,U}} \right\}_{\mathbf{s}} \right). \tag{B43}$$

Here F_Z^L denotes the finite-key decoy-state estimator used for MP-QKD. In the main text it is written as an abstract function in order to keep the notation compact. One possible implementation is the Gaussian-elimination formula used in finite-key MP-QKD. Equivalently, it can be formulated as the following linear program:

The explicit forms of $P_Z(\mathbf{s}|u)$ and $P_V(\mathbf{s}|u)$ are determined by the phase-randomized coherent-state photon-number distribution and the mode-pairing rule of MP-QKD.

$$\begin{aligned}
\text{minimize} \quad & n_{Z,(1,1)}^{\mathrm{ref}} \\
\text{subject to} \quad & N_{\mathrm{pair}} Q_{Z,\mathbf{s}}^{\mathrm{ref,L}} \le \sum_u P_Z(\mathbf{s}|u) n_{Z,u}^{\mathrm{ref}} \le N_{\mathrm{pair}} Q_{Z,\mathbf{s}}^{\mathrm{ref,U}}, \quad \forall \mathbf{s}, \\
& N_{\mathrm{pair}} Q_{V,\mathbf{s}}^{\mathrm{ref,L}} \le \sum_u P_V(\mathbf{s}|u) n_{V,u}^{\mathrm{ref}} \le N_{\mathrm{pair}} Q_{V,\mathbf{s}}^{\mathrm{ref,U}}, \quad \forall \mathbf{s}, \\
& n_{Z,u}^{\mathrm{ref}} \ge 0, \qquad n_{V,u}^{\mathrm{ref}} \ge 0.
\end{aligned} \tag{B44}$$

The optimal value gives $n_{Z1}^{\mathrm{ref,L}}$. Defining

$$q_{Z1}^{\mathrm{ref,L}} = \frac{n_{Z1}^{\mathrm{ref,L}}}{N_{\mathrm{pair}}}, \tag{B45}$$

the Cauchy–Schwarz constraint implies

$$q_{Z1} \geq G_{\delta}^{L}\left(q_{Z1}^{\mathrm{ref,L}}\right). \tag{B46}$$

Thus, the practical finite-key lower bound on the number of single-photon effective Z-pairs is

$$n_{Z1}^{L} = \mathcal{K}_{N_{\mathrm{pair}},\epsilon}^{L}\left[N_{\mathrm{pair}} G_{\delta}^{L}\left(\frac{n_{Z1}^{\mathrm{ref,L}}}{N_{\mathrm{pair}}}\right)\right]. \tag{B47}$$

The phase error rate of single-photon Z-pairs cannot be observed directly. In MP-QKD, it is estimated from the bit-error statistics of single-photon X-pairs by random sampling without replacement. First, the number of single-photon effective X-pairs in the reference protocol is bounded by

$$n_{X1}^{\mathrm{ref,L}} = F_{X}^{L}\left(\left\{Q_{X,\mathbf{s}}^{\mathrm{ref,L}}, Q_{X,\mathbf{s}}^{\mathrm{ref,U}}, Q_{V,\mathbf{s}}^{\mathrm{ref,L}}, Q_{V,\mathbf{s}}^{\mathrm{ref,U}}\right\}_{\mathbf{s}}\right). \tag{B48}$$

The corresponding single-photon bit-error rate of the reference X-pairs is upper bounded by

$$e_{X1}^{\mathrm{bit,ref,U}} = F_{e}^{U}\left(\left\{T_{X,\mathbf{s}}^{\mathrm{ref,L}}, T_{X,\mathbf{s}}^{\mathrm{ref,U}}, Q_{X,\mathbf{s}}^{\mathrm{ref,L}}, Q_{X,\mathbf{s}}^{\mathrm{ref,U}}, Q_{V,\mathbf{s}}^{\mathrm{ref,L}}, Q_{V,\mathbf{s}}^{\mathrm{ref,U}}\right\}_{\mathbf{s}}\right). \tag{B49}$$

The functions F_{X}^{L} and F_{e}^{U} play the same role as in the standard finite-key analysis of MP-QKD. Their only difference here is that their inputs are replaced by the Cauchy–Schwarz-constrained reference intervals $Q_{\Omega,\mathbf{s}}^{\mathrm{ref,L/U}}$ and $T_{X,\mathbf{s}}^{\mathrm{ref,L/U}}$. Equivalently, F_{e}^{U} can be implemented as a linear program that maximizes the number of single-photon bit errors in the X set under the decoy-state constraints on both the gain and error statistics.

According to the random-sampling relation in finite-key MP-QKD, the phase error rate of the reference single-photon Z-pairs satisfies

$$e_{Z1}^{\mathrm{ph,ref,U}} = e_{X1}^{\mathrm{bit,ref,U}} + \Gamma\left(\epsilon_{\mathrm{rs}}, e_{X1}^{\mathrm{bit,ref,U}}, n_{X1}^{\mathrm{ref,L}}, n_{Z1}^{\mathrm{ref,L}}\right), \tag{B50}$$

where ϵ_{rs} is the failure probability of the random-sampling step, and

$$\Gamma(a,b,c,d) = \sqrt{\frac{(c+d)(1-b)b}{cd}\ln\left[\frac{c+d}{2\pi cd(1-b)ba^{2}}\right]}. \tag{B51}$$

For physical consistency, one may take

$$e_{Z1}^{\mathrm{ph,ref,U}} = \min\left\{e_{Z1}^{\mathrm{ph,ref,U}}, \frac{1}{2}\right\}. \tag{B52}$$

To map this reference bound back to the practical protocol, define the reference phase-error probability

$$q_{\mathrm{ph}}^{\mathrm{ref,U}} = \frac{n_{Z1}^{\mathrm{ref,L}}}{N_{\mathrm{pair}}} e_{Z1}^{\mathrm{ph,ref,U}}. \tag{B53}$$

Then the practical phase-error probability satisfies

$$q_{\mathrm{ph}} \le G_\delta^U\left(q_{\mathrm{ph}}^{\mathrm{ref,U}}\right). \tag{B54}$$

Thus, the number of phase errors in the practical protocol is upper bounded by

$$m_{\mathrm{ph}}^U = \mathcal{K}_{N_{\mathrm{pair}},\epsilon}^U\left[N_{\mathrm{pair}} G_\delta^U\left(q_{\mathrm{ph}}^{\mathrm{ref,U}}\right)\right]. \tag{B55}$$

Finally, the practical phase error rate is bounded by

$$e_{Z1}^{\mathrm{ph,U}} = \min\left\{\frac{m_{\mathrm{ph}}^U}{n_{Z1}^L}, \frac{1}{2}\right\}. \tag{B56}$$

Substituting the lower bound n_{Z1}^L and the upper bound $e_{Z1}^{\mathrm{ph,U}}$ into the composable finite-key formula derived in Appendix A gives the final secret key length under Trojan–horse attacks:

$$\ell_{\mathrm{THA}} \le n_{Z1}^L\left[1 - h\left(e_{Z1}^{\mathrm{ph,U}}\right)\right] - \lambda_{\mathrm{EC}} - \log_2\frac{2}{\epsilon_{\mathrm{cor}}} - 2\log_2\frac{1}{\sqrt{2}\hat{\epsilon}\epsilon_{\mathrm{PA}}}. \tag{B57}$$

The corresponding finite-key secret key rate is

$$R_{\mathrm{fin}}^{\mathrm{THA}} = \frac{\ell_{\mathrm{THA}}}{N}, \tag{B58}$$

where N is the total number of transmitted optical pulses.

Author Contributions. Ma prepared the main manuscript, while Shi offered guidance on the writing. Xu and Chen Shen were responsible for collecting relevant information. Haiqiang Ma conducted the final review and served as the corresponding author.

Funding. This work is supported by the National Natural Science Foundation of China Grant No. 12574535, the state Key Laboratory of Information Photonics and Optical Communications (Beijing University of Posts and Telecommunications) GrantNo. IPOC2026ZT12.

Data availability. No datasets were generated or analysed during the current study.

Disclosures

The authors declare no conflicts of interest.

References

[1] Bennett, C.H., Brassard, G.: Quantum cryptography: Public key distribution and coin tossing. In: Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India, pp. 175–179 (1984)

[2] Ekert, A.K.: Quantum cryptography based on bell's theorem. Physical Review Letters **67**(6), 661–663 (1991)

[3] Scarani, V., Bechmann-Pasquinucci, H., Cerf, N.J., Dušek, M., Lütkenhaus, N., Peev, M.: The security of practical quantum key distribution. Reviews of Modern Physics **81**(3), 1301–1350 (2009)

[4] Lo, H.-K., Curty, M., Tamaki, K.: Secure quantum key distribution. Nature Photonics **8**(8), 595–604 (2014)

[5] Portmann, C., Renner, R.: Security in quantum cryptography. Reviews of Modern Physics **94**(2), 025008 (2022)

[6] Renner, R.: Security of quantum key distribution. International Journal of Quantum Information **6**(01), 1–127 (2008)

[7] Lucamarini, M., Yuan, Z., Dynes, J.F., Shields, A.J.: Overcoming the rate–distance limit of quantum key distribution without quantum repeaters. Nature **557**(7705), 400–403 (2018)

[8] Zeng, P., Zhou, H.-Y., Wu, W.-J., Ma, X.: Mode-pairing quantum key distribution. Nature Communications **13**, 3903 (2022)

[9] Wang, Z.-H., Wang, R., Yin, Z.-Q., Wang, S., Lu, F.-Y., Chen, W., He, D.-Y., Guo, G.-C., Han, Z.-F.: Tight finite-key analysis for mode-pairing quantum key distribution. Communications Physics **6**, 265 (2023)

[10] Lu, Z., Wang, G., Li, C., Cao, Z.-Q.: Asymmetric mode-pairing quantum key distribution. Physical Review A **109**(1), 012401 (2024)

[11] Zhang, L., Li, W., Pan, J., Lu, Y., Li, W., Li, Z., Huang, Y., Ma, X., Xu, F., Pan, J.-W.: Experimental mode-pairing quantum key distribution surpassing the repeaterless bound. Physical Review X **15**(2), 021037 (2025)

[12] Shor, P.W., Preskill, J.: Simple proof of security of the bb84 quantum key distribution protocol. Physical Review Letters **85**(2), 441–444 (2000)

[13] Takeoka, M., Guha, S., Wilde, M.M.: Fundamental rate-loss tradeoff for optical quantum key distribution. Nature Communications **5**, 5235 (2014)

[14] Shi, K., Wang, H., Zhang, Y., Ma, D., Ma, F., Ma, H.: Mode-pairing quantum key distribution with the single-photon-added coherent source. Quantum Information Processing **25**(4), 115 (2026) https://doi.org/10.1007/s11128-026-05135-y

[15] Lydersen, L., Wiechers, C., Wittmann, C., al.: Hacking commercial quantum cryptography systems by tailored bright illumination. Nature Photonics **4**(10), 686–689 (2010)

[16] Gisin, N., Fasel, S., Kraus, B., Zbinden, H., Ribordy, G.: Trojan-horse attacks on quantum-key-distribution systems. Physical Review A **73**(2), 022320 (2006)

[17] Vakhitov, A., Makarov, V., Hjelme, D.R.: Large pulse attack as a method of conventional optical eavesdropping in quantum cryptography. Journal of Modern Optics **48**(13), 2023–2038 (2001)

[18] Huang, A., Navarrete, Á., Sun, S.-H., Chaiwongkhot, P., Curty, M., Makarov, V.: Laser-seeding attack in quantum key distribution. Physical Review Applied **12**(6), 064043 (2019)

[19] Jain, N., Anisimova, E., Khan, I., al.: Trojan-horse attacks threaten the security of practical quantum cryptography. New Journal of Physics **16**(12), 123030 (2014)

[20] Lucamarini, M., Choi, I., Ward, M.B., Dynes, J.F., Yuan, Z.L., Shields, A.J.: Practical security bounds against the trojan-horse attack in quantum key distribution. Physical Review X **5**(3), 031030 (2015) https://doi.org/10.1103/PhysRevX.5.031030

[21] Navarrete, Á., Curty, M.: Improved finite-key security analysis of quantum key distribution against trojan-horse attacks. Quantum Science and Technology **7**(3), 035021 (2022) https://doi.org/10.1088/2058-9565/ac74dc

[22] Xu, F., Ma, X., Zhang, Q., Lo, H.-K., Pan, J.-W.: Secure quantum key distribution with realistic devices. Reviews of Modern Physics **92**(2), 025002 (2020)

[23] Makarov, V., Anisimov, A., Skaar, J.: Effects of detector efficiency mismatch on security of quantum cryptosystems. Physical Review A **74**(2), 022313 (2006)

[24] Ding, H.-J., Liu, J.-Y., Zhou, X., Zhang, C.-H., Li, J., Wang, Q.: Improved finite-key security analysis of measurement-device-independent quantum key distribution against a trojan-horse attack. Physical Review Applied **19**(4), 044022 (2023) https://doi.org/10.1103/PhysRevApplied.19.044022

[25] Zhou, Y., Ding, H.-J., Liu, J.-Y., Zhang, C.-H., Zhou, X., Wang, Q.: Finite-key security analysis against the trojan-horse attack on practical reference-frame-independent measurement-device-independent quantum key distribution. Physical Review Applied **23**(2), 024007 (2025) https://doi.org/10.1103/PhysRevApplied.23.024007

[26] Peng, Q., Chen, J., Xing, T., al.: Practical security of twin-field quantum key distribution with optical phase-locked loop under wavelength-switching attack. npj Quantum Information **11**, 7 (2025) https://doi.org/10.1038/s41534-025-00963-9

[27] Luo, T., Liu, Q., Sun, X., Zhang, Z., Huang, C., Xiao, X., Wei, K.: Security analysis against the trojan horse attack on practical polarization-encoding quantum key distribution systems. Physical Review A **109**(4), 042608 (2024) https://doi.org/10.1103/PhysRevA.109.042608

[28] Mizutani, A., Curty, M., Lim, C.C.W., Lo, H.-K., Tamaki, K.: Security of quantum key distribution with imperfect devices. New Journal of Physics **17**(9), 093011 (2015) https://doi.org/10.1088/1367-2630/17/9/093011

[29] Sun, S.-H., Xu, F., Jiang, M.-S., Ma, X.-C., Lo, H.-K., Liang, L.-M.: Effect of source tampering in the security of quantum cryptography. Physical Review A **92**(2), 022304 (2015) https://doi.org/10.1103/PhysRevA.92.022304

[30] Pereira, M., Kato, G., Mizutani, A., Curty, M., Tamaki, K.: Quantum key distribution with correlated sources. Science Advances **6**(37), 4487 (2020) https://doi.org/10.1126/sciadv.aaz4487

[31] Xu, F., Curty, M., Qi, B., Lo, H.-K.: Protocols for quantum key distribution with side channels. New Journal of Physics **17**(10), 103012 (2015)

[32] Tamaki, K., Curty, M., Kato, G., Lo, H.-K., Azuma, K.: Loss-tolerant quantum cryptography with imperfect sources. Physical Review A **90**(5), 052314 (2014) https://doi.org/10.1103/PhysRevA.90.052314

[33] Wang, W., Tamaki, K., Curty, M.: Finite-key security analysis for quantum key distribution with leaky sources. New Journal of Physics **20**(8), 083027 (2018) https://doi.org/10.1088/1367-2630/aad839

[34] Koashi, M.: Simple security proof of quantum key distribution based on complementarity. New Journal of Physics **11**(4), 045018 (2009)

[35] Tomamichel, M., Lim, C.C.W., Gisin, N., Renner, R.: Tight finite-key analysis for quantum cryptography. Nature Communications **3**, 634 (2012)

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

[37] Curty, M., Azuma, K., Lo, H.-K.: Finite-key analysis for twin-field quantum key distribution. npj Quantum Information **5**(1), 64 (2019)

[38] Lim, C.C.W., Curty, M., Walenta, N., Xu, F., Zbinden, H.: Concise security bounds for practical decoy-state quantum key distribution. Physical Review A **89**(2), 022307 (2014)

[39] Hwang, W.-Y.: Quantum key distribution with high loss: toward global secure communication. Physical Review Letters **91**(5), 057901 (2003)

[40] Lo, H.-K., Ma, X., Chen, K.: Decoy state quantum key distribution. Physical Review Letters **94**(23), 230504 (2005)

[41] Wang, X.-B.: Beating the photon-number-splitting attack in practical quantum cryptography. Physical Review Letters **94**(23), 230503 (2005)