



OPEN

Measurement-device-independent quantum key distribution with leaky sources

Weilong Wang^{1,2,3✉}, Kiyoshi Tamaki⁴ & Marcos Curty¹

Measurement-device-independent quantum key distribution (MDI-QKD) can remove all detection side-channels from quantum communication systems. The security proofs require, however, that certain assumptions on the sources are satisfied. This includes, for instance, the requirement that there is no information leakage from the transmitters of the senders, which unfortunately is very difficult to guarantee in practice. In this paper we relax this unrealistic assumption by presenting a general formalism to prove the security of MDI-QKD with leaky sources. With this formalism, we analyze the finite-key security of two prominent MDI-QKD schemes—a symmetric three-intensity decoy-state MDI-QKD protocol and a four-intensity decoy-state MDI-QKD protocol—and determine their robustness against information leakage from both the intensity modulator and the phase modulator of the transmitters. Our work shows that MDI-QKD is feasible within a reasonable time frame of signal transmission given that the sources are sufficiently isolated. Thus, it provides an essential reference for experimentalists to ensure the security of implementations of MDI-QKD in the presence of information leakage.

In theory, quantum key distribution (QKD)^{1–4} provides an information-theoretically secure way to distribute secret keys between two distant parties (commonly known as Alice and Bob). In practice, however, this is not the case. This is so because real devices do not typically conform to the requirements imposed by the security proofs. Indeed, various types of quantum hacking attacks have been proposed and experimentally demonstrated recently, which exploit device imperfections in practical QKD systems⁴. To tackle these implementation security loopholes, many efforts have been made, among which device-independent (DI) QKD^{5–7} and measurement-device-independent (MDI) QKD⁸ are two prominent approaches. The security of DI-QKD relies on the violation of a Bell inequality^{9,10} and no knowledge about the inner working of the quantum apparatuses is needed given that the apparatuses are ‘honest’¹¹, i.e., given that they follow the prescriptions of the protocol and not those of Eve. DI-QKD is, however, difficult to implement experimentally with current technology, especially for long distances^{12–14}. On the other hand, thanks to its feasibility, MDI-QKD has attracted great attention and has been widely experimentally demonstrated in recent years^{15–22}. In terms of security, MDI-QKD closes all side-channels in the detection unit, which significantly simplifies the path towards achieving implementation security in QKD, as now one only needs to secure the source. MDI-QKD requires, however, that certain assumptions on the sources are satisfied.

A common assumption is that Alice’s and Bob’s transmitters do not leak any unwanted information out of their security zones. Inspired by the results introduced in^{23–25}, which study the information leakage problem in standard decoy-state QKD systems, here we relax such an unrealistic requirement and perform a finite-key security analysis of MDI-QKD with leaky sources. In particular, we focus on information leakage from two main apparatuses within the transmitters, the intensity modulator (IM), which is used to generate decoy states, and the phase modulator (PM), which is used to encode the basis and bit information. For instance, such information leakage might be due to a Trojan-horse attack (THA)²⁶ performed by Eve. In this framework, we evaluate the security of two prominent MDI-QKD protocols: the symmetric three-intensity decoy-state MDI-QKD scheme²⁷, and the efficient four-intensity decoy-state MDI-QKD protocol introduced in²⁸, which has recently been implemented over a distance of 404 km²⁰. As expected, our results show that MDI-QKD is more sensitive to information leakage than standard decoy-state QKD. Still, we show that MDI-QKD is feasible within a reasonable time frame of signal transmission given that Alice’s and Bob’s sources are sufficiently isolated.

¹El Telecomunicación, Department of Signal Theory and Communications, University of Vigo, 36310 Vigo, Spain. ²State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, Henan, China. ³Henan Key Laboratory of Network Cryptography Technology, Zhengzhou 450001, Henan, China. ⁴Faculty of Engineering, University of Toyama, Gofuku 3190, Toyama 930-8555, Japan. ✉email: wwang@com.uvigo.es

Methods

The symmetric three-intensity decoy-state MDI-QKD protocol. We begin by describing the specific steps of the symmetric three-intensity decoy-state MDI-QKD protocol. Here, we consider a sifting strategy which protects the protocol against the sifting attack²⁹. This is so because the total number of pulses sent by Alice and Bob is fixed a priori and, moreover, the termination condition is basis independent³⁰. The assumptions that we make on the users' devices in the absence of information leakage can be found in the Supplementary Information 1. The steps of the protocol are as follows:

1. *State preparation:* The first two steps of the protocol are repeated N times, where N is a prefixed number. In each round, Alice and Bob select a basis $\chi \in \{Z, X\}$ with probabilities p_Z and $p_X = 1 - p_Z$, and select an intensity setting γ^{j_A} and γ^{j_B} with $j_A, j_B \in \{s, v, w\}$, with probability p_{j_A} and p_{j_B} , respectively. Afterwards, each of them encodes a random bit in a phase-randomized WCP of the chosen intensity in the chosen basis by using, for instance, the polarization encoding scheme employed in Ref.⁸ and sends it to the untrusted relay via the quantum channel. Note that our analysis is valid for any other encoding scheme.
2. *Measurement:* The untrusted relay is supposed to perform a Bell state measurement (BSM) on the states received from Alice and Bob and then record the measurement outcomes. For concreteness, below we shall assume that the untrusted relay uses the BSM introduced in Ref.⁸, which is based on linear optical elements and can distinguish two Bell states. In reality, however, the relay can behave as Eve decides.
3. *Announcement of the measurement outcomes and random data post-selection:* Once the N rounds of steps 1 and 2 have finished, the relay announces in which rounds he obtained successful measurements together with the corresponding measurement outcomes. For each successful measurement event, Alice selects a fictitious basis Z_{A_c} or X_{A_c} with probability $p_{Z_{A_c}}$ and $p_{X_{A_c}} = 1 - p_{Z_{A_c}}$, respectively, and then she announces her fictitious basis choices.
4. *Sifting:* If Alice's choice is the X_{A_c} basis, Bob announces his state preparation basis choice but Alice does not announce hers and then they discard the corresponding data. If Alice's choice is the Z_{A_c} basis, both Alice and Bob announce their state preparation basis choices as well as their intensity settings. We denote by $Z^{j_A j_B}$ ($X^{j_A j_B}$) the set of indexes that identify the successful measurement events when Alice and Bob select the intensity settings γ^{j_A} and γ^{j_B} , respectively, Alice chooses the fictitious basis Z_{A_c} , and both of them select the Z (X) basis. If the sifting conditions $|Z^{j_A j_B}| \geq N_Z^{j_A j_B}$ and $|X^{j_A j_B}| \geq N_X^{j_A j_B}$ are satisfied for all $j_A, j_B \in \{s, v, w\}$, where $N_Z^{j_A j_B}$ and $N_X^{j_A j_B}$ are prefixed threshold values, Alice and Bob proceed to execute the next steps of the protocol. If the sifting conditions are not satisfied, the protocol aborts.
5. *Parameter estimation:* Alice and Bob estimate a lower bound, which we denote by $N_{\text{click},00,ss|Z}^L$ ($N_{\text{click},11,ss|Z}^L$), on the number of successful measurement events in the sifted key data set Z^{ss} , in which both of them sent vacuum (single-photon) pulses. Also they use all the data in the sets $Z^{k_A k_B}$ and $X^{j_A j_B}$, except that in the set Z^{ss} , to estimate an upper bound on the single-photon phase error rate in the sifted key data set Z^{ss} , which we denote by e_{ph}^U .
6. *Information reconciliation and privacy amplification:* Alice and Bob perform an error correction step for a predetermined quantum bit error rate (QBER), which we denote by E_Z^{ss} . Then Alice computes a hash of the sifted key data in Z^{ss} by using a random universal₂ hash function³¹ and sends Bob the hash value together with the hash function. Bob uses the hash function to compute a hash of his corrected sifted key data and checks if the hash value coincides with that of Alice. If both hash values coincide, this error verification step guarantees that they share identical keys after error correction except for an exponentially small probability. Moreover, if this step succeeds, then they perform a privacy amplification step by applying a random universal₂ hash function to distill the final secret key.

Note that the sifting condition in Step 4 of the above protocol is only for data processing, and it is not related to the termination of the quantum communication steps, i.e., Steps 1 and 2, which is basis independent. Therefore, as indicated above, the protocol is secure against the sifting attack³⁰.

Parameter estimation method for the three-intensity protocol with leaky sources. In this section we briefly explain the general idea of our method to estimate the relevant parameters that are required to evaluate the secret key rate formula in the presence of information leakage. For concreteness, we consider the security analysis introduced in³², which provides a lower bound on the secret key length, ℓ , given by

$$\ell \geq N_{\text{click},00,ss|Z}^L + N_{\text{click},11,ss|Z}^L \left[1 - H\left(e_{\text{ph}}^U\right) \right] - \text{leak}_{\text{EC}} - \log_2 \frac{2}{\varepsilon_{\text{sec}}^2 - \varepsilon} - \log_2 \frac{2}{\varepsilon_{\text{cor}}}, \quad (1)$$

where $H(x) = -x \log_2(x) - (1-x) \log_2(1-x)$ is the binary Shannon entropy function. The parameter leak_{EC} is the amount of syndrome information declared by Alice in the error correction step of the protocol, given by $\text{leak}_{\text{EC}} = |Z^{ss}| f_{\text{EC}} H(E_Z^{ss})$ for simplicity, where the parameter f_{EC} is the efficiency of the error correction code. The quantities ε_{sec} and ε_{cor} are the secrecy and correctness parameters of the protocol, respectively, and $\varepsilon \leq 1 - \varepsilon_{Z,00} \varepsilon_{Z,11} \varepsilon_{\text{ph},11}$ with $\varepsilon_{Z,00}$, $\varepsilon_{Z,11}$ and $\varepsilon_{\text{ph},11}$ being defined as the success probabilities when estimating the quantities $N_{\text{click},00,ss|Z}^L$, $N_{\text{click},11,ss|Z}^L$ and e_{ph}^U , respectively. In other words, ε denotes the failure probability that at least one of the estimations of $N_{\text{click},00,ss|Z}^L$, $N_{\text{click},11,ss|Z}^L$ and e_{ph}^U is incorrect.

In the following we explain how to estimate the quantities $N_{\text{click},00,ss|Z}^L$, $N_{\text{click},11,ss|Z}^L$ and e_{ph}^U in the presence of information leakage. The detailed calculations can be found in the Supplementary Information 1. For

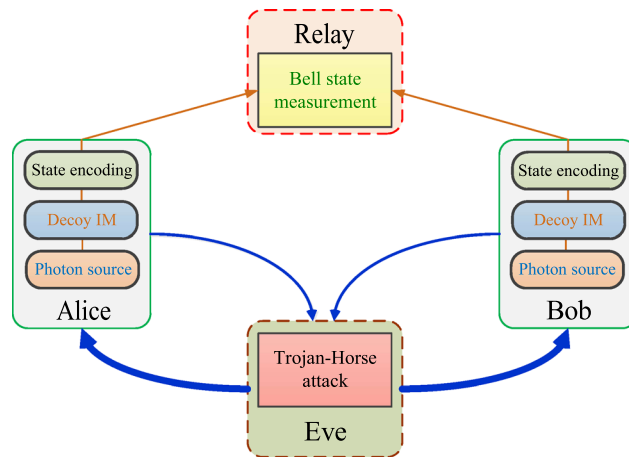


Figure 1. Each of Alice and Bob uses a photon source to prepare phase-randomised WCPs. Decoy states are generated by means of an intensity modulator (Decoy IM). The bit and basis information of the pulses are encoded with a state encoding setup (Encoding PM). The relay is supposed to perform a Bell state measurement on the incoming pulse pairs. In a THA, Eve actively sends bright light pulses (thick blue arrows) into Alice's and Bob's devices to trigger the emission of side-channel signals. Then, Eve measures the back-reflected light (thin blue arrows) to extract information about Alice's and Bob's internal settings. Note that since the relay is untrusted (i.e., it can be even Eve), in this figure we consider that it is the relay who performs the THA.

concreteness, we shall assume that the information leakage is due to a THA performed by an active Eve. In this THA against the MDI-QKD system, Eve separately sends bright light into Alice's and Bob's devices and then measures the back-reflected light. In so doing, she can obtain partial information about Alice's and Bob's internal settings for each experimental trial. See Fig. 1 for an illustration of Eve's THA. We remark, however, that our method is general and can be applied to analyze passive information leakage scenarios as well.

THA against the intensity modulator. Here, we briefly indicate the key ideas to analyze a THA targeted against the intensity modulator (IM), which is used to generate decoy states. The detailed calculations can be found in the Supplementary Information 1. In particular, we first consider an asymptotic scenario where Alice and Bob send an infinite number of pulses. In this scenario, we mainly apply the trace distance argument^{24,25,33} to relate the detection and error events arising from different intensity settings of Alice and Bob and obtain some linear relations between them. Then, by applying Azuma's inequality³⁴, the relations can be extended to the realistic regime where Alice and Bob send a finite number (N) of pulses. Finally, given the constraints provided by the mathematical relations obtained in the previous step, the relevant parameters which are needed to evaluate Eq. (1) can be estimated by using, for instance, linear programming techniques³⁵.

THA against the phase modulator. A THA against the phase modulator (PM) might render Alice's and Bob's output states (which now also contain Eve's systems due to the THA) *basis dependent*. As a result, Eve might be able to learn partial information about Alice's and Bob's basis and bit value choices each given time. The security of the standard BB84 protocol with a basis-dependent flaw has been analyzed in a previous work³⁶ by using the idea of a quantum coin^{37,38}. This idea was then generalized to phase encoding schemes for MDI-QKD where both Alice and Bob have basis-dependent flaws³⁹. Here, to estimate the phase error rate in the presence of a THA against the PM, we apply the method introduced in Ref³⁹ to our protocol.

More specifically, to simplify the analysis, we first consider a scenario where Alice's and Bob's light sources are both ideal single-photon sources. Also, we assume that Alice's and Bob's basis choices are random and do not depend on the IM or on the state of previous emitted pulses. Precisely, we consider a virtual entanglement scenario where each of Alice and Bob prepares a bipartite entangled state and then measures one of the two systems to actually prepare the states that are sent to the untrusted relay. We then apply the Bloch sphere bound⁴⁰ to this fictitious scenario and obtain the mathematical relation between the expected number of events, which contains the expected number of phase errors in the asymptotic limit. Next, we extend it to the finite-key regime by using Azuma's inequality, which contains the actual number of phase errors. Finally, the upper bound on the number of phase errors can be numerically estimated by simply using the optimization toolbox of Matlab, and thus we obtain the upper bound on the phase error rate. More details can be found in the Supplementary Information 1.

The four-intensity decoy-state MDI-QKD protocol. We now consider the four-intensity decoy-state MDI-QKD protocol introduced in²⁸, which has been recently implemented over a distance of 404 km²⁰. In this protocol, each of Alice and Bob uses one intensity setting γ^s for the Z basis states, and three intensity settings γ^v , γ^w and $\gamma^0 = 0$ for the X basis states. This is motivated by the fact that in order to increase the number of single-photon pulses emitted in the Z basis used for key generation, the intensity of the signal states, γ_s , needs to be close to one, while in order to have a tight estimation of the relevant parameters, the intensities in the X

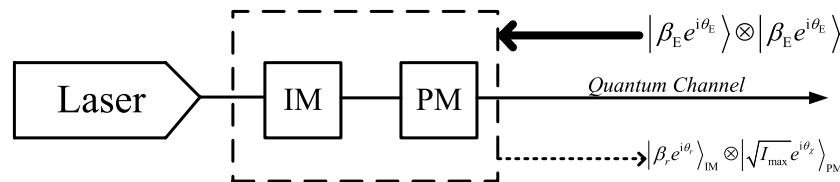


Figure 2. Example of a THA against the IM and the PM of Alice (Bob). For simplicity, we assume that Eve sends Alice (Bob) two high intensity single-mode coherent pulses, each of which is denoted by $|\beta_E e^{i\theta_E}\rangle$. One of them targets the IM and the other one targets the PM. We further assume also for simplicity that the back-reflected light from the IM and the PM to Eve is in a product state of two coherent states. One comes from the IM, which we denote by $|\beta_r e^{i\theta_r}\rangle$, and the other comes from the PM, which has the form $|\sqrt{I_{\max}} e^{i\theta_\chi}\rangle$, where r and χ refer to the intensity setting and basis choice, respectively, with $r \in \{s, v, w\}$ and $\chi \in \{Z, X\}$. Eve can learn partial information about the intensity settings and the basis choices by separately measuring the states $|\beta_r e^{i\theta_r}\rangle$ and $|\sqrt{I_{\max}} e^{i\theta_\chi}\rangle$.

basis used for parameter estimation need to be much weaker. With the four-intensity decoy-state MDI-QKD protocol, one can optimize the intensities for key generation and parameter estimation independently. The probabilities to select the corresponding intensities are p_s , p_v , p_w and p_0 , respectively, with $p_s + p_v + p_w + p_0 = 1$. Note that the probability to choose the Z basis is now $p_Z = p_s$ and the probability to choose the X basis is given by $p_X = p_v + p_w + p_0$.

Parameter estimation method for the four-intensity protocol with leaky sources. The security analysis of this protocol against information leakage from the IM and the PM is slightly different from that in the previous section. This is because of the following. Since now the intensity setting in the Z basis is unique and it is typically different from the intensity settings in the X basis, by analyzing the information leakage from the IM Eve can also learn partial information about the users' basis choices. Similarly, by analyzing the information leakage from the PM Eve can learn partial information about the users' intensity settings as well. That is, the information leakage from the IM and the PM of each user is now correlated. Fortunately, a general procedure to estimate the relevant parameters has already been briefly introduced in Ref²⁴. Here, we adapt it to the scenario of the four-intensity decoy-state MDI-QKD protocol.

Note that, in general, when the IM and the PM are correlated, the yields associated with different photon number states can also depend on the bit value²⁴. However, for simplicity, in the model above we assume that the back-reflected light does not carry information about the bit value but only about the basis. The specific calculations for the relevant parameters to evaluate Eq. (1) can be found in the Supplementary Information 1.

Results

The secret key rates in the presence of information leakage can be simulated given the security analysis summarized above. In this section, we show and compare the results for the three-intensity and four-intensity protocols.

Simulation results for the three-intensity decoy-state MDI-QKD protocol. In the simulation, only for illustration purposes, we assume a particular example of THA, which is shown in Fig. 2. Eve sends Alice (Bob) two high intensity single-mode coherent pulses, each of which is denoted by $|\beta_E e^{i\theta_E}\rangle$, with β_E representing the amplitude and θ_E the phase of the coherent state. One of them targets the IM and the other one targets the PM. For simplicity, we shall also assume that the back-reflected light from both the IM and the PM to Eve is still a coherent state. In so doing, as we show in the Supplementary Information 1, we can obtain simply analytical expressions for those quantities where we apply the trace distance argument. Moreover, we further assume that the back-reflected light from the IM has the form $|\beta_r e^{i\theta_r}\rangle$, where the values of the parameters β_r and θ_r depend on Alice's and Bob's intensity settings each given time with $r \in \{s, v, w\}$, and the back-reflected light from the PM is given by $|\sqrt{I_{\max}} e^{i\theta_\chi}\rangle$, where I_{\max} is the maximum intensity of the back-reflected light and $\chi \in \{Z, X\}$ refers to the basis choice. Note that, here, for simplicity, and in order to compare our simulation results to those in²⁵, we assume that Eve's back-reflected light from the PM only contains the basis information, as already mentioned above. That is, we assume that $|\Psi_{0,Z}^i\rangle_{A,E} = |\Psi_{0,Z}^i\rangle_{A'} \otimes |\phi_Z\rangle_E$ and $|\Psi_{1,Z}^i\rangle_{A,E} = |\Psi_{1,Z}^i\rangle_{A'} \otimes |\phi_Z\rangle_E$, where the state $|\phi_Z\rangle_E = |\sqrt{I_{\max}} e^{i\theta_Z}\rangle$ of Eve's back-reflected light is the same for both bit values (and similarly for the X basis). Here, the state $|\Psi_{0,Z}^i\rangle_{A,E}$ ($|\Psi_{1,Z}^i\rangle_{A,E}$) denotes the joint state of Alice and Eve when Alice uses the Z basis to encode the bit value 0 (1) in the i th round of the protocol and the state $|\Psi_{0,Z}^i\rangle_{A'}$ ($|\Psi_{1,Z}^i\rangle_{A'}$) denotes the state of Alice in such scenario. Likewise, we assume a similar situation at Bob's side. Further details can be found in the Supplementary Information 1. To learn partial information about the intensity settings, Eve can measure the state $|\beta_r e^{i\theta_r}\rangle$, and to learn partial information about the basis choices, Eve can measure the state $|\sqrt{I_{\max}} e^{i\theta_\chi}\rangle$. We emphasize, however, that this is just a particular model of a THA that we use it as an example to evaluate the secret key rate in a simple way. It is important to emphasize, however, that our security analysis can be applied to any THA. It remains a very important open question to determine the optimal state that Eve can send to Alice and Bob, as well as to experimentally characterize the identity of the back-reflected light. These questions are generally setup dependent and are beyond the scope of this paper.

In the presence of information leakage, the actual secret key length, ℓ' , is bounded by

e_d	p_d	η_{det}	α	f_{EC}
1%	5×10^{-6}	0.25	0.2	1.2

Table 1. Experimental parameters used in the simulations. The parameter e_d is the intrinsic error rate due to the misalignment of the MDI-QKD system; p_d is the dark count rate of the relay's detectors, which we assume is equal for all of them; η_{det} is the overall detection efficiency of the relay's receiver; α is the loss coefficient of the channel measured in dB/km; and f_{EC} is the efficiency of the error correction code.

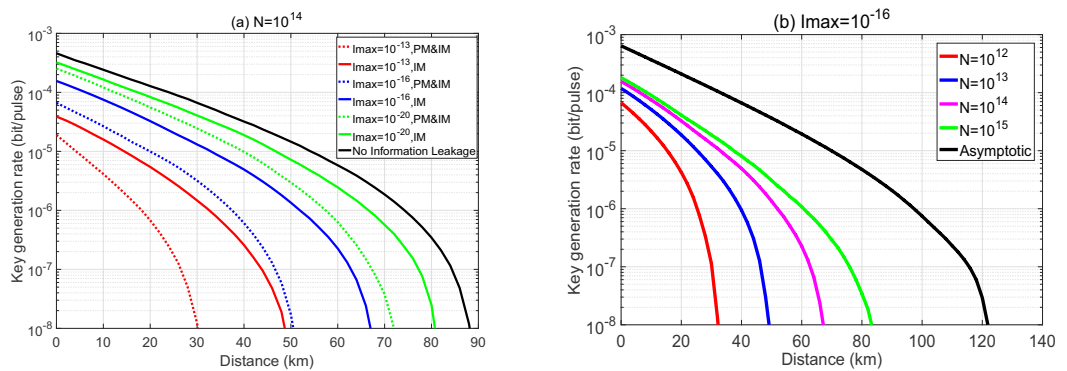


Figure 3. The secret key rate in logarithmic scale as a function of the distance in Case 1 for the three-intensity protocol. **(a)** Here we consider a fixed value of the total number of transmitted pulses, $N = 10^{14}$ and various values for the intensity I_{max} . **(b)** Here we fix $I_{\text{max}} = 10^{-16}$ and consider various values for N . Moreover, we evaluate information leakage from the IM only.

$$\ell' \geq \max_{\Gamma_{\text{AB}}} \min_{\Gamma_{\text{E}}} \ell, \quad (2)$$

where ℓ is given by Eq. (1). Here, Γ_{AB} and Γ_{E} denote the spaces of the parameters controlled by Alice and Bob, and by Eve, respectively. In the simulation, we assume a practically reasonable value for the weakest decoy state, $\gamma^w = 5 \times 10^{-4}$, and, without loss of generality, we assume that $\theta_s = 0$. The experimental parameters used in the simulations are listed in Table 1. Below we present the simulation results of the secret key rates in three practical cases within the framework of the THA described above. Each case corresponds to a particular model for the back-reflected light.

Case 1. In the framework of the THA considered, it is clear that the higher the intensity of the back-reflected light is, the more information Eve can extract. In this first example, we evaluate a worst-case scenario, where Alice and Bob may overestimate the intensity of the back-reflected light leaked to Eve. In particular, we suppose that the intensity β_r^2 is always upper bounded by a certain value I_{max} for all r and we conservatively assume that

$$\beta_s^2 = \beta_v^2 = \beta_w^2 = I_{\text{max}}. \quad (3)$$

The simulation result of the secret key rate, ℓ'/N , as a function of the transmission distance between Alice and Bob in this case is shown in Fig. 3a for a fixed value of the total number of transmitted pulses, $N = 10^{14}$. In this figure, the black solid line represents the key rate in the situation where there is no information leakage, namely $I_{\text{max}} = 0$, and the different colored lines correspond to different amounts of information leakage. More precisely, the colored solid lines represent the key rates in the presence of a THA against only the IM. If we compare these results to the longest achievable distance without information leakage, which is about 88 km, we find that now the secret key rate vanishes at about 48 km even when I_{max} is as small as 10^{-13} . The colored dotted lines represent the secret key rates in the presence of a THA against both the IM and the PM. Now the secret key rates are obviously lower than the ones corresponding to a THA against only the IM. For example, when $I_{\text{max}} = 10^{-13}$ the secret key rate now vanishes at only 30 km. These results highlight the strong effect that information leakage (even when is very tiny) can have on the performance of MDI-QKD.

As already observed in the finite-key analysis for decoy-state QKD²⁵, here we also find that in MDI-QKD Alice and Bob need to discard part of their data (on average about $Np_{\text{X}_{\text{Ac}}}$ pulse pairs) to estimate the phase error rate when there is information leakage from the PM. In our simulation, we find that the optimal value of $p_{\text{Z}_{\text{Ac}}}$ typically lies in the interval $[0.65, 0.9]$. Note that, compared to the simulation result in²⁵, we have that the value of $p_{\text{Z}_{\text{Ac}}}$ is typically smaller in the MDI-QKD protocol, which means that MDI-QKD has to sacrifice a bigger proportion of data than in the case of the standard decoy-state QKD protocol to estimate the phase error rate.

Also, we find that MDI-QKD seems to be more sensitive to information leakage. In order to obtain a certain performance, the value of I_{max} in MDI-QKD is roughly the square of that in standard decoy-state QKD due to the fact that in MDI-QKD there are two leaky sources (Alice and Bob) instead of only one leaky source. Thus, to

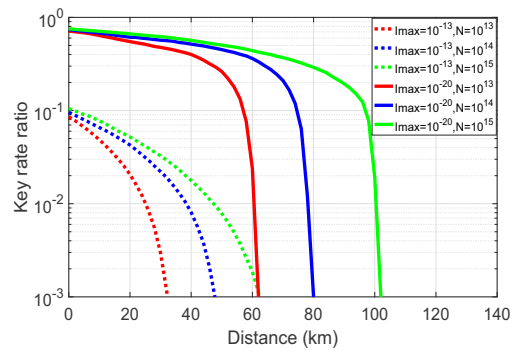


Figure 4. The ratio ($\ell'_{I_{\max}>0}/\ell'_{I_{\max}=0}$) between the secret key rates in logarithmic scale with and without information leakage as a function of the distance for two fixed positive values of $I_{\max} = \{10^{-13}, 10^{-20}\}$ and various values for N .

implement the MDI-QKD protocol, both Alice and Bob need to carefully isolate their devices from the external environment to guarantee the security of the system.

In Fig. 3b, the different colored lines show the secret key rate as a function of the distance for a fixed value $I_{\max} = 10^{-16}$ and for different total numbers of transmitted pulses. Here, for simplicity, we only plot the key rates against information leakage from the IM and omit the results when there is also information leakage from the PM as they are similar to those shown in Fig. 3b. That is, in this figure we can see the effect of the information leakage as a function of the number of transmitted pulses. For example, when $I_{\max} = 10^{-16}$, the longest achievable distance is about 84 km when the total number of transmitted pulses is $N = 10^{15}$. However, when $N = 10^{12}$, this distance decreases to 32 km. Our results indicate that the finite-key effect has a much bigger impact on the secret key rate in the presence of information leakage²⁷. The reason for this is mainly that, in order to estimate the statistical fluctuations for a finite sampling size in the presence of information leakage from the IM, our methodology relies on applying Azuma's inequality³⁴ to the total number of *transmitted pulses*. In contrast, when there is no information leakage from the IM, one can apply Azuma's inequality to the number of *pulses detected*. This is so because in this latter case, one can assume a counterfactual scenario where Alice and Bob select their intensity settings a posteriori, i.e., after the relay has detected the successful events. As a consequence, the performance of MDI-QKD in the finite-key regime is comparatively worse in the presence of information leakage from the IM. Note that for the case of information leakage from the PM, we actually apply Azuma's inequality to the number of the detected events.

To further illustrate how the information leakage affects the secret key rate as a function of the number of transmitted pulses, in Fig. 4 we plot the ratio ($\ell'_{I_{\max}>0}/\ell'_{I_{\max}=0}$) between the secret key rates for two fixed positive values of information leakage, $I_{\max} = \{10^{-13}, 10^{-20}\}$ and those when $I_{\max} = 0$ (i.e., when there is no information leakage) for different values of N . Here, for simplicity, we disregard again the information leakage from the PM. From Fig. 4 one can see that given a fixed distance and a fixed value of N , the ratio when $I_{\max} = 10^{-13}$ is at least one order of magnitude lower than that when $I_{\max} = 10^{-20}$. And the ratio when $I_{\max} = 10^{-13}$ drops faster as the distance increases than that when $I_{\max} = 10^{-20}$. For instance, if we focus on the red lines, from 0 to 30 km, the ratio drops from about 10^{-1} to 10^{-3} when $I_{\max} = 10^{-13}$ (i.e., two orders of magnitude) while the ratio drops only from 0.71 to 0.49 (i.e., of the same order of magnitude) when $I_{\max} = 10^{-20}$. This suggests that the effect of information leakage increases when N decreases, and the finite-size effect is amplified when the amount of information leakage increases. We remark that the simulation results for the other two cases that we consider next are analogous to those of Fig. 4 and thus we omit them in the next two subsections.

Case 2. In the previous case, we considered a conservative scenario for Alice and Bob, where the intensity of the back-reflected light is maximal and independent of the settings of the IM. Thus, the amount of information leaked might be overestimated, which results in a relatively pessimistic lower bound on the secret key rate. However, in practice, the input light of Eve may also go through the IM. As a consequence, the back-reflected light could be modulated in the same manner as the senders' pulses during the state preparation process. In this case, we have that

$$\beta_s^2 = \frac{\gamma^s}{\gamma^v} \beta_v^2 = \frac{\gamma^s}{\gamma^w} \beta_w^2 = I_{\max}. \quad (4)$$

That is, here we assume that the maximum amount of information leakage comes from the largest intensity setting of the senders, namely $I_{\max} = \beta_s^2$. The intensity of the back-reflected light corresponding to the other intensity settings fulfills the conditions: $\beta_s^2/\beta_v^2 = \gamma^s/\gamma^v$ and $\beta_s^2/\beta_w^2 = \gamma^s/\gamma^w$.

The simulation results of the secret key rate are shown in Fig. 1 in the Supplementary Information. The behavior of the curves is very similar to those in Case 1, and in the simulation we find that the optimized value of $p_{Z_{Ac}}$ is similar as well. One main difference is that with the same experimental parameters the secret key rate

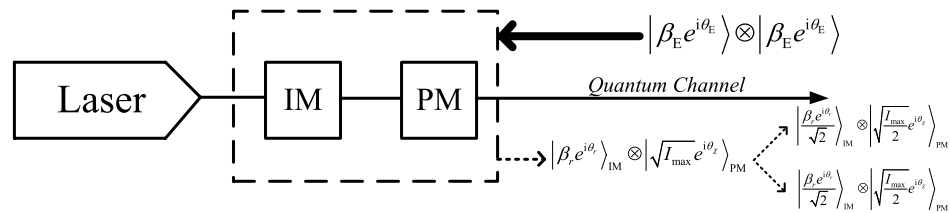


Figure 5. Example of a THA against correlated IM and PM of Alice (Bob). The simulation is similar to that in Fig. 2 but now $r \in \{s, v, w, 0\}$. Moreover, we assume that Eve splits the joint back-reflected light by means of a 50:50 beamsplitter, one part is to learn information about the intensity settings and the other part is used to learn information about the basis choices.

is now a little bit higher and can go a bit further than that in Case 1. For example, when the total number of transmitted pulses is 10^{14} and the maximum intensity of the back-reflected light is $I_{\max} = 10^{-13}$, we find that the secret key is positive up to about 54 km while in Case 1 this distance is 48 km in the presence of information leakage only from the IM.

Case 3. In this case we consider a more favorable situation for Alice and Bob where they implement an additional step to randomize the phase of each signal going out of their transmitters including the back-reflected light to Eve. Moreover, we optimistically assume that there is no information leakage from this phase randomization step. Furthermore, we suppose that the amplitudes β_k still satisfy Eq. (4) like in the previous case. Then, we have that the state of Eve's back-reflected light from the IM and the PM are given by:

$$\begin{aligned}\rho_{\gamma^k} &= e^{-(\beta_k)^2} \sum_{n=0}^{\infty} \frac{(\beta_k)^2}{n!} |n\rangle \langle n|, \\ \rho_{I_{\max}} &= e^{-I_{\max}} \sum_{n=0}^{\infty} \frac{I_{\max}}{n!} |n\rangle \langle n|,\end{aligned}\quad (5)$$

respectively.

This means that the information about Alice's and Bob's inner settings can only be leaked via the amplitudes of the back-reflected light but Eve cannot obtain any information from its phase. We remark, however, that here we consider a model which is slightly different from the ones considered in previous works^{24,25}. To be precise, in Refs.^{24,25} the authors consider that the phase randomization step is only applied to the back-reflected light from the IM. However, here we consider that this step is applied to the back-reflected light from both the IM and the PM. This means that, now Eve cannot exploit any information leakage from the PM, but only information leakage from the IM as the state $\rho_{I_{\max}}$ does not depend on the basis choice.

The simulation results of the secret key rate are shown in Fig. 2 in the Supplementary Information. Here, we find that the typical interval where $p_{Z_{Ac}}$ lies is $[0.71, 0.93]$. Compared to the secret key rate shown in the previous two cases, now the secret key rate is obviously improved. For example, when the total number of transmitted pulses is $N = 10^{14}$ and $I_{\max} = 10^{-7}$, the secret key rate remains positive up to about 62 km. In comparison, the maximum achievable distance with the same number of transmitted pulses and assuming an I_{\max} as low as 10^{-13} is only about 36 km in Case 2, and it is even worse in Case 1.

In practice, however, Eve might also perform a THA against the phase randomization step to obtain some information about the random phase applied by Alice and Bob each given time. This will obviously reduce the benefit of the phase randomization step. One could also analyze this last scenario with the techniques in this paper, but for simplicity we omit it here.

Simulation results for the four-intensity decoy-state MDI-QKD protocol. In what follows, for illustration purposes we consider a particular example of the THA considered in the previous section, which is shown in Fig. 5. Now, however, the back-reflected light from the IM has the form $|\beta_r e^{i\theta_r}\rangle$ with $r \in \{s, v, w, 0\}$. Moreover, since the IM and the PM are correlated, Eve can jointly measure the states $|\beta_r e^{i\theta_r}\rangle$ and $|\sqrt{I_{\max}} e^{i\theta_\chi}\rangle$, which is the back-reflected light from the PM with $\chi \in \{Z, X\}$, to extract partial information about both the intensity settings and the basis choices. Particularly, we shall consider that Eve splits the joint back-reflected light $|\beta_r e^{i\theta_r}\rangle \otimes |\sqrt{I_{\max}} e^{i\theta_\chi}\rangle$ into two parts by means of a 50:50 beamsplitter, one part is used to learn partial information about the intensity settings and the other part is used to learn partial information about the basis choices. We remark, however, that our method to estimate the phase error rate could be applied to any strategy applied by Eve. Importantly, to have a fair comparison with the simulation results shown in the previous section, we assume that the amount of information leaked to Eve in both protocols is the same. That is, we assume that the intensity of the back-reflected light is equal in both cases.

Note that since the information leakage from the IM and the PM is correlated, in the following figures, we plot the secret key rates in the presence of information leakage from both devices.

Case 1. The simulation result of the secret key rate, ℓ'/N , as a function of the transmission distance between Alice and Bob in this case is shown in Fig. 6a for a fixed value of the total number of transmitted pulses, $N = 10^{14}$.

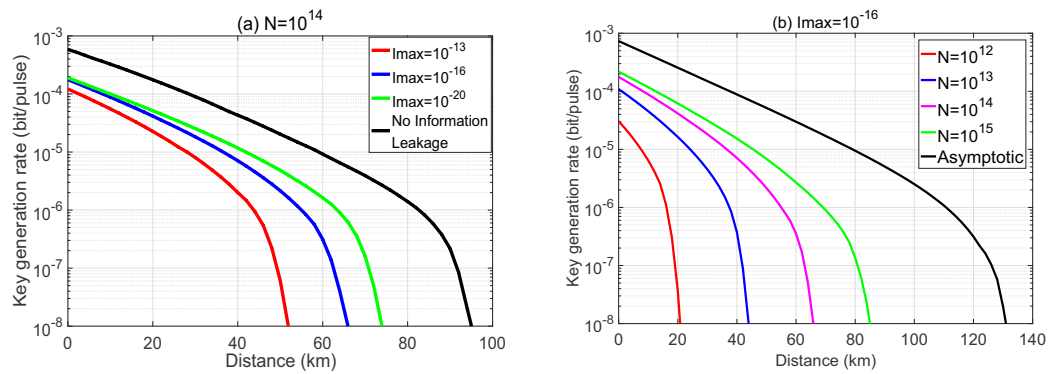


Figure 6. The secret key rate in logarithmic scale as a function of the distance in Case 1 for the four-intensity protocol. **(a)** Here we consider a fixed value of the total number of transmitted pulses, $N = 10^{14}$ and various values for the intensity I_{\max} . **(b)** Here we fix $I_{\max} = 10^{-16}$ and consider various values for N .

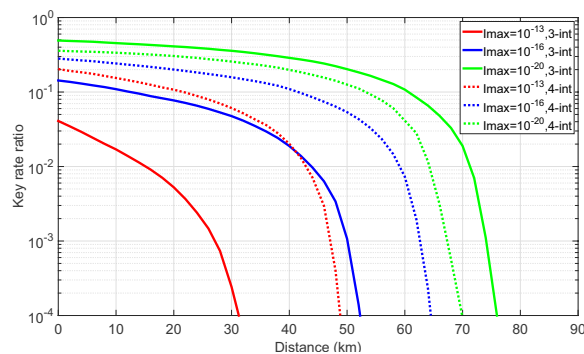


Figure 7. The ratio $(\ell'_{I_{\max} > 0} / \ell'_{I_{\max} = 0})$ between the secret key rates in logarithmic scale with and without information leakage as a function of the distance for a fixed number, $N = 10^{14}$, of transmitted pulses in the two protocols (3-int and 4-int represent the three-intensity decoy state MDI-QKD protocol and the four-intensity decoy-state MDI protocol that we consider, respectively).

The black solid line represents the key rate in the situation where there is no information leakage, and the different colored lines correspond to different amounts of information leakage. The longest achievable distance without information leakage is about 96 km. When $I_{\max} = 10^{-13}$, the secret key rate vanishes at about 52 km. In the simulation, we find that in this case the optimized value of $p_{Z_{Ac}}$ typically lies in the interval $[0.75, 0.94]$. That is, in this protocol Alice and Bob can sacrifice a smaller proportion of the data than that in the symmetric three-intensity decoy-state MDI-QKD protocol (where, as we have shown in the previous section, the typical interval of the optimized value of $p_{Z_{Ac}}$ is $[0.65, 0.9]$).

Figure 6b shows the secret key rates as a function of the distance for a fixed value $I_{\max} = 10^{-16}$ for different total numbers of transmitted pulses. For example, the longest achievable distance is about 84 km when the total number of transmitted pulses is $N = 10^{15}$. However, when $N = 10^{12}$, this distance decreases to 21 km.

To further compare the effect of the information leakage on the secret key rate in the two MDI-QKD protocols that we consider, we plot the ratio $(\ell'_{I_{\max} > 0} / \ell'_{I_{\max} = 0})$ between the secret key rates for different positive values of information leakage, I_{\max} , and the secret key rate when there is no information leakage, i.e., $I_{\max} = 0$, given a fixed total number of transmitted pulses, say, $N = 10^{14}$ in Fig. 7. The solid and dotted lines represent the ratios in the symmetric three-intensity decoy-state MDI-QKD protocol and in the four-intensity decoy-state MDI-QKD protocol, respectively. In the following, for simplicity, let us denote these two protocols by ‘3-int’ and ‘4-int’, respectively. The result in Fig. 7 indicates that when the amount of information leakage is small enough, for instance, $I_{\max} = 10^{-20}$, the impact of the information leakage on the 3-int protocol is smaller than that on the 4-int protocol as the green solid line is always above the green dotted line. However, the key rate ratio drops much faster as the amount of information leakage increases in the 3-int protocol than that in the 4-int protocol. From Fig. 7, we find that when $I_{\max} = 10^{-16}$ and $I_{\max} = 10^{-13}$, the ratio in the 4-int protocol is bigger than that in the 3-int protocol. That is, when I_{\max} increases, the effect of information leakage becomes more relevant on the 3-int protocol than that on the 4-int protocol given a fixed total number of transmitted pulses.

The intuition for this behaviour could be the following: from Figs. 2 and 5, we can see that the back-reflected light from the PM is the same for the 3-int and 4-int protocol. Now suppose that in the 4-int protocol Eve measures the back-reflected light from the IM and the PM independently instead of splitting the back-reflected light with a 50:50 BS. Then she learns the same information from the PM in both protocols. However, it may be more

difficult for Eve to learn the information from the IM in the 4-int protocol than in the 3-int protocol because she needs to distinguish between four states in the former but she only needs to distinguish between three states in the latter. In this case, the 4-int protocol is more robust against information leakage than the 3-int protocol for all values of I_{\max} . Nevertheless, if Eve exploits the correlations between the back-reflected light from the IM and the PM, then which protocol is more robust seems to depend on the value of I_{\max} . In addition, note that the results illustrated in Fig. 7 consider the case where Eve splits the back-reflected light with a 50:50 BS, which might not be the optimal option for the example of THAs evaluated.

Case 2. The simulation results of the secret key rate as a function of the transmission distance are shown in Fig. 3 in the Supplementary Information. The behavior of the curves is very similar to those in case 1, and in the simulation we find that the optimized value of $p_{Z_{Ac}}$ is also similar. One main difference is that with the same experimental parameters the secret key rate is a little higher and the achievable distance is a little longer than those in Case 1. For example, when the total number of transmitted pulses is $N = 10^{14}$ and the maximum intensity of the back-reflected light is $I_{\max} = 10^{-13}$, now we find that the secret key is positive up to about 57 km while in Case 1 this distance is 52 km.

Here we omit the comparison of the key rate ratios between the two protocols as the result in this case is similar to that shown in Fig. 7. And for the same reason, we omit such comparison in Case 3 as well.

Case 3. The simulation results of the secret key rate as a function of the transmission distance are shown in Fig. 4 in the Supplementary Information. Here, we find that the typical interval that $p_{Z_{Ac}}$ lies in is $[0.86, 0.99]$. Compared to the secret key rates shown in the previous two cases, now it is obviously improved. For example, when the total number of transmitted pulses is $N = 10^{14}$ and $I_{\max} = 10^{-7}$, the secret key rate remains positive up to about 66 km. In comparison, the maximal achievable distance with the same number of transmitted pulses and assuming an I_{\max} as low as 10^{-13} is only about 57 km (52 km) in Case 2 (Case 1). As discussed previously, this is because the phase randomization step removes the information leaked in the phase of the output states to Eve.

Conclusion and discussion

In this paper, we have quantitatively analyzed the security of two decoy-state MDI-QKD protocols with leaky sources in the finite-key regime. Specially, we have simulated the secret key rate under three particular examples of THA, where Eve sends coherent pulses of light to probe the intensity modulators and phase modulators of the legitimate parties. Similar to the analysis presented in²⁵, we have introduced an additional post-processing step in the actual protocol where Alice and Bob sacrifice part of their data. This step is necessary for the security proof to go through. Our simulation results suggest that MDI-QKD is more sensitive to information leakage than standard decoy-state QKD, but is possible to distill secure keys from leaky sources within a reasonable time frame of signal transmission given that Alice's and Bob's sources are sufficiently isolated. Furthermore, we have found that when the amount of information leakage is small enough, the effect of information leakage has a bigger impact on the four-intensity decoy-state MDI-QKD protocol than on the symmetric three-intensity decoy-state MDI-QKD protocol. However, when the amount of information leakage increases, the four-intensity MDI-QKD protocol becomes more robust against information leakage than the symmetric three-intensity MDI-QKD protocol.

We note that Ref.⁴¹ introduced a security analysis for MDI-QKD which does not have to characterize the states emitted by only assuming that the generated signals live in a qubit space. While this analysis might certainly have some advantages in some scenarios (e.g., when evaluating state preparation flaws), it cannot be applied to the situation we study here with leaky sources. Indeed, due to the presence of side-channels, the emitted signals are not qubits but higher dimensional signals. This means that, in its current formulation the work in Ref.⁴¹ does not apply to the scenario that we evaluate and cannot take information leakage into consideration.

We emphasize that the methods introduced in this paper are completely general and can be applied to any information leakage, not necessarily in the form of coherent states. We have assumed this particular model only for simplicity in order to perform simulations.

In this context it would be interesting to consider a stronger THA, where Eve sends entangled probe states to Alice's and Bob's sources instead of sending them independent bright pulses. In such a scenario, by performing a joint measurement on the outgoing states as well as on her ancilla system, Eve might be able to extract more information about Alice's and Bob's internal settings than what has been presented in this paper. This case, however, is beyond the scope of this work but could be evaluated with the techniques that have been introduced in this paper.

Received: 4 July 2020; Accepted: 28 December 2020

Published online: 18 January 2021

References

1. Bennett, C. H. & Brassard, G. *Quantum Cryptography: Public Key Distribution and Coin Tossing*, 175–179 (1984).
2. Ekert, A. K. Quantum cryptography based on bell theorem. *Phys. Rev. Lett.* **67**, 661 (1991).
3. Scarani, V. *et al.* The security of practical quantum key distribution. *Rev. Mod. Phys.* **81**, 1301 (2009).
4. Lo, H.-K., Curty, M. & Tamaki, K. Secure quantum key distribution. *Nat. Photon.* **8**, 595–604 (2014).
5. Barrett, J., Hardy, L. & Kent, A. No signaling and quantum key distribution. *Phys. Rev. Lett.* **95**, 010503 (2005).
6. Acín, A. *et al.* Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.* **98**, 230501 (2007).
7. Vazirani, U. & Vidick, T. Fully device-independent quantum key distribution. *Phys. Rev. Lett.* **113**, 140501 (2014).
8. Lo, H.-K., Curty, M. & Qi, B. Measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **108**, 130503 (2012).
9. Bell, J. S. On the einstein podolsky rosen paradox. *Phys. Phys. Fizika* **1**, 195 (1964).

10. Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V. & Wehner, S. Bell nonlocality. *Rev. Mod. Phys.* **86**, 419 (2014).
11. Curty, M. & Lo, H.-K. Foiling covert channels and malicious classical post-processing units in quantum key distribution. *NPJ Quantum Inf.* **5**, 14 (2019).
12. Gisin, N., Pironio, S. & Sangouard, N. Proposal for implementing device-independent quantum key distribution based on a heralded qubit amplifier. *Phys. Rev. Lett.* **105**, 070501 (2010).
13. Curty, M. & Moroder, T. Heralded-qubit amplifiers for practical device-independent quantum key distribution. *Phys. Rev. A* **84**, 010304 (2011).
14. Zapatero, V. & Curty, M. Long-distance device-independent quantum key distribution. *Sci. Rep.* **9**, 17749 (2019).
15. Rubenok, A., Slater, J. A., Chan, P., Lucio-Martinez, I. & Tittel, W. Real-world two-photon interference and proof-of-principle quantum key distribution immune to detector attacks. *Phys. Rev. Lett.* **111**, 130501 (2013).
16. da Silva, T. F. *et al.* Proof-of-principle demonstration of measurement-device-independent quantum key distribution using polarization qubits. *Phys. Rev. A* **88**, 052303 (2013).
17. Liu, Y. *et al.* Experimental measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **111**, 130502 (2013).
18. Tang, Z. *et al.* Experimental demonstration of polarization encoding measurement-device-independent quantum key distribution. *Phys. Rev. Lett.* **112**, 190503 (2014).
19. Tang, Y.-L. *et al.* Measurement-device-independent quantum key distribution over 200 km. *Phys. Rev. Lett.* **113**, 190501 (2014).
20. Yin, H.-L. *et al.* Measurement-device-independent quantum key distribution over a 404 km optical fiber. *Phys. Rev. Lett.* **117**, 190501 (2016).
21. Tang, Y.-L. *et al.* Measurement-device-independent quantum key distribution over untrusted metropolitan network. *Phys. Rev. X* **6**, 011024 (2016).
22. Comandar, L. *et al.* Quantum key distribution without detector vulnerabilities using optically seeded lasers. *Nat. Photon.* **10**, 312 (2016).
23. Lucamarini, M. *et al.* Practical security bounds against the trojan-horse attack in quantum key distribution. *Phys. Rev. X* **5**, 031030 (2015).
24. Tamaki, K., Curty, M. & Lucamarini, M. Decoy-state quantum key distribution with a leaky source. *New J. Phys.* **18**, 065008 (2016).
25. Wang, W., Tamaki, K. & Curty, M. Finite-key security analysis for quantum key distribution with leaky sources. *New J. Phys.* **20**, 083027 (2018).
26. Gisin, N., Fasel, S., Kraus, B., Zbinden, H. & Ribordy, G. Trojan-horse attacks on quantum-key-distribution systems. *Phys. Rev. A* **73**, 022320 (2006).
27. Curty, M. *et al.* Finite-key analysis for measurement-device-independent quantum key distribution. *Nat. Commun.* **5**, 3732 (2014).
28. Zhou, Y.-H., Yu, Z.-W. & Wang, X.-B. Making the decoy-state measurement-device-independent quantum key distribution practically useful. *Phys. Rev. A* **93**, 042324 (2016).
29. Pfister, C., Lütkenhaus, N., Wehner, S. & Coles, P. J. Sifting attacks in finite-size quantum key distribution. *New J. Phys.* **18**, 053001 (2016).
30. Tamaki, K. *et al.* Security of quantum key distribution with iterative sifting. *Quantum Sci. Technol.* **3**, 014002 (2018).
31. Carter, J. L. & Wegman, M. N. Universal classes of hash functions. *J. Comput. Syst. Sci.* **18**, 143–154 (1979).
32. Mizutani, A., Curty, M., Lim, C. C. W., Imoto, N. & Tamaki, K. Finite-key security analysis of quantum key distribution with imperfect light sources. *New J. Phys.* **17**, 093011 (2015).
33. Nielsen, M. A. & Chuang, I. L. *Quantum Information and Quantum Computation* (Cambridge University Press, Cambridge, 2000).
34. Azuma, K. Weighted sums of certain dependent random variables. *Tohoku Math. J.* **19**, 357–367 (1967).
35. Vanderbei, R. J. *et al.* *Linear Programming* (Springer, Berlin, 2015).
36. Lo, H.-K. & Preskill, J. Security of quantum key distribution using weak coherent states with nonrandom phases. *Quantum Inf. Comput.* **7**, 431–458 (2007).
37. Gottesman, D., Lo, H.-K., Lütkenhaus, N. & Preskill, J. Security of quantum key distribution with imperfect devices. *Quantum Inf. Comput.* **5**, 20 (2004).
38. Koashi, M. Simple security proof of quantum key distribution based on complementarity. *New J. Phys.* **11**, 045018 (2009).
39. Tamaki, K., Lo, H.-K., Fung, C.-H.F. & Qi, B. Phase encoding schemes for measurement-device-independent quantum key distribution with basis-dependent flaw. *Phys. Rev. A* **85**, 042307 (2012).
40. Tamaki, K., Koashi, M. & Imoto, N. Unconditionally secure key distribution based on two nonorthogonal states. *Phys. Rev. Lett.* **90**, 167904 (2003).
41. Yin, Z. *et al.* Measurement-device-independent quantum key distribution with uncharacterized qubit sources. *Phys. Rev. A* **88**, 062322 (2013).

Acknowledgements

This work was supported by the Galician Regional Government, consolidation of Research Units: AtlantTIC), the Spanish Ministry of Economy and Competitiveness (MINECO), the Fondo Europeo de Desarrollo Regional (FEDER) through Grant TEC2017-88243-R and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant agreement No. 675662 (Project QCALL). W. W. gratefully acknowledges support from the National Natural Science Foundation of China (Grants nos. 61701539, 61972413, 61901525) and the National Cryptography Development Fund (mmjj20180107, mmjj20180212). K.T. acknowledges support from JSPS KAKENHI Grant numbers JP18H05237 18H05237 and JST-CREST JPMJCR 1671.

Author contributions

All authors contributed to conceive the original idea. W.W. did the simulations and wrote the first draft. All authors reviewed and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-81003-2>.

Correspondence and requests for materials should be addressed to W.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021