Decoy-state quantum-key-distribution-based quantum private query with error tolerance bound

Bin Liu,[1,2,3,*] Zhonghao Liang,[1,†] Wei Huang,[1,4,‡] Fei Gao,[2,§] Jiayi Yuan,[1,‖] and Bingjie Xu[4,¶]

[1]*College of Computer Science, Chongqing University, Chongqing 400044, China*

[2]*The State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*

[3]*Center for Network and Information, Shihezi University, Shihezi, Xinjiang 832003, China*

[4]*Science and Technology on Communication Security Laboratory, Institute of Southwestern Communication, Chengdu 610041, China*

(Received 19 February 2024; accepted 3 May 2024; published 30 May 2024)

Quantum private query (QPQ) faces many challenges in practical applications. At present, some scholars have made substantial work on overcoming channel loss, channel noise, and nonideal light sources, respectively. However, a protocol that would overcome all three of these problems has yet to materialize. We review a practical QPQ protocol that can actually work in noisy channels. This protocol has an upper bound on tolerable errors based on the required level of security and reliability. Then, we study the security of the above QPQ protocol under weak coherent pulses. The results indicate that the multiphoton pulses have induced significant vulnerabilities, which seriously threatens the privacy of users. Finally, we propose a decoy-state method to solve the serious threat to user security caused by multiphoton pulses. The analysis shows that the decoy-state method significantly improves the security of the QPQ protocol under weak coherent pulses. The improved protocol can not only tolerate transmission losses and channel noise, but also overcome the security vulnerability caused by a nonideal light source.

DOI: 10.1103/PhysRevA.109.052442

I. INTRODUCTION

With the development of informatization, advances in quantum computing theory and experiments have challenged classical cryptographic protocols based on assumptions of computational complexity. How to ensure the security of the communication process and the privacy of the communication parties has become one of the key research topics today. Over the past few decades, quantum communication technologies have emerged, including quantum key distribution (QKD) [1–12], quantum secret sharing [13], quantum secure direct communication [14–16], quantum digital signatures [17,18], and more. These technologies hold promising application potential in various fields, notably in e-commerce [19,20]. Notably, QKD has undergone substantial experimental development, demonstrating information-theoretic security and finding practical applications in protecting information. Compared with public key systems such as RSA, the great success of quantum cryptography in secure communication offers a solution to the serious threats facing the increasing power of quantum computing. The research shows that quantum cryptography can achieve higher security in theory [21,22] than traditional cryptosystems.

During a data query, the user (Alice) typically wants to know a database element held by the database provider (Bob), but does not want him to know which element she is interested in. At the same time, Bob wants to limit the amount of information (usually one item) Alice can gain about the database to prevent her from getting additional database information. The task described above is called symmetrically private information retrieval (SPIR). In an ideal SPIR protocol, Bob cannot get any information about the address of the item Alice queried and Alice cannot get any items other than the item she queried. In essence, SPIR implements as one out of N oblivious transfer [23]. According to the no-go theorem [24], the ideal SPIR cannot be realized in quantum cryptography either. At present, the most practical approach is to loosen the security requirement in SPIR to the level of "deception sensitivity" (that is, all effective deception will have a nonzero probability of detection). In the field of quantum cryptography, quantum private query (QPQ) is the practical way to realize the SPIR task. Due to the properties of quantum mechanics, QPQ achieves higher security than traditional SPIR protocols. In recent decades, a large number of QPQ protocols have been proposed. At present, the implementation of QPQ protocol mainly includes QPQ based on quantum computation [25–27] and QPQ based on QKD [28–39]. In 2008, the first QPQ protocol (GLM protocol) [25] was given by Italian scholar Giovannetti *et al.* This kind of QPQ protocol encoding database information to the unitary operation is significant in theory. However, it is not practical due to the large dimensionality of the unitary operation. In contrast, QKD-based QPQ is easier to implement under the current technology. The research [28] shows that the QKD-based QPQ protocol not only resists the transmission losses, but also has the same

*liubin31416@gmail.com
†Zhonghao_Liang@cqu.edu.cn
‡Corresponding author: huangwei096505@aliyun.com
§gaof@bupt.edu.cn
‖1498203126@qq.com
¶xbjpku@163.com

difficulty as QKD in physical implementation. This kind of QPQ protocol is an important part of quantum cryptography and has gradually become a quantum scheme for implementing SPIR tasks.

The QKD-based QPQ protocol has received extensive attention in the past few years due to the realistic prospects. QPQ protocols based on various QKD schemes have been proposed [30–34]. However, there are still several problems to be solved before its practical application. Real-world quantum cryptosystems always have some drawbacks of real apparatuses (transmission loss, channel noise, unideal sources, etc.). In 2011, Jacobi *et al.* [28] used the SARG04 QKD scheme [40] to design a QPQ protocol (J-protocol), which is completely immune to transmission losses and can be implemented with existing QKD technology. Then, in order to solve the problem caused by channel noise, some scholars have proposed error correction schemes with practical channels for several QKD-based QPQ protocols [32,41]. After analyzing the error correction work of the former, Wei *et al.* [39] put forward a scheme that comprehensively considers reliability, user privacy, and database security.

The above several works have made great contributions to the practical application of QKD-based QPQ. However, there still are differences between quantum communication devices under existing technical conditions and theoretical models. These differences are mainly manifested in the light source and the detector side of two aspects. On the one hand, many theoretical protocols assume that the light source is a perfect single-photon source, while the actual light source will inevitably appear as multiphoton or no-photon signals. On the other hand, the detectors in the theoretical protocol are perfect. In practice, the detector has some problems such as uneven efficiency and dark counting. In fact, there are many security vulnerabilities caused by equipment defects in the practical application of quantum cryptography protocol. Security vulnerabilities on the detector side can be resolved by the measurement-device-independent communication mode [35,36], while there is a scarcity of practical security analysis on the light source side. Aiming at the problem of nonideal light source, we studied the practical security of users and databases at the theoretical level in the early stage [42–44]. However, the communication protocol in this study cannot deal with the problem caused by channel noise and cannot communicate in the real channel. Therefore, we analyze the possible security defects of QKD-based QPQ protocol with weak coherent pulse and the improved scheme of decoy-state method is also given. The improved protocol can really work via noisy channel and overcome the vulnerability of light source caused by multiphoton pulses. The improved scheme is a step forward for the practical application of QKD-based QPQ.

The remainder of this paper is organized as follows. We review the QKD-based QPQ protocol proposed by Wei *et al.* [39] in Sec. II. In Sec. III, we analyze the practical security of the above protocol under weak coherent pulses. Section IV gives the decoy-state method of QKD-based QPQ protocol. Finally, we propose a rearrangement of decoy-state QKD-based QPQ protocol and make a brief conclusion in Sec. V.

II. REVIEW OF THE ERROR TOLERANCE BOUND IN QKD-BASED QUANTUM PRIVATE QUERY

During quantum communication, significant transmission losses often occur when information is exchanged between the communicating parties. These losses arise when the receiver fails to detect the signal sent by the sender, necessitating multiple retransmissions from the sender's end. A concern arises if these losses are caused by unauthorized tapping of the communication line, as it allows the eavesdropper to learn the sender's message undetected. J-protocol [28] addressed this issue by implementing a postprocessing mechanism to mitigate transmission losses. In 2021, we conducted a comprehensive review of the J-protocol and identified security vulnerabilities associated with the use of weak coherent pulses [42]. Through our enhancements, the revised J-protocol effectively addresses these vulnerabilities related to multiphoton pulses.

However, several challenges remain before QKD-based QPQ protocols can be deployed in real-world scenarios. Notably, the improved J-protocol still grapples with significant channel noise issues in practical applications. The presence of channel noise gives rise to various quantum signal errors, including bit errors, flip errors, and phase errors. Consequently, the receiver may receive incorrect information and the sender may exploit these errors to conceal any cheating behavior.

Here, we analyze a protocol that can combat channel noise and improve it under the condition of weak coherent pulses. The error tolerance bound (ETB) protocol proposed by Wei *et al.* [39] is a suitable choice. Here's a recap of the protocol.

There are two default parameters: database size N and the forecasted upper limit ε for the error rate of the raw key.

(1) The user (Alice) sends the database (Bob) a long sequence of photons, each of which is a random state in $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$. Here,

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \tag{1}$$

$$|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \tag{2}$$

State $|0\rangle$ denotes horizontal polarization and $|1\rangle$ stands for vertical polarization.

(2) Bob randomly measures the received qubits in either Z basis (i.e., $\{|0\rangle, |1\rangle\}$) or X basis (i.e., $\{|+\rangle, |-\rangle\}$) and codes the Z (X) basis as 0 (1). Bob announces which qubits have been successfully detected and then ignores the undetected qubits.

(3) Bob randomly selects some photons and informs Alice to publish states in which they are sent to detect the error rate. If the error rate is within the threshold value ε, the protocol continues; otherwise, the protocol terminates.

(4) They discard the photons used for detection. If the rest of the photons' measurement output is $\{|0\rangle, |+\rangle\}$ ($\{|1\rangle, |-\rangle\}$), Bob announces bit 0 (1).

(5) According to the results published by Bob and the photon states prepared by herself, Alice can successfully infer the basis Bob used in step (2) with a probability of 1/4. For example, if Alice sends $|1\rangle$ and Bob announces 0, she knows that Bob's output is $|+\rangle$ and he measures it in the X basis. Thus Alice can obtain a conclusive bit 1. Then they record all the conclusive and inconclusive bits and share an oblivious

raw key K_r. Obviously, (1) Bob knows every bit in K_r, (2) Alice knows only 1/4 of its bits, and (3) Bob does not know which bits are conclusive bits to Alice.

(6) They transmit enough qubits successfully so that k_1k_2N bits are eventually retained as K_r. K_r is divided into k_2 groups, each containing k_1 N-bit substrings.

(7) Alice declares a random shift $i \in \{1, 2, \ldots, N\}$ to align the qubits she knows in substrings so that she can get the ith bit in them.

(8) Bob bitwise adds the substrings of each group to obtain a middle key in which Alice knows the ith bit (maybe more bits). Here, they obtain k_2 middle keys.

(9) An N-bit string is selected randomly by Bob as the final key K_f and encrypted by the k_2 middle keys, respectively. Then, Bob sends the k_2 ciphertexts to Alice.

(10) Alice uses the known ith bit in each middle key to decrypt these ciphertexts and obtains k_2 bits. Then they enter into one of two modes as follows.

(a) If the k_2 bits are identical (each of them would equal K_f^i, the ith bit of K_r), they execute the retrieval mode. This case indicates that Alice obtains correct K_f^i and a high probability of no error occurs. Alice announces a shift $s(s = i - j)$, if she wants to query the jth item from database. Bob encrypts the database with K_f shifted by s and sends the encrypted database to Alice. Alice then can use K_f^i to decrypt the jth item.

(b) If the k_2 bits are not identical, which means at least one error occurs, they execute the checking mode. Bob is required to announce the whole raw key. If the error rate is higher than threshold ε, Bob is identified as a cheater and the protocol aborts. Otherwise, the protocol restarts (executed no more than $N_{\max}$ times).

ETB protocol achieves three crucial objectives. First, it constrains the number of erroneous items the database receives by a repetition code with overwhelming probability. And in step (3), Bob is allowed to check the states of the carrier qubits to detect outside eavesdroppers. These measures allow Alice to determine that Bob is cheating when the error rate exceeds the threshold ε, which greatly ensures user privacy. Second, even when subjected to a joint-measurement attack and exploiting the redundancy of error correction, Alice can only access a limited number of database items. This substantially guarantees database security. Third, the two sides of communication enter two different modes, which can realize real-time detection. In this way, it is allowed that Alice detects Bob's virtual attack before the protocol ends, which makes more sense than detecting it after the protocol ends. At the same time, Bob's virtual attack can be detected with a significant probability (exceeding 70% when Bob induces an error rate larger than 3%).

Consequently, the probability of Bob being able to cheat in ETB protocol is limited, contributing to the fulfillment of practical application requirements. By balancing protocol reliability, database security, and user privacy, ETB protocol also establishes an exact upper bound ε on tolerable errors. Furthermore, two aspects of the work hold substantial significance. One involves embedding error filtering technology into the real-time check, substantially reducing the error rate when retrieving database items. The other introduces a method to quantify Bob's advantage. According to information theory, the entropy, that is,

$$-\sum_{i=1}^{N} p_i \log_2 p_i, \tag{3}$$

can be used to quantify information. For dishonest Bob, his advantage can be quantified as

$$I_{\text{Bob}} = \log_2 N - \left(-\sum_{i=1}^{N} p_i \log_2 p_i\right). \tag{4}$$

Here, p_i denotes the probability that the ith bit is selected as the retrieval address. Each p_i equals $1/N$ for honest Bob, but for the dishonest one, the value of p_i will be different, which means that he will gain an additional advantage.

III. SECURITY ANALYSIS OF ERROR TOLERANCE BOUND IN QUANTUM-KEY-DISTRIBUTION-BASED QUANTUM PRIVATE QUERY WITH WEAK COHERENT PULSE

ETB protocol deals with channel noise in QKD-based QPQ to improve the practicability of QPQ. The protocol works really well with channel noise and inherits many advantages of QKD-based QPQ protocol. However, there is still a crucial practical problem that has not been solved, namely the theoretical difference between the practical device and its ideal mode. Like most existing QKD-based QPQ protocols, the raw key distribution of ETB protocol also utilizes BB84 protocol particles, which is designed for ideal quantum communication devices. However, it will lead to practical security vulnerabilities when implementing theoretical protocols with existing technology. Due to the uncertainty of the light source, there is no perfect single-photon light source under experimental conditions (for example, in weak coherent pulses, multiphoton pulses are inevitable). The actual multiphoton pulse is different from the theoretical single-photon pulse, which will lead to the photon number splitting attack in the QKD protocol [45]. In fact, multiphoton pulses also pose a serious threat to the security of QKD-based QPQ protocols, especially on the light source side. Since the security vulnerabilities from the light source side do not introduce vulnerabilities on the database in ETB protocol, we only analyze the possible security vulnerabilities on the light source side.

In order to extract Alice's retrieval address, dishonest Bob typically needs to identify which raw key bits are Alice's conclusive bits. Wei *et al.* analyzed the optimal individual attack for dishonest Bob and designed a specific measurement basis to control the probability of Alice obtaining conclusive raw key bits (see Sec. IV, subsection D in [39]). This method may introduce errors into the raw key string, which can be detected by real-time checking mechanisms. Actually, in the theoretical case, Bob's attack is bound to introduce errors. The underlying reason lies in the fact that the mixed-state density operators formed by the ensembles $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$ are both 1/2 that are completely indistinguishable. Furthermore, the mixed states formed by the ensembles $\{|0\rangle, |+\rangle\}$ and $\{|1\rangle, |-\rangle\}$ cannot be unambiguously discriminated.

However, the situation is different for multiphoton pulses. In most quantum communication processes, the two parties only concern the probability of obtaining different results.

In the multiphoton case, Bob can control the probability of Alice obtaining a certain result by positive operator-valued measurement (POVM) without errors introduced. Since the channel loss is large in practice, Bob can use the low-loss channel to reduce the original signal loss, then retain only the multiphoton pulses sent by Alice and claim that no other pulse has been successfully received. In this case, Alice's practical security is extremely vulnerable. Taking the example of two-photon pulses, we next introduce three ways in which Bob can control the probability of Alice getting a conclusive bit.

A. Method for controlling Alice only getting inconclusive bits

For two-photon pulses, Alice actually sends $|00\rangle$, $|11\rangle$, $|++\rangle$, and $|--\rangle$, instead of $|0\rangle$, $|1\rangle$, $|+\rangle$, and $|-\rangle$. We can then define the following states:

$$|\Phi_0\rangle = \frac{1}{\sqrt{6}}(2|00\rangle + |01\rangle + |10\rangle), \tag{5}$$

$$|\Phi_1\rangle = \frac{1}{\sqrt{6}}(-|01\rangle - |10\rangle + 2|11\rangle). \tag{6}$$

Obviously,

$$\langle\Phi_0|11\rangle = \langle\Phi_0|--\rangle = 0, \tag{7}$$

$$\langle\Phi_1|00\rangle = \langle\Phi_1|++\rangle = 0. \tag{8}$$

Based on the above states, a set of POVM elements can be constructed:

$$\mathbf{\Pi_0} = \mathbf{p}|\Phi_0\rangle\langle\Phi_0|, \tag{9}$$

$$\mathbf{\Pi_1} = \mathbf{q}|\Phi_1\rangle\langle\Phi_1|, \tag{10}$$

$$\mathbf{\Pi}_?^1 = \mathbf{I} - \mathbf{\Pi_0} - \mathbf{\Pi_1}. \tag{11}$$

In order to satisfy the semipositive definite of $\mathbf{\Pi}_?^1$ ($\mathbf{\Pi}_?^1 \geqslant 0$), $\mathbf{p}$, $\mathbf{q}$ are equal to $3/4$. For the ensembles $\{|00\rangle, |++\rangle\}$ and $\{|11\rangle, |--\rangle\}$, we can calculate

$$\mathbf{tr}\left(\mathbf{\Pi_0}\frac{1}{2}(|00\rangle\langle 00| + |++\rangle\langle++|)\right) = \frac{1}{2}, \tag{12}$$

$$\mathbf{tr}\left(\mathbf{\Pi_1}\frac{1}{2}(|11\rangle\langle 11| + |--\rangle\langle--|)\right) = \frac{1}{2}. \tag{13}$$

The probability that Bob can distinguish whether the qubits sent by Alice belong to $\{|00\rangle, |++\rangle\}$ or $\{|11\rangle, |--\rangle\}$ is $1/2$. When Bob measures with this POVM and only keeps the qubits that have been identified by which ensemble they come from, he can then publish the corresponding 0 or 1, and at this position make it impossible for Alice to obtain the conclusive bit. For example, Alice sends $|00\rangle$ or $|++\rangle$ and Bob recognizes that the qubit belongs to $\{|00\rangle, |++\rangle\}$. Bob then announces bit 0, which means Alice cannot obtain a conclusive bit. The same result will be obtained while Alice sends $|11\rangle$ or $|--\rangle$. Through this POVM, Bob can control Alice to obtain inconclusive bits without introducing errors.

B. Method for controlling Alice getting conclusive bits with the probability of 1/2

Similar to subsection A, we define the following states:

$$|\Phi_z\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \tag{14}$$

$$|\Phi_x\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle). \tag{15}$$

Then, we can get

$$\langle\Phi_z|++\rangle = \langle\Phi_z|--\rangle = 0, \tag{16}$$

$$\langle\Phi_x|00\rangle = \langle\Phi_x|11\rangle = 0. \tag{17}$$

Based on the above states, another set of POVM elements can be constructed:

$$\mathbf{\Pi_z} = \mathbf{m}|\Phi_z\rangle\langle\Phi_z|, \tag{18}$$

$$\mathbf{\Pi_x} = \mathbf{n}|\Phi_x\rangle\langle\Phi_x|, \tag{19}$$

$$\mathbf{\Pi}_?^2 = \mathbf{I} - \mathbf{\Pi_z} - \mathbf{\Pi_x}. \tag{20}$$

In order to satisfy the semipositive definite of $\mathbf{\Pi}_?^2$ ($\mathbf{\Pi}_?^2 \geqslant 0$), $\mathbf{m}$, $\mathbf{n}$ are equal to 1. For the ensembles $\{|00\rangle, |11\rangle\}$ and $\{|++\rangle, |--\rangle\}$, we can calculate

$$\mathrm{tr}\left(\mathbf{\Pi_z}\frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)\right) = \frac{1}{2}, \tag{21}$$

$$\mathrm{tr}\left(\mathbf{\Pi_x}\frac{1}{2}(|++\rangle\langle++| + |--\rangle\langle--|)\right) = \frac{1}{2}. \tag{22}$$

The same result will be obtained; the probability that distinguishing whether the qubits sent by Alice belong to $\{|00\rangle, |11\rangle\}$ or $\{|++\rangle, |--\rangle\}$ is also $1/2$. When Bob measures the received two-photon pulses with this POVM and only keeps the qubits that can be identified by which ensemble they come from, whether he announces bit 0 or 1, Alice will get the conclusive bit at this position with the probability of $1/2$. For example, Bob distinguishes the received qubits as belonging to $\{|00\rangle, |11\rangle\}$. Bob announces bit 0, which means his measurement output should be $\{|00\rangle, |++\rangle\}$. In this case, if Alice sent $|00\rangle$, she cannot get a conclusive bit; if she sent $|11\rangle$, then she can infer that Bob measures it in the X basis and get a conclusive bit of 1. The same result will be obtained when Bob announces 1. As a result, whether Bob announces 1 or 0, Alice will get the conclusive bit 1 (Alice inferred Bob measured with the X basis) with a probability of $1/2$. Through this POVM, Bob can control Alice to obtain conclusive bits with the probability of $1/2$ and no error is introduced.

C. Combined attack on ETB protocol

The two POVMs above will make the proportion of Alice's conclusive bits in the raw key smaller or larger correspondingly, which will make Alice aware of Bob's cheating. To keep the proportion of Alice's conclusive bits in the raw key as $1/4$, we can combine the above two POVMs. Bob can then control conclusiveness of each raw key bit by randomly using the above two POVMs. In all POVMs, Bob disposes of all unidentifiable measurements. In the first POVM, Alice will get no conclusive bits and, in the second POVM, Alice

will get $1/2$ proportion of conclusive bits. In this way, the overall proportion of Alice's conclusive bits in the raw key is $1/4$ and no error is introduced. Consequently, the real-time detection in the protocol will not be able to detect Bob's cheating behavior.

In the following we analyze the amount of information Bob can obtain about Alice's privacy by the above attacks. Suppose the i'th bit in the final key K_f is the conclusive bit to Alice. Obliviously, the i'th bit in each of the k substrings is conclusive to Alice, where $k = k_1 \cdot k_2$ and k_1 and k_2 are the parameters introduced in step (6) of the ETB protocol. For any other bit in K_f except the i'th one, for example, the jth bit, if Bob is not sure that it is inconclusive to Alice, Bob should not be sure about the jth bit in each of the k substrings. The above happens with probability $1/2^k$. Therefore, for any bit other than the i'th one in K_f, Bob can confirm that Alice is inconclusive with the probability $1 - 1/2^k$. Then, the probability that Bob can confirm only the i'th bit could be conclusive to Alice is

$$\left[1 - \left(\frac{1}{2}\right)^k\right]^{N-1}. \tag{23}$$

In this case, the amount of information Bob gains is $\log_2 N - \log_2 1$. For the situation that there are m bits other than the i'th one in the final key K_f that Bob cannot confirm to be inconclusive for Alice, the probability can be calculated as

$$\binom{N-1}{m}\left[1 - \left(\frac{1}{2}\right)^k\right]^{N-1-m}\left(\frac{1}{2}\right)^{km}. \tag{24}$$

And the amount of information Bob gains changes to $\log_2 N - \log_2(m+1)$. Consequently, the mathematical expectation of the amount of information Bob gains can be calculated as

$$\begin{aligned} I_{\text{Bob}} &= \sum_{m=0}^{N-1}\binom{N-1}{m}\left[1 - \left(\frac{1}{2}\right)^k\right]^{N-1-m} \\ &\quad \times \left(\frac{1}{2}\right)^{k\cdot m}\log_2\left(\frac{N}{m+1}\right) \\ &> \left[1 - \left(\frac{1}{2}\right)^k\right]^{N-1}\log_2 N. \end{aligned} \tag{25}$$

By the above calculation, we arrive at the minimum value of I_{Bob}. In practice, I_{Bob} is very close to $\log_2 N$ because of the large value of the parameter k. According to ETB protocol, we compute the appropriate parameter values (see Sec. IV, subsection B in [39]) and give a simulation to estimate dishonest Bob's advantage in Algorithm 1. The concrete results of the simulation are presented in Table I.

Since no new errors are introduced under the above attack, we do not need to consider the effect of the error rate in the raw key on the protocol. The result will be the same as discussed in ETB protocol. For the advantage of Bob, as shown in Table I, I_{Bob} is very close to $\log_2 N$, which means Bob has full access to Alice's privacy. The last column of data in Table I is the theoretical lower bound of the ratio to $\log_2 N$ determined by Eq. (25), which is very close to 1 and agrees with the simulation results. The above results show a huge security vulnerability of ETB protocol on the light source side under multiphoton pulses. Next, we introduce the decoy method to confront the security threat of multiphoton pulses.

ALGORITHMS 1. Evaluation of Bob's advantage.

Input: N, p_1, p_2, k_1, k_2
Output: I_{Bob}

(P1) Define the input parameters. Here, N is the database size, p_1 is the proportion of the first POVM in all retained measurements, and p_2 is the proportion of the second POVM in all retained measurements. k_1 and k_2 are proper parameters for ETB protocol. Set $(k = k_1 \cdot k_2)$, then the raw key should contain kN bits and can be denoted by a $k \times N$ matrix $R = (r_{ij})$ with elements 0 or 1

(P2) Generate Bob's attacking strategy. A $k \times N$ matrix $B = (b_{ij})$ is generated according to p_1 and p_2. Based on the preceding attack strategy, set $p_1 = p_2 = 0.5$. For each element b_{ij}, set the value 0 and 0.5 to it with probability p_1 and p_2, respectively

(P3) Generate the conclusiveness of raw key. A $k \times N$ matrix $A = (a_{ij})$ is generated according to Bob's strategy B, that is, the value of a_{ij} is set with value 1 (0) with probability b_{ij} $(1 - b_{ij})$. Here, $a_{ij} = 1$ (0) means that Alice obtains a conclusive (inconclusive) (r_{ij}) in the raw key

(P4) Shift according to matrix *A*. Select a random positive integer $i' \in \{1, 2, \ldots, N\}$. For $i = 1, 2, \ldots, k$, select a value $j \in \{1, 2, \ldots, N\}$ such that $a_{ij} = 1$ (that is, r_{ij} is Alice's conclusive bit), then set $s = i' - j$ and finally shift the ith row of B by s bits

(P5) Compute Bob's advantage. Compute $q_j = \prod_{i=1}^{k} b_{ij}$ for $j = 1, 2, \ldots, N$, then set $p_i = q_i / \sum_{j=1}^{N} q_j$ for $i = 1, 2, \ldots, N$, and finally output $I_{\text{Bob}} = \log_2 N - (-\sum_{i=1}^{N} p_i \log_2 p_i)$

IV. DECOY-STATE QKD-BASED QUANTUM PRIVATE QUERY WITH ERROR TOLERANCE BOUND

A. Theoretical elaboration and calculation

Inspired by the idea [11,42] that the sender divides the transmitted pluses into signal state and decoy state according to the intensity of the light source in the decoy-state QKD protocol, we propose a decoy-state method for QPQ to help Alice detect whether Bob has performed the multiphoton attack. We take the weak coherent source used in most decoy-state QKD protocols for analysis. Weak coherent source (WCS) is an attenuated laser source with few vacuum signals and

TABLE I. For different database size N, proper choices of k_1 and k_2 when the error rate in the raw key (e_r) is equal to 0.03. I_{Bob} is the advantage Bob obtains under the condition of multiphoton pulses, collected in 100 runs of Algorithm 1.

N	k_1	k_2	I_{Bob}	$\log_2 N$	$[(1 - 1/2^k)]^{N-1}$
10^3	6	3	9.965784	9.965784	0.996196
5×10^3	8	4	12.287712	12.287712	0.999999
10^4	9	5	13.287712	13.287712	0.999999
5×10^4	10	5	15.609640	15.609640	0.999999
10^5	11	5	16.609640	16.609640	0.999999

multiphoton signals. At present, basically most quantum secure communication protocols using a single photon as the light source use a weak coherent source for attenuation to simulate a single photon light source and carry out relevant experiments.

In the decoy-state method, Alice has three kinds of light sources and randomly chooses light sources to emit the pulses. These three sources are the vacuum S_0, a weak decoy source S_ν, and a signal source S_μ, where the subscript is the parameter of the Poisson distribution and $\nu \ll 1$ and $\mu = O(1)$. Since the number of photons contained in a pulse in a weak coherent source follows the Poisson distribution with the parameter μ, the probability of a pulse containing n photons can be expressed as

$$P_n = \frac{\mu^n}{n!} e^{-\mu}. \tag{26}$$

Assuming the probability that the receiver detects a single photon is η, then the transmission efficiency of the n-photon pulse is

$$\eta_n = 1 - (1 - \eta)^n. \tag{27}$$

When Bob receives a pulse, he can only discriminate the number of photons contained in the pulse, not which light source the pulse came from. Suppose the yields of the three sources are γ_0, γ_ν, and γ_μ, corresponding to S_0, S_ν, and S_μ, respectively. Note that γ_0 is also the dark count rate of the system. We can get the yield of two light sources with different parameters:

$$\gamma_\nu = e^{-\nu}\gamma_0 + \nu\, e^{-\nu}\gamma_1 + o(\nu\, e^{\nu}), \tag{28}$$

$$\gamma_\mu = e^{-\mu}\gamma_0 + \mu\, e^{-\mu}\gamma_1 + (1 - e^{-\mu} - \mu\, e^{-\mu})\gamma_m, \tag{29}$$

where γ_1 is the yield of the single-photon pulses, i.e., the rate of the pulses, which Bob claims to have successfully received in the single-photon pulses sent by Alice. γ_m is the yield of the multiphoton pulses. The multiphoton pulses of S_ν can be ignored because $\nu \ll 1$. Note that the yield definition here is the rate that Bob claims he has successfully received, not the rate of the signals that he has actually received. This means that γ_0, γ_ν, and γ_μ may not be the actual yields of the three light sources for dishonest Bob. Based on the successfully detected pulses claimed by Bob, Alice will obtain the yields γ_ν and γ_μ of the light sources S_ν and S_μ in the transmission. Therefore, Alice can calculate γ_1 and γ_m based on Eqs. (28) and (29), which means that she can estimate the proportion of multiphoton pulses in the sequence provided to her by Bob.

However, since γ_1 and γ_m are relevant, Alice is able to calculate the theoretical value of γ_m based only on the dark count rate and γ_1. In fact, the yield of the pulse can be divided into two parts—the transmission efficiency corresponding to the number of photons and the dark count rate γ_0. Thus γ_1 and γ_m are also able to be formulated as

$$\gamma_1 = \eta + \gamma_0 - \eta\gamma_0, \tag{30}$$

$$\gamma_n = \eta_n + \gamma_0 - \eta_n\gamma_0. \tag{31}$$

According to Eqs. (28), (30), and (31), we can get a theoretical value of γ_m by γ_0:

$$\begin{aligned}\gamma_m' &= \frac{\sum_{n=2}^{\infty} \frac{\mu^n}{n!} e^{-\mu}\gamma_n}{1 - e^{-\mu} - \mu\, e^{-\mu}} \\ &= 1 - \frac{1 - e^{-\mu(1-\eta)} - \mu(1-\eta)e^{-\mu(1-\eta)}}{1 - e^{-\mu} - \mu\, e^{-\mu}} e^{(1-\eta)}(1 - \gamma_0).\end{aligned} \tag{32}$$

By comparing the difference between γ_m and γ_m', Alice can estimate the probability of Bob performing a multiphoton attack through parameter estimation and hypothesis testing theory. For the light source S_μ, suppose Alice emits N pulses; then the number of multiphoton pulses in them is

$$N_m = (1 - e^{-\mu} - \mu\, e^{-\mu})N. \tag{33}$$

According to the central-limit theorem, since N is large enough, the binomial distribution can be approximated as a normal distribution:

$$X = \frac{N_m(\gamma_m - \gamma_m')}{\sqrt{N_m\gamma_m'(1 - \gamma_m')}} \sim N(0, 1). \tag{34}$$

Based on the statistic X, Alice can estimate the probability of Bob cheating with the help of the standard normal distribution table.

B. Practical security analysis of decoy-state method for ETB protocol

When γ_m and γ_m' are close, the proportion of multiphoton pulses will be limited. In order to give Bob a greater advantage, we assume that the multiphoton pulses are all inconclusive bits of Alice under Bob's control. For the single-photon signals, we use the attack method mentioned in [39], that is, Bob uses basis $\{|0'\rangle, |1'\rangle\}$ to measure the single-photon portion, where

$$|0\rangle' = \cos\left(\frac{\pi}{8}\right)|0\rangle + \sin\left(\frac{\pi}{8}\right)|1\rangle, \tag{35}$$

$$|1\rangle' = \sin\left(\frac{\pi}{8}\right)|0\rangle - \cos\left(\frac{\pi}{8}\right)|1\rangle. \tag{36}$$

In single-photon measurements, Bob can control the conclusiveness of each raw key bit with probability 0.8536 at the cost of inducing an error rate of 0.5 for the attacked raw key bits (see Sec. IV, subsection B in [39]). Suppose P_m is the proportion of the multiphoton pulses and, for the single-photon pulses, Bob makes proportion p_1 (p_2) of the raw key bits known by Alice with probability 0.8536 (0.1464). The following formulas should be satisfied:

$$\frac{0.5(0.8536p_1 + 0.1464p_2)}{0.8536p_1 + 0.1464p_2 + 0.25(1 - p_m - p_1 - p_2)} \leqslant \varepsilon, \tag{37}$$

$$0.8536p_1 + 0.1464p_2 + 0.25(1 - p_m - p_1 - p_2) = \frac{1}{4}. \tag{38}$$

Equation (37) is to make the overall error rate lower than the threshold ε so that the protocol can proceed normally and Eq. (38) is to ensure that the proportion of Alice's conclusive bits in the raw key is consistent with the normal protocol so as not to cause suspicion.

TABLE II. Advantage of Bob under different e_r's and p_m. Here $N = 10000$, $k_1 = 9$, and $k_2 = 5$. The data in the table are the maximum, minimum, and average values of Bob's advantage from top to bottom. Collected in 100 runs of the modified Algorithm 1.

e_r		0.01	0.02	0.03	0.04	0.05	0.06
p_m	1%	1.5225	2.4687	3.5167	5.7485	11.2242	12.1353
		1.1649	1.6174	1.9516	2.3917	2.8068	3.0409
		1.2687	1.8647	2.4156	2.9790	3.6270	4.3288
	0.5%	1.6637	2.0228	3.9913	4.4178	8.7629	7.3593
		0.8804	1.3085	1.8993	2.2397	2.6182	2.9907
		0.9931	1.6139	2.3066	2.8947	3.6098	4.0974
	0.1%	1.3586	2.4208	3.6477	3.7113	6.8099	8.3920
		0.6131	1.0919	1.6784	1.9607	2.5225	2.8628
		0.6721	1.2639	1.8660	2.3588	3.0082	3.5643

In the following, we analyze the advantages of Bob in the improved protocol using simulations. Slightly different from the description of Algorithm 1 in Sec. III, in this attack, there are four kinds of bits in the raw key whose proportions correspond to p_m, p_1, p_2, and $p_3 = 1 - p_m - p_1 - p_2$. (Here, p_3 is the proportion of single-photon pulses that are not attacked.) In the multiphoton pulses, Bob controls the corresponding bits to be Alice's inconclusive bits. The matrix B in Algorithm 1 is generated according to p_m, p_1, p_2, and p_3. We choose $p_1, p_2 \in [0, 1 - p_m]$ with step length $\frac{1-p_m}{100}$ so that Bob can gain more advantage, under the restriction of Eqs. (37) and (38). For $N = 10\,000$, we simulate for 100 times.

The advantages gained by Bob under different p_m are shown in Table II. Obviously, same as the ETB protocol, I_{Bob} grows with the increase of e_r due to the allowed error rate induced by Bob's attack in the single-photon pluses. We can see that when $p_m = 1\%$ and $e_r \ll 0.03$, Bob generally will get less than 4 bits information about the retrieval address, while the retrieval address here contains $\log_2 N = 13.2877$ bits. In this case, according to [39], it is difficult for Bob to threaten user Alice's privacy.

In order to verify the practicability of the decoy-state method, we then compare the simulation results of Bob's advantages under the decoy-state method with those (Table IV in [39]) in ETB protocol. We can see the details in Fig. 1. When $P_m = 0.1\%$, the mean of the advantages obtained by Bob are approximate for the above described attack and the single-photon attack only (i.e., the attack in [39]). Obviously, under the defense of the decoy-state method, Bob gains an acceptable advantage. However, in attacks that retain only multiphoton pulses (see Table I), Bob gains an advantage close to $\log_2 N$. The results show that the decoy-state method can greatly reduce the vulnerability caused by multiphoton pulses.

Next, to find the maximum qualified value of p_m, we present the trend of the advantage obtained by Bob under different P_m's (see Fig. 2). In the figure, we can see that I_{Bob} grows linearly as p_m grows. To satisfy that the average advantage obtained by Bob is less than 4 bits, p_m should be bounded to 0.03 when $e_r = 0.03$. The upper bounds of the corresponding p_m's for different e_r's are given in Table III.

For the choice of the parameters μ and ν, Table 5 in [42] shows the details. When $\eta \geqslant 0.1$, we consider μ to be 0.3

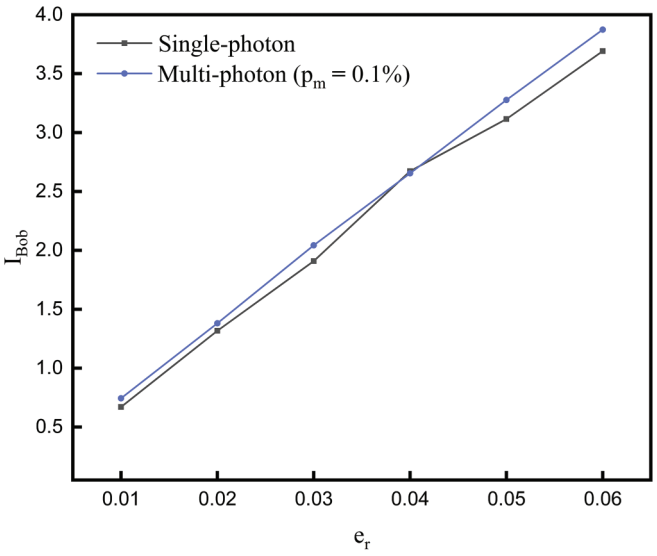

FIG. 1. Relationship between the e_r and I_{Bob} under the condition of single-photon pulses only and multiphoton pulses ($p_m = 0.1\%$). Here, $N = 10\,000$, $k_1 = 9$, and $k_2 = 5$. The error rate in the row key $e_r = 0.03$.

to balance the transmission efficiency and the proportion of multiphoton pulses. In addition, ν should be less than 0.01, so that the multiphoton pulse proportion of S_ν is less than 1%, which is negligible.

Compared to previously proposed protocols aiming to reconcile discrepancies between quantum communication devices under current technological constraints and theoretical models in QKD-based QPQ, our current work exhibits superior comprehensiveness. As evident from Table IV, our protocol stands alone in simultaneously addressing security vulnerabilities stemming from transmission loss, channel noise, and nonideal light sources.

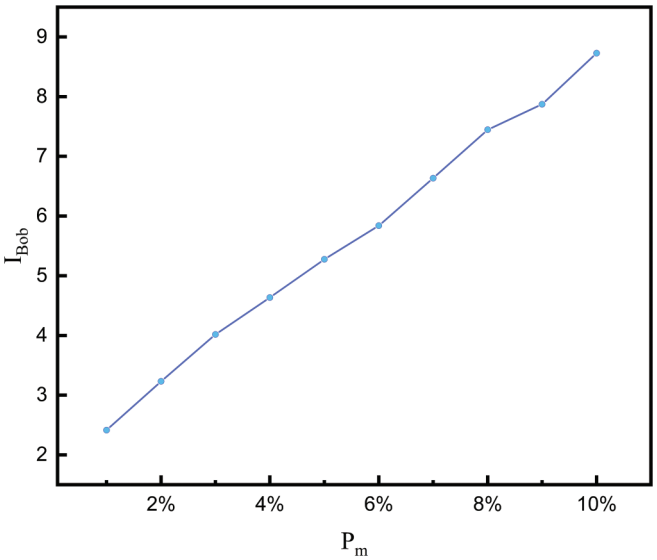

FIG. 2. Relationship between the p_m and I_{Bob} under multiphoton pulses. Here, $N = 10\,000$, $k_1 = 9$, and $k_2 = 5$. The error rate in the row key $e_r = 0.03$.

TABLE III. Average advantage obtained by Bob under different e_r's. p_m is the corresponding maximum proportion of multiphoton pulses in the transmission. Here $N = 10000$, $k_1 = 9$, and $k_2 = 5$.

e_r	0.01	0.02	0.03	0.04	0.05	0.06
I_{Bob}	4.0148	3.9463	4.0172	3.8536	3.7471	3.9053
p_m	5%	4%	3%	2%	1%	0.4%

TABLE IV. Comparison of our protocol with the aforementioned QKD-based QPQ protocols.

	Transmission loss	Nonideal sources	Channel noise
Protocol of [28]	√	×	×
Protocol of [39]	√	×	√
Protocol of [42]	√	√	×
Our protocol	√	√	√

V. CONCLUSION

Through the above analysis, we rearrange the steps of the decoy-state QKD-based QPQ protocol.

In step (1), Alice has three light sources and randomly selects one of them to emit weak coherent pulses into the channel, where S_0 is the vacuum, S_μ is the signal source, and S_ν is the weak decoy source.

In step (4), Alice checks γ_0, γ_ν, and γ_μ after they discard the records of checking qubits. Based on the statistics in Sec. IV, Alice estimates the proportion of multiphoton pulses p_m and the probability of Bob having cheated. If p_m is unacceptable, the protocol aborts.

The other steps are the same as the original protocol.

Due to the differences between practical quantum communication devices and theoretical models, the QKD-based QPQ protocol still needs practical security analysis before application. Our current work focuses on the practical security of the QKD-based QPQ protocol. We first review the ETB scheme that is able to deal with the channel noise. Then we carry out the practical security analysis of the ETB scheme under weak coherent pulses and find that multiphoton pulses will bring a great threat to the light source side. Finally, we propose the decoy-state method to resist the multiphoton attacks during transmission. The analysis confirms that the decoy-state method is effective and is significant to protect the security of the user in the ETB scheme. The results show that the decoy-state method can greatly limit the threat of multiphoton pulses to the light source and can continue to maintain the corresponding "cheat-sensitivity" security. The improved ETB scheme can not only work well via noisy channel, but also keep the balance of the reliability, database security, and user privacy under a nonideal light source. This implementation will memorably promote the practical application process of the QKD-based QPQ protocol.

Our present work theoretically realizes QPQ protocols tailored for authentic physical light sources and real-world channels. Nonetheless, achieving a truly practical QPQ protocol demands more. There remains a noticeable gap in exploring the defector's perspective. Prior studies have tackled security vulnerabilities on the detector side by employing the measurement-device-independent communication mode [35,36]. The subsequent step ought to involve integrating the advancements from both fronts to forge a more pragmatic QPQ protocol.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (Grant No. 2022YFC3801700), the National Natural Science Foundation of China (Grant No. 62171418), Open Foundation of State key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) (Grant No. SKLNST-2022-1-03), Sichuan Science and Technology Program (Grant No. 2023JDRC0017), and Natural Science Foundation of Henan Province (Grant No. 212300410062).

[1] C. H. Bennett *et al.*, *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing* (IEEE, New York, 1984).
[2] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).
[3] C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).
[4] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, Phys. Rev. A **72**, 012326 (2005).
[5] T.-G. Noh *et al.*, Phys. Rev. Lett. **103**, 230501 (2009).
[6] T. Sasaki, Y. Yamamoto, and M. Koashi, Nature (London) **509**, 475 (2014).
[7] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, Rev. Mod. Phys. **92**, 025002 (2020).
[8] G.-J. Fan-Yuan, S. Wang, Z.-Q. Yin, W. Chen, D.-Y. He, G.-C. Guo, and Z.-F. Han, Quantum Eng. **2**, e56 (2020).
[9] L.-C. Kwek, L. Cao, W. Luo, Y. Wang, S. Sun, X. Wang, and A. Q. Liu, AAPPS Bull. **31**, 15 (2021).
[10] C. Zhou, X. Wang, Z. Zhang, S. Yu, Z. Chen, and H. Guo, Sci. China Phys. Mech. Astron. **64**, 260311 (2021).
[11] W. Zhang, T. van Leent, K. Redeker, R. Garthoff, R. Schwonnek, F. Fertig, S. Eppelt, W. Rosenfeld, V. Scarani, C. C.-W. Lim *et al.*, Nature (London) **607**, 687 (2022).
[12] M. V. Panduranga Rao and M. Jakobi, Phys. Rev. A **87**, 012331 (2013).
[13] M. Hillery, V. Bužek, and A. Berthiaume, Phys. Rev. A **59**, 1829 (1999).
[14] F. Wu, G. Yang, H. Wang, J. Xiong, F. Alzahrani, A. Hobiny, and F. Deng, Sci. China Phys. Mech. Astron. **60**, 120313 (2017).
[15] G.-L. Long and X.-S. Liu, Phys. Rev. A **65**, 032302 (2002).
[16] F.-G. Deng and G.-L. Long, Phys. Rev. A **68**, 042315 (2003).
[17] R. Amiri, P. Wallden, A. Kent, and E. Andersson, Phys. Rev. A **93**, 032325 (2016).
[18] H.-L. Yin, Y. Fu, C.-L. Li, C.-X. Weng, B.-H. Li, J. Gu, Y.-S. Lu, S. Huang, and Z.-B. Chen, Natl. Sci. Rev. **10**, nwac228 (2023).
[19] K. Thapliyal and A. Pathak, Quantum Inf. Process. **18**, 191 (2019).

[20] X.-Y. Cao, B.-H. Li, Y. Wang, Y. Fu, H.-L. Yin, and Z.-B. Chen, Sci. Adv. **10**, eadk3258 (2024).
[21] P. W. Shor, *Proceedings 35th Annual Symposium on Foundations of Computer Science* (IEEE, New York, 1994), pp. 124–134.
[22] Y. Dulek, A. B. Grilo, S. Jeffery, C. Majenz, and C. Schaffner, *Annual International Conference on the Theory and Applications of Cryptographic Techniques* (Springer, Berlin, 2020), pp. 729–758.
[23] M. O. Rabin, Technical Report TR-81, Aiken Computation Lab, Harvard University, 1981.
[24] H.-K. Lo, Phys. Rev. A **56**, 1154 (1997).
[25] V. Giovannetti, S. Lloyd, and L. Maccone, Phys. Rev. Lett. **100**, 230502 (2008).
[26] L. Olejnik, Phys. Rev. A **84**, 022313 (2011).
[27] F. Yu, D. Qiu, X. Wang, Q. Li, L. Li, and J. Gruska, Theor. Comput. Sci. **807**, 330 (2020).
[28] M. Jakobi, C. Simon, N. Gisin, J.-D. Bancal, C. Branciard, N. Walenta, and H. Zbinden, Phys. Rev. A **83**, 022301 (2011).
[29] F. Gao, B. Liu, Q.-Y. Wen, and H. Chen, Opt. Express **20**, 17411 (2012).
[30] J.-L. Zhang, F.-Z. Guo, F. Gao, B. Liu, and Q.-Y. Wen, Phys. Rev. A **88**, 022334 (2013).
[31] Y.-G. Yang, S.-J. Sun, P. Xu, and J. Tian, Quantum Inf. Process. **13**, 805 (2014).
[32] F. Gao, B. Liu, W. Huang, and Q.-Y. Wen, IEEE J. Sel. Top. Quantum Electron. **21**, 98 (2014).
[33] B. Liu, F. Gao, W. Huang, and Q. Wen, Sci. China-Phys. Mechanics Astronomy **58**, 100301 (2015).
[34] S.-W. Xu, Y. Sun, and S. Lin, Quantum Inf. Process. **15**, 3301 (2016).
[35] L.-Y. Zhao, Z.-Q. Yin, W. Chen, Y.-J. Qian, C.-M. Zhang, G.-C. Guo, and Z.-F. Han, Sci. Rep. **7**, 39733 (2017).
[36] A. Maitra, G. Paul, and S. Roy, Phys. Rev. A **95**, 042344 (2017).
[37] C.-Y. Wei, X.-Q. Cai, B. Liu, T.-Y. Wang, and F. Gao, IEEE Trans. Comput. **67**, 2 (2018).
[38] F. Gao, S. Qin, W. Huang, and Q. Wen, Sci. China Phys. Mech. Astron. **62**, 070301 (2019).
[39] C.-Y. Wei, X.-Q. Cai, T.-Y. Wang, S.-J. Qin, F. Gao, and Q.-Y. Wen, IEEE J. Sel. Areas Commun. **38**, 517 (2020).
[40] V. Scarani, A. Acin, G. Ribordy, and N. Gisin, Phys. Rev. Lett. **92**, 057901 (2004).
[41] P. Chan, I. Lucio-Martinez, X. Mo, C. Simon, and W. Tittel, Sci. Rep. **4**, 5233 (2014).
[42] B. Liu, S. Xia, D. Xiao, W. Huang, B. Xu, and Y. Li, Sci. China Phys. Mech. Astron. **65**, 240312 (2022).
[43] L. Qin, B. Liu, F. Gao, W. Huang, B. Xu, and Y. Li, Physica A **633**, 129427 (2024).
[44] Y.-F. Jiao, W. Huang, B. Liu, W.-Z. Shao, Z.-D. Shen, and B.-J. Xu, Quantum Inf. Process. **23**, 133 (2024).
[45] B. Huttner, N. Imoto, N. Gisin, and T. Mor, Phys. Rev. A **51**, 1863 (1995).