Source attack of decoy-state quantum key distribution using phase information

Yan-Lin Tang,[1] Hua-Lei Yin,[1] Xiongfeng Ma,[2,*] Chi-Hang Fred Fung,[3,†] Yang Liu,[1] Hai-Lin Yong,[1] Teng-Yun Chen,[1,‡] Cheng-Zhi Peng,[1] Zeng-Bing Chen,[1] and Jian-Wei Pan[1,§]

[1]*Hefei National Laboratory for Physical Sciences at Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei, Anhui, China*

[2]*Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China*

[3]*Department of Physics and Center of Theoretical and Computational Physics, University of Hong Kong, Pokfulam Road, Hong Kong*

(Received 22 March 2013; published 8 August 2013)

Quantum key distribution (QKD) utilizes the laws of quantum mechanics to achieve information-theoretically secure key generation. This field is now approaching the stage of commercialization, but many practical QKD systems still suffer from security loopholes due to imperfect devices. In fact, practical attacks have successfully been demonstrated. Fortunately, most of them only exploit detection-side loopholes, which are now closed by the recent idea of measurement-device-independent QKD. On the other hand, little attention is paid to the source, which may still leave QKD systems insecure. In this work, we propose and demonstrate an attack that exploits a source-side loophole existing in qubit-based QKD systems using a weak coherent state source and decoy states. Specifically, by implementing a linear-optics unambiguous state discrimination measurement, we show that the security of a system without phase randomization—which is a step assumed in conventional security analyses but sometimes neglected in practice—can be compromised. We conclude that implementing phase randomization is essential to the security of decoy-state QKD systems under current security analyses.

DOI: 10.1103/PhysRevA.88.022308 PACS number(s): 03.67.Dd, 03.67.Ac, 03.67.Hk, 42.50.−p

I. INTRODUCTION

Quantum key distribution (QKD) aims at offering information-theoretical security for secret key expansion [1,2] that is guaranteed by quantum mechanics. Commercial QKD systems have emerged on the market and are now under rapid development. Despite theoretical security proofs, various quantum hacking strategies targeting practical QKD systems have been proposed, with some of them demonstrated in experiments. These attacks exploit certain imperfections in the devices used to build QKD systems. Except for the phase-remapping attack [3,4], which aims at the source of plug-and-play QKD systems, until now, most practical attacks have been launched on the detection side of QKD systems, including the fake-state attack [5,6], the time-shift attack [7,8], and the detector-blinding attack [9,10].

In order to achieve security when imperfect (untrusted) devices are present, QKD schemes [11,12] that are fully device independent (without assumptions on either the detector or the source) have been proposed. However, these schemes suffer from their unrealistic requirement for a high transmission efficiency (with the lowest to date being 75% [13]), which limits their use in practice. Recently, the newly proposed measurement-device-independent QKD (MDI-QKD) scheme [14], whose security does not rely on any assumptions on the detection system, can defeat all aforementioned detection-side attacks. On the source side, the security proof of the MDI-QKD scheme relies on a trusted-source scenario, whose security concern is relatively less explored. Besides, many other security proofs [15,16] also rely on stringent assumptions on the source side. Any deviation from these assumptions may lead to loopholes that can be exploited for eavesdropping.

The use of different sources directly affects the security of QKD systems. Systems implementing the popular Bennett-Brassard 1984 (BB84) protocol [1] often use a weak coherent state (WCS) source instead of a single-photon source for the transmission of quantum states. Fortunately, such a substitution of the source is safe but its security proof [15,17] assumes that the phase of the source is randomized, without which the security would be weakened [18]. Later security proof by Lo and Preskill [16] eliminated the need for phase randomization, but with the performance substantially diminished.

Even though the security of the BB84 protocol with WCS is proven, there is a substantial performance gap between it and the case of a single-photon source. A significant achievement was made with the proposal of the decoy-state technique [15,19,20], which greatly improves the performance of the BB84 protocol with WCS, and thus decoy-state BB84 with WCS has become one of the most popular schemes for practical implementation. Again, phase randomization is assumed in the current security proof [15], and a security analysis for decoy-state QKD without phase randomization is not yet available [16]. This fact can easily be overlooked and QKD system designers often neglect the implementation of phase randomization without realizing the danger of opening up a security loophole. Indeed, we experimentally demonstrate that this is a major security loophole. We propose and demonstrate an attack on the source part of a decoy-state QKD system with WCS when phase randomization is not implemented. By using a combination of an unambiguous-state-discrimination (USD) measurement and a photon-number-splitting (PNS) attack [21], we show that the final key generated by the non-phase-randomized system can be compromised.

*xma@tsinghua.edu.cn
†chffung@hku.hk
‡tychen@ustc.edu.cn
§pan@ustc.edu.cn

II. HACKING STRATEGY

The essence of our hacking strategy is as follows. Since Alice prepares her pulses without phase randomization, we may assume that Eve knows the overall phase of every transmitted state by Alice in the worst-case scenario. Then, from Eve's point of view, the states sent by Alice are drawn from an ensemble of pure states (corresponding to the signal and decoy states). Thus, quantum mechanics allows Eve to distinguish between the signal and decoy states by a USD measurement. In this way, one of the foundations of the security proof of decoy-state QKD, the photon number channel model [15], is violated.

Eve's attack is composed of two parts, a USD measurement and a PNS attack strategy [21], as shown in Fig. 1. First, Eve performs a positive operator-valued measurement (POVM) to distinguish between a signal state and a decoy state without disturbing the quantum state sent by Alice (see Appendix A for details). Then she measures the photon number. Conditioned on her measurement results for the signal/decoy information, Eve may forward some photons to Bob so as to preserve the statistics of a normal quantum channel. For a single-photon state, Eve either blocks it or directly forwards it without knowing the qubit information, while for a multiphoton state, Eve can keep one copy and forward the rest, giving her the full qubit information (see Appendix B for details).

An attack is considered successful if Eve is able to trick Alice and Bob into accepting an insecure key. To show that our attack is successful, we compare the following two key rates: (i) a lower bound on the secure key rate from the perspective of Alice and Bob, who overlook phase randomization and apply the conventional decoy-state postprocessing, denoted R^l, and (ii) an upper bound on the secure key rate taking into account our attack, denoted R^u. The former situation assumes that phase randomization is performed (but is actually not performed) and uses the postprocessing scheme presented in Ref. [22]; the key rate lower bound for this case (details given in Appendix E) is

$$R^l = -Q_\mu H(E_\mu) + Y_1^s \mu e^{-\mu}\left[1 - H\left(e_1^s\right)\right], \tag{1}$$

where Q_μ and E_μ are the overall gain and quantum bit error rate (QBER); μ denotes the expected photon number of the signal state; Y_1^s and e_1^s are the yield and the error rate of the single-photon signal state, respectively,

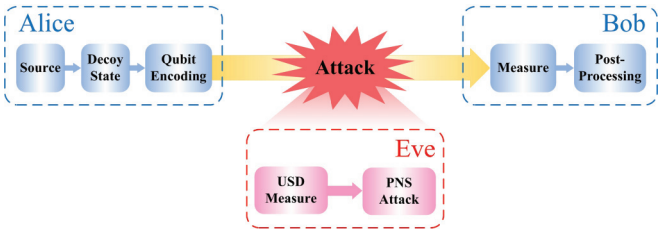

FIG. 1. (Color online) Schematic diagram of the USD + PNS attack on a decoy-state QKD system without phase randomization. Eve intercepts the quantum states sent by Alice and then performs a USD measurement to distinguish the signal/decoy state succeeded by a PNS attack. Conditioned on her results of signal/decoy information and photon numbers, she sets the yields smartly so that the detection statistics on Bob's side remains the same as the case without attacks.

which are estimated by the decoy-state method; and $H(e) = -e\log_2(e) - (1-e)\log_2(1-e)$ is the binary Shannon entropy function. Here, we assume that Alice and Bob run the efficient BB84 [23] and take the basis sift factor to be 1. On the other hand, our USD + PNS attack sets an upper bound on the key rate (details given in Appendix C):

$$R^u = Y_1^s e^{-\mu}\mu. \tag{2}$$

Note that different values of Y_1^s are used in the computation of R^l and R^u. The value of Y_1^s used in the lower bound, R^l, is the one estimated by Alice and Bob using conventional decoy-state processing, and the value used in the upper bound, R^u, is the one chosen by Eve in the attack. Since the lower bound represents the key rate at which Alice and Bob generate a new key that they think is secure, if this rate is higher than what is allowed after taking into account our attack, some of the new key must be insecure and Eve has some information about it. Thus, our attack is successful if

$$R^l > R^u \quad \text{(attack successful)}. \tag{3}$$

Eve's attack aims to preserve the measurement statistics of Bob in order to conceal the attack. We form the attack strategy as an optimization problem subject to preserving the gain statistics. On the other hand, we do not constrain the error statistics here since the error rate introduced by our attack demonstration is negligible. To get more details of this discussion, please see Appendix F.

III. EXPERIMENT SETUP

The PNS attack of our USD + PNS hacking strategy, which requires a quantum nondemolition measurement [24–27] on the photon numbers, a lossless channel, and the ability to control Bob's detector efficiency, is beyond current technology. Thus, we assume in the analysis that Eve has the ability to perform the PNS part and we only realize the USD measurement in our experiment.

To demonstrate the attack, we use a phase-encoding BB84 QKD system with strong phase-stabilization pulses [28–30]. The experimental setup is shown in Fig. 2. At Alice's site, a distributed feedback (DFB) diode with a central wavelength of 1550.12 nm and a pulse duration of 1 ns operates at a repetition rate of 4 MHz. Then the laser pulse passes through two asymmetric Mach-Zehnder interferometers (AMZIs). The first one splits the coherent pulse generated by the laser source into two time bins, one for the phase stabilization and the other for the quantum signal. Then, the decoy state is prepared by randomly modulating the intensity of the quantum signal. The intensities of the signal and decoy states, μ and ν, are set to 0.5 and 0.1, respectively. In the second AMZI, the phase modulator encodes the quantum pulse with one of the four BB84 phases, and it performs no action on the strong phase-stabilization pulse.

At Eve's site, both the phase-stabilization pulse and the quantum signal are split into two pulses with a 99:1 beam splitter of Eve's AMZI, which have the same splitting ratio and the same path-length difference as Alice's first AMZI. The fraction of the phase-stabilization light, passing through the arm with the intensity modulator, interferes with the fraction of the quantum signal, passing through the arm with the

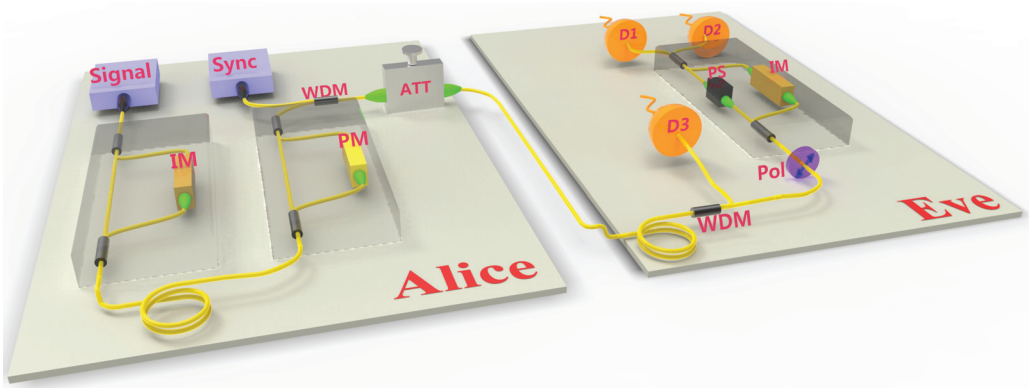

FIG. 2. (Color online) Schematic of our experiment setup for USD measurement demonstration. Alice's first AMZI splits the signal pulse into two pulses: a strong one for phase stabilization and a weak one for the quantum signal modulated by the intensity modulator (IM) to be either a decoy state or a signal state. The second AMZI encodes the BB84 states with a phase modulator (PM). Then a synchronization pulse is coupled with the signal pulses into a signal fiber and sent to Bob. At Eve's site, she utilizes a polarization controller and polarizer (Pol) to compensate for the polarization change and a phase shifter (PS) to compensate for the phase drift. Then she uses the same AMZI setup as Alice's first one to make the quantum pulse interfere with the strong phase-stabilization pulse modulated by the IM. The interference result either indicates the identity of the quantum pulse (signal or decoy) or is inconclusive.

phase shifter. With the intensity modulator, Eve can choose to measure either the signal or the decoy state. Eve's AMZI cancels the path delay between the phase-stabilization pulse and the quantum signal pulse, which is set by Alice's first AMZI, and makes the two pulses meet and interfere with each other. Identical AMZIs will make perfect interference, and in our experiment, we have achieved a high visibility of 500:1. The interference results can be divided into two cases: (i) if the pulses in the two arms have the same intensity, detector D2 does not click and the USD measurement result is inconclusive; and (ii) when detector D2 clicks, Eve can identify the state sent by Alice. The second case corresponds to a successful USD measurement outcome for Eve.

IV. RESULTS

The performance of our USD experiment is characterized by two sets of parameters, as listed in Table I. The first one is related to success probabilities. Since the overlap between the signal and decoy states is nonzero, Eve's USD measurement cannot succeed with unity probability and we denote the success probability when Alice sends a signal (decoy) state by q_μ (q_ν). The second set of parameters is related to error probabilities. Note that even when the USD measurement succeeds, experimental imperfection may cause the USD to report the wrong state. We denote the probability of correctly identifying the input state conditioned on a successful USD by ξ_μ (ξ_ν) when Alice sends the signal (decoy) state. These key parameters q_μ, q_ν, ξ_μ, and ξ_ν characterize the effectiveness of our USD attack from Eve's perspective and her ability to compromise the security of the QKD system. Details of these definitions can be found in Appendix C.

Several aspects of our experiment affect the success probabilities q_μ and q_ν. First is the fundamental indistinguishability of nonorthogonal quantum states. Our USD measurement acts only on the first pulse, and not on the second pulse, which encodes the phase information. The optimal USD to distinguish the two possibilities of the first pulse, $|\sqrt{\frac{\mu}{2}}\rangle$ and $|\sqrt{\frac{\nu}{2}}\rangle$, has maximal success probability (details given in Appendix A)

$$q_{\text{opt}} = 1 - \exp\left(-\tfrac{1}{4}|\sqrt{\nu} - \sqrt{\mu}|^2\right). \tag{4}$$

On the other hand, our linear-optics USD setup, shown in Fig. 2, achieves only $q_{\max} = q_{\text{opt}}/2$ even when the devices are perfect. It is an interesting question with regard to how to implement a USD measurement to achieve q_{opt}, especially with linear optics. When $\mu = 0.5$ and $\nu = 0.1$, one obtains $q_{\max} = 1.87\%$ and $q_{\text{opt}} = 3.75\%$. Experimental imperfections and inefficient detectors further reduce the actual success probability.

The measurement results for q_μ, q_ν, ξ_μ, and ξ_ν over a time period of 748 s are listed in Table II. Note that ξ_μ (ξ_ν) is near 100%, which indicates that we almost never made a mistake in identifying the state.

TABLE I. Ideal and experimental probabilities of the USD attack, conditioned on signal/decoy states sent by Alice.

State of	Ideal case			Experimental case		
Alice	Signal	Decoy	Failure	Signal	Decoy	Failure
Signal	$q_{\max}$	0	$1 - q_{\max}$	$q_\mu\xi_\mu$	$q_\mu(1-\xi_\mu)$	$1 - q_\mu$
Decoy	0	$q_{\max}$	$1 - q_{\max}$	$q_\nu(1-\xi_\nu)$	$q_\nu\xi_\nu$	$1 - q_\nu$

TABLE II. Experimental results. The standard deviations (in the time domain) of q_μ, q_ν, ξ_μ, and ξ_ν are, respectively, 4.3×10^{-5}, 4.0×10^{-5}, 0.98%, and 0.40%, which shows that the attack system is robust.

q_μ	q_ν	ξ_μ	ξ_ν	q_{opt}
1.18×10^{-3}	1.16×10^{-3}	96.90%	98.37%	3.75%

Using the experimental values for these parameters, we can derive the key rate upper bound as a function of the transmission loss between Alice and Bob, which is shown in Fig. 3. Also shown is the the lower bound of the key rate that Alice and Bob thought to be achievable with the assumption of phase randomization, which adopts some realistic parameters of Bob's setup with superconducting single-photon detectors [31]: dark count $Y_0 = 10^{-7}$, detection efficiency 5%, and misalignment error rate $e_d = 2.0\%$. The key result is that when the overall transmission loss is beyond 36.3 dB, the upper bound is below the lower bound as shown in Eq. (3), and thus our attack allows Eve to successfully steal the secret key.

To illustrate the potential power of the ideal USD + PNS attack, we consider the ideal USD measurement and take the success probabilities q_μ and q_ν to be the theoretically maximum of 3.75% and the correct distinguishing probabilities, ξ_μ and ξ_ν, to be 1. This upper bound curve is shown in Fig. 3, which indicates that when the overall loss between Alice and Bob is only beyond 21.2 dB, the decoy-state BB84 protocol with phase-nonrandomized WCS is insecure. For comparison,

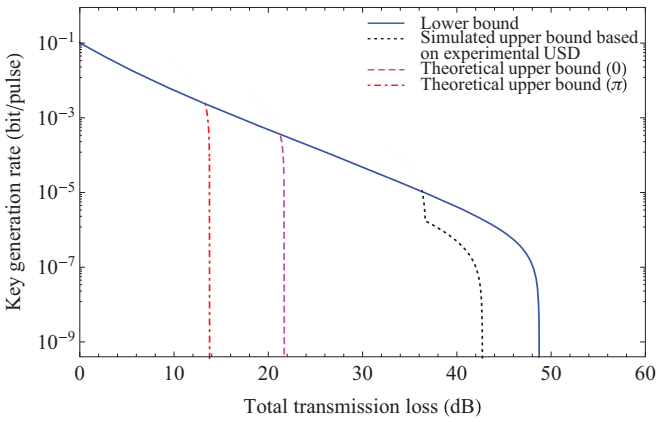

FIG. 3. (Color online) Bounds on the key generation rate. The lower bound, given in Eq. (1), is computed by ignoring the phase randomization problem. The simulated upper bound is evaluated by Eq. (2) and the data listed in Table II. The region for which the upper bound is below the lower bound corresponds to the secret key being stolen by Eve. Also shown are two best theoretical upper bounds using ideal values $q_\mu = q_\nu = 23.0\%, 3.75\%$ and $\xi_\mu = \xi_\nu = 1$, where the dash-dotted (dashed) curve corresponds to a setup with a relative phase of π (0) between signal and decoy pulses giving rise to $q_\mu = q_\nu = 23.0\%$ (3.75%). Upper bounds corresponding to other relative phases fall between these two curves. Note that in our experiment, the relative phase is 0. Our attack is successful when the lower bound is higher than the upper bound, which occurs when the transmission loss is larger than 36.3 dB (for our experiment), 21.2 dB (for the ideal situation with zero relative phase), and 13.3 dB (for the ideal situation with π relative phase).

an upper bound curve for the success probability of 23.0% corresponding to the relative phase between signal and decoy pulses of π is also shown. In essence, this figure shows that the potential impact of our attack can be significant and phase randomization cannot be neglected in decoy-state QKD.

V. DISCUSSION AND CONCLUSION

By exploiting the phase information of the signal and decoy states, our experimental attack succeeds in stealing the final secret key when the transmission loss is over a certain threshold. We prove that phase randomization cannot be neglected in decoy-state QKD using WCS, unless a new security proof is available. Our result also answers a long-standing question. Before our work, it was unclear whether performing phase randomization improves the key rate performance of decoy-state BB84 using a WCS. Our result implies that performing phase randomization is strictly better. We remark that our attack is not limited to the phase-encoding system with strong phase-stabilization pulses [28–30] on which our experiment is based. As long as the phase of each state, be it a signal or a decoy state, is known by Eve, she does not need the strong phase reference from Alice. Eve can simply prepare an auxiliary pulse with the corresponding phase. Therefore, this attack can be launched to hack a regular decoy-state QKD system without phase randomization.

A key feature of our experiment is the implementation of USD with linear optics. Even with only linear optics, this attack system can efficiently compromise the security of the key. Our work on applying USD with linear optics in quantum information opens an avenue to full linear-optics-based implementation of general quantum measurements, extending previous results [32–34]. For future work, it is an interesting perspective topic to study the case where Eve knows partial information on the phases. For example, in a QKD system with active phase randomization [35], the phase may not be perfectly random in practice. A related question will be whether a fully randomized phase is necessary for Alice and Bob to guarantee the security.

ACKNOWLEDGMENTS

We thank Y. Cao, W. F. Cao, L. J. Wang, and F. Zhou for enlightening discussions. This work was supported by the National Natural Science Foundation of China, the National Basic Research Program of China Grants, the CAS and USTC Special Grant for Postgraduate Research, Innovation and Practice, and Quantum Communication Technology Co., Ltd., Anhui. X.M. gratefully acknowledges the financial support from National Basic Research Program of China Grants No. 2011CBA00300 and No. 2011CBA00301, National Natural Science Foundation of China Grant No. 61033001, and the 1000 Youth Fellowship program in China. C.-H.F.F. gratefully acknowledges the financial support of RGC Grant No. 700712P from the HKSAR Government.

APPENDIX A: USD MEASUREMENT

Considering the one-decoy state protocol [22], where two coherent states are used by Alice, a signal state, $|\sqrt{\mu}e^{i\theta_s}\rangle$, and

a weak decoy state, $|\sqrt{\nu}e^{i\theta_d}\rangle$. Here, μ and ν are the intensities of the signal and decoy states, respectively, with $\mu > \nu$; θ_s and θ_d are the phases of the signal and decoy states, respectively. Since no phase randomization is performed and Alice does not intentionally put any phase difference between signal and decoy pulses, their phases are naturally the same and thus we take $\theta_s = \theta_d = 0$. (For the case where $\theta_s - \theta_d \neq 0$, the success probability for Eve's attack will increase as shown in Fig. 3.)

For the phase-encoding scheme, Alice encodes qubits in the relative phases of the two pulses separated by the second Mach-Zehnder interferometer, as shown in Fig. 2. For the BB84 protocol, she encodes $\phi \in \{0,\pi/2,\pi,3\pi/2\}$ randomly and sends out

$$\begin{aligned} |\psi_s\rangle &= \left|\sqrt{\frac{\mu}{2}}\right\rangle \left|e^{i\phi}\sqrt{\frac{\mu}{2}}\right\rangle, \\ |\psi_d\rangle &= \left|\sqrt{\frac{\nu}{2}}\right\rangle \left|e^{i\phi}\sqrt{\frac{\nu}{2}}\right\rangle \end{aligned} \tag{A1}$$

for signal and decoy states, respectively.

Since the signal and decoy states are not orthogonal (i.e., $\langle\psi_s|\psi_d\rangle \neq 0$), Eve cannot perfectly distinguish them with unity probability. Instead, she performs a USD measurement to perfectly distinguish them with probability <1. We impose that our USD measurement acts only on the first pulse and we leave the second pulse, which encodes the phase information, intact. This is because a measurement on the second pulse may destroy the qubit encoded in the relative phase. In this case, the failure probability corresponding to the optimal USD [36–38] (assuming equal *a prior* probabilities of $|\psi_s\rangle$ and $|\psi_d\rangle$, which is the case in our experiment) is given by the overlap between the states to be identified:

$$p_f = \left|\left\langle\sqrt{\frac{\nu}{2}}\middle|\sqrt{\frac{\mu}{2}}\right\rangle\right| = \exp\left(-\frac{1}{2}\left|\sqrt{\frac{\nu}{2}} - \sqrt{\frac{\mu}{2}}\right|^2\right). \tag{A2}$$

The corresponding POVM is

$$\begin{aligned} \hat{E}_\mu &= \frac{1}{(1+p_f)(1-p_f^2)} P\left(\left|\sqrt{\frac{\mu}{2}}\right\rangle - \left\langle\sqrt{\frac{\nu}{2}}\middle|\sqrt{\frac{\mu}{2}}\right\rangle\left|\sqrt{\frac{\nu}{2}}\right\rangle\right), \\ \hat{E}_\nu &= \frac{1}{(1+p_f)(1-p_f^2)} P\left(\left|\sqrt{\frac{\nu}{2}}\right\rangle - \left\langle\sqrt{\frac{\mu}{2}}\middle|\sqrt{\frac{\nu}{2}}\right\rangle\left|\sqrt{\frac{\mu}{2}}\right\rangle\right), \\ \hat{E}_f &= I - \hat{E}_\mu - \hat{E}_\nu, \end{aligned} \tag{A3}$$

where we define the projection function $P(|\varphi\rangle) = |\varphi\rangle\langle\varphi|$ for some state $|\varphi\rangle$. The measurement outcome $\hat{E}_\alpha$ indicates that the input state is $|\sqrt{\alpha/2}\rangle$, where $\alpha = \mu,\nu$; and the outcome $\hat{E}_f$ is inconclusive about whether the input state is a signal state or a decoy state. Note that the success probability corresponding to the optimal USD is

$$q_{\text{opt}} \triangleq \left\langle\sqrt{\frac{\mu}{2}}\middle|\hat{E}_\mu\middle|\sqrt{\frac{\mu}{2}}\right\rangle = \left\langle\sqrt{\frac{\nu}{2}}\middle|\hat{E}_\nu\middle|\sqrt{\frac{\nu}{2}}\right\rangle = 1 - p_f, \tag{A4}$$

and error does not occur since $\langle\sqrt{\mu/2}|\hat{E}_\nu|\sqrt{\mu/2}\rangle = \langle\sqrt{\nu/2}|\hat{E}_\mu|\sqrt{\nu/2}\rangle = 0$. The ideal probabilities of this optimal USD measurement outcomes are summarized in Table I. In our experiment, we implement the USD measurement with linear optics as shown in Fig. 2, and its maximal success probability, assuming 100% efficiency detectors, is given by

$$\begin{aligned} q_{\max} &= \frac{1 - \left|\langle 0|\sqrt{\frac{\mu}{4}} - \sqrt{\frac{\nu}{4}}\right|^2}{2} \\ &= \frac{1 - \exp\left(-\frac{|\sqrt{\mu}-\sqrt{\nu}|^2}{4}\right)}{2} = \frac{q_{\text{opt}}}{2}. \end{aligned} \tag{A5}$$

It is an interesting question to find a way to implement the optimal USD measurement corresponding to Eq. (A3) using linear optics.

APPENDIX B: PNS ATTACK

After the USD measurement, Eve measures the photon number of Alice's pulse and launches the PNS attack. The photon numbers of the two WCSs follow the Poisson distribution:

$$P_i^s = \frac{\mu^i e^{-\mu}}{i!}, \quad P_i^d = \frac{\nu^i e^{-\nu}}{i!}. \tag{B1}$$

Define the gains, Q_μ and Q_ν, respectively, to be the probabilities for Bob to get a detection event given that Alice sends signal and decoy states,

$$\begin{aligned} Q_\mu &= Y_0^s P_0^s + Y_1^s P_1^s + Y_2^s P_2^s + \cdots, \\ Q_\nu &= Y_0^d P_0^d + Y_1^d P_1^d + Y_2^d P_2^d + \cdots. \end{aligned} \tag{B2}$$

where Y_i is the yield of the i-photon state or the conditional probability for Bob to get a detection given that Alice sends out an i-photon state; the superscripts, s and d, denote the signal state and decoy state, respectively.

In the postprocessing of decoy-state QKD, the yield of the single-photon-state component can be inferred from the detection statistics of signal and decoy states on Bob's side. The underlying assumption of the photon number channel model for the security proof of decoy-state QKD [15] can be described as

$$Y_i^s = Y_i^d, \tag{B3}$$

which holds when the phases of signal and decoy states are randomized. In our USD + PNS attack, Eq. (B3) is violated when Eve is able to distinguish between the signal and decoy states by a USD measurement given the phase information of the coherent states. (Note that Eq. (B3) may also be violated even when only partial phase information is known to Eve [39].) She smartly chooses these proportions (Y_i^s and Y_i^d) so that her attack will not be detected. This is achieved by maintaining the same observed gain statistics (Q_μ and Q_ν) as in the normal situation.

APPENDIX C: KEY RATE UPPER BOUND

In the USD attack, Eve performs the POVM as shown in Fig. 2. Conditioned on these results, Eve sets a different yield (detection probability) for each i-photon state. Define q_μ (q_ν) to be the conditional probability for Eve to result in a successful measurement outcome when Alice sends out a signal (decoy) state. Both experimental success probabilities, q_μ and q_ν, are limited by the theoretical maximum,

$$q_\mu, q_\nu \leqslant q_{\max}, \tag{C1}$$

TABLE III. Probabilities of these POVM outcomes, conditioned on different intensity states sent by Alice, and yields for different POVM outcomes and photon numbers i.

POVM	E_μ	E_ν	E_f
Signal	$q_\mu\xi_\mu$	$q_\mu(1-\xi_\mu)$	$1-q_\mu$
Decoy	$q_\nu(1-\xi_\nu)$	$q_\nu\xi_\nu$	$1-q_\nu$
Yields	Z_i^μ	Z_i^ν	X_i

where $q_{\max}$ given in Eq. (A5) can be achieved when Eve's detection efficiency is 100%.

In practice, even when Eve obtains a successful measurement outcome, she might make an error in determining whether the state is a signal or a decoy state. Define

$$\xi_\mu \triangleq \text{Prob}(E_\mu|\text{signal}), \quad \xi_\nu \triangleq \text{Prob}(\text{E}_\nu|\text{decoy}) \tag{C2}$$

to be the conditional probabilities for Eve to guess Alice's state correctly when Eve obtains successful measurement outcome and Alice sends out a signal (decoy) state. The relationship between these probabilities when Alice sends out a signal and decoy state is shown in Table III. This table appears as part of Table I, which also shows the ideal probabilities of the optimal USD measurement.

Define Z_i^μ, Z_i^ν, and X_i to be the yields of the i-photon state, conditioned on Eve getting the measurement outcome E_μ, E_ν, and E_f, respectively, as listed in the last row in Table III. The yield is the probability that a valid detection occurs when Eve sends a pulse to Bob after the attack. We assume that Eve sets Z_0^μ, Z_0^ν, and X_0 to 0. This is because if Eve forwards any photon to Bob when she gets a vacuum state, she may introduce errors. The quantum no-cloning theorem tells us that when the photon number is 1, Eve is unable to keep a copy of the qubit information, then the yields Z_i^μ, Z_i^ν, and X_i will enable Bob to generate secure keys from this one-photon state component.

In our attack, we set $X_i = 0$ for all i, for the following reason. Since our implementation of the USD measurement destroys the quantum reference pulse, if the USD outcome is inconclusive (i.e., E_f), Eve cannot always choose the right intensity for the regenerated reference pulse. This increases the error rate for Bob, which alerts Alice and Bob to Eve's presence. Thus, in order to avoid this, we design Eve's attack so that when she fails to learn the state intensity, she does not forward any pulse to Bob in order to emulate a channel loss. This means that $X_i = 0$ for all i, which we assume for the remaining analysis. Note that if a nondemolition method is used to identify the intensity, no additional errors are introduced and thus there is no need to set $X_i^\mu = X_i^\nu = 0$ for all i.

Thus, the yields Y_i^s and Y_i^d, from Bob's point of view, are composed of the two successful outcomes of Eve as listed in Table III:

$$\begin{aligned} Y_i^s &= q_\mu\left[\xi_\mu Z_i^\mu + (1-\xi_\mu)Z_i^\nu\right], \\ Y_i^d &= q_\nu\left[\xi_\nu Z_i^\nu + (1-\xi_\nu)Z_i^\mu\right]. \end{aligned} \tag{C3}$$

Then, by inserting Eq. (C3) into Eq. (B2), the gains of signal and decoy states are given by

$$\begin{aligned} Q_\mu &= \sum_{i=1}^{\infty} q_\mu\left[\xi_\mu Z_i^\mu + (1-\xi_\mu)Z_i^\nu\right]e^{-\mu}\frac{\mu^i}{i!}, \\ Q_\nu &= \sum_{i=1}^{\infty} q_\nu\left[\xi_\nu Z_i^\nu + (1-\xi_\nu)Z_i^\mu\right]e^{-\nu}\frac{\nu^i}{i!}. \end{aligned} \tag{C4}$$

For a normal quantum channel, Alice and Bob should get

$$Q_\mu = 1-e^{-\eta\mu}, \quad Q_\nu = 1-e^{-\eta\nu}. \tag{C5}$$

If Eve does not want to disturb the detection statistics on Bob's side, she should choose Z_i^μ and Z_i^ν smartly, so that Eqs. (C4) and (C5) are satisfied.

In the decoy-state postprocessing [22], the secure key is only derived from the single-photon component. Then, the upper bound of the key rate is given by [40]

$$R^u = Y_1^s e^{-\mu}\mu = q_\mu\left[\xi_\mu Z_1^\mu + (1-\xi_\mu)Z_1^\nu\right]e^{-\mu}\mu. \tag{C6}$$

Now, Eve needs to optimize Z_i^μ and Z_i^ν in order to minimize the upper bound of the key rate, Eq. (C6). The optimization problem can be stated as follows:

$$\min_{\{Z_i^\mu,\, Z_i^\nu\}} Y_1^s \tag{C7}$$

subject to

$$\begin{aligned} Q_\mu &= 1-e^{-\eta\mu} = \sum_{i=1}^{\infty} q_\mu\left[\xi_\mu Z_i^\mu + (1-\xi_\mu)Z_i^\nu\right]e^{-\mu}\frac{\mu^i}{i!}, \\ Q_\nu &= 1-e^{-\eta\nu} = \sum_{i=1}^{\infty} q_\nu\left[\xi_\nu Z_i^\nu + (1-\xi_\nu)Z_i^\mu\right]e^{-\nu}\frac{\nu^i}{i!}, \end{aligned} \tag{C8}$$

where all Z_i^μ and Z_i^ν are in the regime $[0,1]$. In the detection statistics equations, μ and ν are given by Alice's intensity choice, and q_μ, q_ν are determined both by the overlap between the signal and decoy states and by Eve's USD measurement efficiency. For a given overall efficiency η between Alice and Bob, we can calculate the key rate upper bound.

APPENDIX D: EXPERIMENTAL DETAILS

In our experimental demonstration, the laser source is produced by a DFBdiode with a central wavelength of 1550.12 nm and a pulse duration of 1 ns, operating at a repetition rate of 4 MHz. Alice sets $\mu = 0.5$ and $\nu = 0.1$ and randomly modulates the signal and decoy states with uniform probabilities. Eve performs the USD measurement for the signal and decoy states randomly with equal probabilities as well.

The experiment results are collected over an operation time of 748 s. The fluctuation of the attack performance through time is shown in Fig. 4, where we can see that the results are very stable.

In our experiment, the phases of signal pulses and decoy pulses are both 0 because the phase reference of every pulse is passed to Eve. In some other systems, the relative phase between signal and decoy pulses may not be 0 (i.e., $\theta_s - \theta_d \neq 0$) and it can be shown that the success probability of USD is maximized when the relative phase is π and minimized when

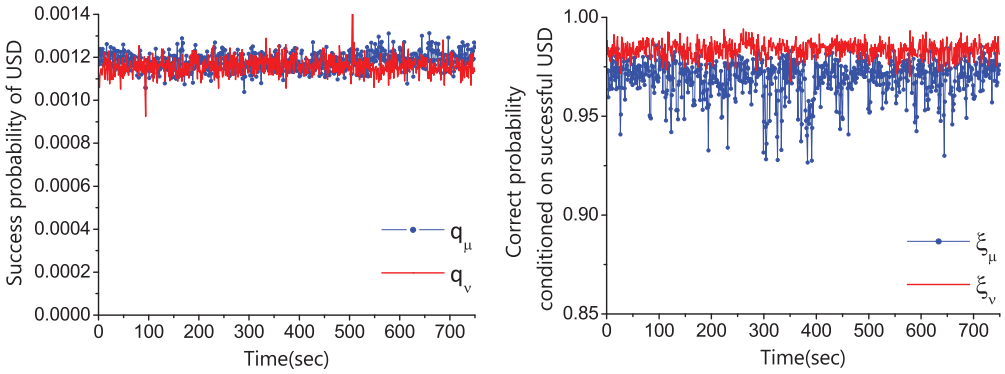

FIG. 4. (Color online) Experimental results over time for the USD attack demonstration. Note that ξ_μ (ξ_ν) is close to 100%, which indicates that we have a nearly perfect interferometer with a high visibility achieving 500:1.

it is 0, corresponding to 23.0% and 3.75%, respectively. With a higher success probability, Eve is more capable of stealing the final key and thus the upper bound of the key rate becomes lower. Figure 3 shows the two upper bounds corresponding to the two success probabilities.

APPENDIX E: REVIEW OF ONE-DECOY STATE

The key assumption in the security proof of decoy-state QKD [15] is the equivalence between phase-randomized coherent states and the photon number channel model. A WCS can be described as a superposition of photon number (Fock) states

$$|\alpha\rangle = |\sqrt{\mu}e^{i\theta}\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle, \tag{E1}$$

where μ and θ are the intensity and phase of the coherent state, respectively. Since the eavesdropper has no knowledge of phase θ, from her point of view, the density matrix of the state should be written as [15]

$$\rho_\mu = \int_0^{2\pi}\frac{d\theta}{2\pi}|\sqrt{\mu}e^{i\theta}\rangle\langle\sqrt{\mu}e^{i\theta}| = e^{-\mu}\sum_{n=0}^{\infty}\frac{\mu^n}{n!}|n\rangle\langle n|. \tag{E2}$$

As shown, the state is a Poisson distributed mixture of photon number state $|n\rangle$. Then, the channel between Alice and Bob can be understood as a photon number channel. Alice uses channel n with a probability of $e^{-\mu}\frac{\mu^n}{n!}$ to send out an n-photon state to carry the qubit information.

Based on the photon number channel model, we briefly review the postprocessing for the one-decoy state protocol [22]. The lower bound of the key rate when Alice and Bob are unaware of Eve's attack is given by

$$R^l = -Q_\mu H(E_\mu) + Y_1\mu e^{-\mu}[1 - H(e_1)], \tag{E3}$$

where Q_μ and E_μ are the overall gain and QBER, and Y_1 and e_1 are the yield and the error rate of the single-photon state, which are estimated by the decoy states. From the analysis using the one-decoy state protocol [22], one can derive the lower bound of Y_1 and upper bound of e_1:

$$\begin{aligned} Y_1 &\geqslant \frac{\mu}{\mu\nu - \nu^2}\left(Q_\nu e^\nu - Q_\mu e^\mu\frac{\nu^2}{\mu^2} - E_\mu Q_\mu e^\mu\frac{\mu^2 - \nu^2}{e_0\mu^2}\right),\\ e_1 &\leqslant \frac{E_\mu Q_\mu e^\mu}{Y_1^{L,\mu,0}\mu}. \end{aligned} \tag{E4}$$

The gains, Q_μ and Q_ν, are given in Eq. (C5). We model the overall QBER in the normal quantum channel to be

$$E_\mu Q_\mu = e_0 Y_0 + e_d(1 - e^{-\eta\mu}), \tag{E5}$$

where $e_0 = 1/2$ is the error rate of the background count; Y_0 is the background count rate, which includes the detector dark count and other background contributions; and e_d is the probability that a photon triggers the incorrect detector and is due to the misalignment and instability of the optical system. As we did not implement Bob's system in our experiment, we adopt some realistic parameters of a setup with superconducting single-photon detectors [31]: $Y_0 = 10^{-7}$, $e_d = 2.0\%$, and a detection efficiency of 5%.

APPENDIX F: ERROR STATISTICS ANALYSIS

In this attack, we form the attack strategy as an optimization problem subject to preserving the gain statistics without maintaining the error statistics, since the attack induces only a low error rate and is not noticeable by looking at the error rate statistics. Here we analyze in detail the more rigorous results of the key rate upper bound, when Eve strictly maintains the gain statistics and the error statistics simultaneously. We can see that even considering the error rate introduced by our attack demonstration that Alice and Bob might check strictly, Eve can still successfully steal the secure key in the same channel loss regime.

To maintain the error statistics, Eve should satisfy the equations

$$\begin{aligned} E_\mu Q_\mu &= \frac{1}{2}Y_0 + e_d(1 - e^{-\eta\mu})\\ &= \frac{1}{2}Z_0^\mu + \sum_{i=1}^{\infty}q_\mu\left[\epsilon_i^\mu\xi_\mu Z_i^\mu + \frac{1}{2}(1 - \xi_\mu)Z_i^\nu\right]e^{-\mu}\frac{\mu^i}{i!}, \end{aligned}$$

$$E_\nu Q_\nu = \frac{1}{2}Y_0 + e_d(1-e^{-\eta\nu}) = \frac{1}{2}Z_0^\nu + \sum_{i=1}^{\infty} q_\nu \left[\epsilon_i^\nu \xi_\nu Z_i^\nu + \frac{1}{2}(1-\xi_\nu)Z_i^\mu\right] e^{-\nu}\frac{\nu^i}{i!}, \tag{F1}$$

where $\epsilon_i^{\mu(\nu)}$ is the error Eve sets when Alice sends a signal (decoy) state and Eve gets the correct USD measurement result, and $Z_0^{\mu(\nu)}$ is the dark count Eve sets when there is no photon in the signal (decoy) state Alice sends. Note that Eve simply sets the QBER error to be the upper bound $\frac{1}{2}$ here, when Eve gets the incorrect USD measurement results.

Similarly to Appendix C, the optimization problem of minimizing the key rate upper bound can be stated as follows:

$$\min_{\{Z_i^\mu, Z_i^\nu\}} Y_1^s \tag{F2}$$

subject to

$$Q_\mu = 1-e^{-\eta\mu} = \sum_{i=1}^{\infty} q_\mu \left[\xi_\mu Z_i^\mu + (1-\xi_\mu)Z_i^\nu\right] e^{-\mu}\frac{\mu^i}{i!},$$

$$Q_\nu = 1-e^{-\eta\nu} = \sum_{i=1}^{\infty} q_\nu \left[\xi_\nu Z_i^\nu + (1-\xi_\nu)Z_i^\mu\right] e^{-\nu}\frac{\nu^i}{i!},$$

$$E_\mu Q_\mu = \frac{1}{2}Y_0 + e_d(1-e^{-\eta\mu}) \geqslant \sum_{i=1}^{\infty}\frac{1}{2}q_\mu(1-\xi_\mu)Z_i^\nu e^{-\mu}\frac{\mu^i}{i!},$$

$$E_\nu Q_\nu = \frac{1}{2}Y_0 + e_d(1-e^{-\eta\nu}) \geqslant \sum_{i=1}^{\infty}\frac{1}{2}q_\nu(1-\xi_\nu)Z_i^\mu e^{-\nu}\frac{\nu^i}{i!}, \tag{F3}$$

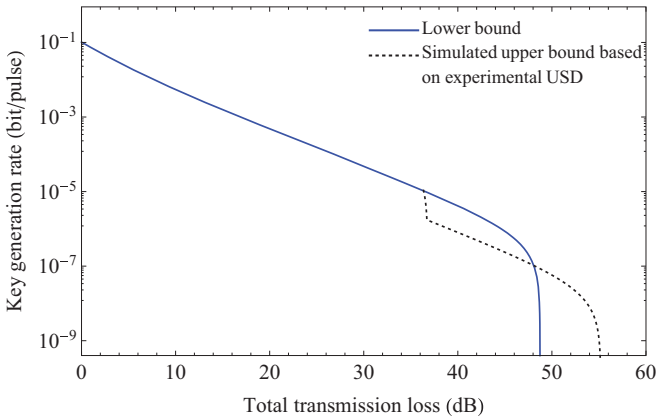

FIG. 5. (Color online) Bounds on the key generation rate. Our attack is successful when the lower bound is higher than the upper bound, which occurs when the transmission loss is between 36.3 and 48.1 dB (for our experiment).

where all Z_i^μ and Z_i^ν are in the regime $[0,1]$. In the gain and error statistics equations, ξ_μ, ξ_ν and q_μ, q_ν are set as the experimental results. For a given overall efficiency η between Alice and Bob, we can calculate the key rate upper bound, as shown in Fig. 5.

Theoretically, since the error rate of USD measurement is 0, the error statistics is easy to maintain or even optimize. Therefore, the gap between the lower bound in Eq. (E3) and the upper bound in Eq. (C6) will be the same as the one shown in Fig. 3.

[1] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing* (IEEE Press, New York, 1984), pp. 175–179.
[2] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[3] C.-H. F. Fung, B. Qi, K. Tamaki, and H.-K. Lo, Phys. Rev. A **75**, 032314 (2007).
[4] F. Xu, B. Qi, and H.-K. Lo, New J. Phys. **12**, 113026 (2010).
[5] V. Makarov, A. Anisimov, and J. Skaar, Phys. Rev. A **74**, 022313 (2006).
[6] V. Makarov and J. Skaar, Quantum Inf. Comput. **8**, 0622 (2008).
[7] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, Quantum Inf. Comput. **7**, 073 (2007).
[8] Y. Zhao, C.-H. F. Fung, B. Qi, C. Chen, and H.-K. Lo, Phys. Rev. A **78**, 042333 (2008).
[9] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nature Photonics **4**, 686 (2010).
[10] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Nature Commun. **2**, 349 (2011).
[11] D. Mayers and A. Yao, in *FOCS, 39th Annual Symposium on Foundations of Computer Science* (IEEE, Computer Society Press, Los Alamitos, CA, 1998), p. 503.
[12] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).
[13] M. Lucamarini, G. Vallone, I. Gianani, P. Mataloni, and G. Di Giuseppe, Phys. Rev. A **86**, 032325 (2012).
[14] H.-K. Lo, M. Curty, and B. Qi, Phys. Rev. Lett. **108**, 130503 (2012).
[15] H.-K. Lo, X. Ma, and K. Chen, Phys. Rev. Lett. **94**, 230504 (2005).
[16] H.-K. Lo and J. Preskill, Quantum Inf. Comput. **7**, 0431 (2007).
[17] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, Quantum Inf. Comput. **4**, 325 (2004).
[18] M. Dušek, M. Jahma, and N. Lütkenhaus, Phys. Rev. A **62**, 022306 (2000).
[19] W.-Y. Hwang, Phys. Rev. Lett. **91**, 057901 (2003).
[20] X.-B. Wang, Phys. Rev. Lett. **94**, 230503 (2005).
[21] G. Brassard, N. Lütkenhaus, T. Mor, and B. C. Sanders, Phys. Rev. Lett. **85**, 1330 (2000).
[22] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, Phys. Rev. A **72**, 012326 (2005).
[23] H.-K. Lo, H. F. Chau, and M. Ardehali, J. Cryptol. **18**, 133 (2005).

[24] N. Imoto, H. A. Haus, and Y. Yamamoto, Phys. Rev. A **32**, 2287 (1985).
[25] M. Brune, S. Haroche, V. Lefevre, J. M. Raimond, and N. Zagury, Phys. Rev. Lett. **65**, 976 (1990).
[26] M. J. Holland, D. F. Walls, and P. Zoller, Phys. Rev. Lett. **67**, 1716 (1991).
[27] P. Grangier, J. Levenson, and J. Poizat, Nature **396**, 537 (1998).
[28] Z. Yuan and A. Shields, Opt. Express **13**, 660 (2005).
[29] T. Chen, H. Liang, Y. Liu, W. Cai, L. Ju, W. Liu, J. Wang, H. Yin, K. Chen, Z. Chen *et al.*, Opt. Express **17**, 6540 (2009).
[30] M. Peev, C. Pacher, R. Alléaume, C. Barreiro, J. Bouda, W. Boxleitner, T. Debuisschert, E. Diamanti, M. Dianati, J. F. Dynes *et al.*, New J. Phys. **11**, 075001 (2009).
[31] K. Yoshino, M. Fujiwara, A. Tanaka, S. Takahashi, Y. Nambu, A. Tomita, S. Miki, T. Yamashita, Z. Wang, M. Sasaki *et al.*, Opt. Lett. **37**, 223 (2012).
[32] R. B. M. Clarke, A. Chefles, S. M. Barnett, and E. Riis, Phys. Rev. A **63**, 040305 (2001).
[33] S. J. van Enk, Phys. Rev. A **66**, 042313 (2002).
[34] C. Wittmann, U. L. Andersen, M. Takeoka, D. Sych, and G. Leuchs, Phys. Rev. Lett. **104**, 100505 (2010).
[35] Y. Zhao, B. Qi, and H. K. Lo, Appl. Phys. Lett. **90**, 044106 (2007).
[36] I. Ivanovic, Phys. Lett. A **123**, 257 (1987).
[37] D. Dieks, Phys. Lett. A **126**, 303 (1988).
[38] A. Peres, Phys. Lett. A **128**, 19 (1988).
[39] S.-H. Sun, M. Gao, M.-S. Jiang, C.-Y. Li, and L.-M. Liang, Phys. Rev. A **85**, 032304 (2012).
[40] X. Ma, C.-H. F. Fung, F. Dupuis, K. Chen, K. Tamaki, and H.-K. Lo, Phys. Rev. A **74**, 032330 (2006).