Physics Letters A 380 (2016) 1589–1592

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

Trojan horse attacks on counterfactual quantum key distribution ☆

Xiuqing Yang [a,b,*], Kejin Wei [c], Haiqiang Ma [c], Shihai Sun [d,**], Yungang Du [b], Lingan Wu [e]

[a] School of Science, Beijing Jiaotong University, Beijing 100044, China
[b] College of Science, Inner Mongolia University of Technology, 010051 Hohhot, China
[c] School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
[d] Department of Physics, National University of Defense Technology, Changsha 410073, China
[e] Laboratory of Optical Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China

ARTICLE INFO

Article history:
Received 21 March 2015
Received in revised form 9 September 2015
Accepted 12 September 2015
Available online 2 November 2015
Communicated by A. Eisfeld

Keywords:
Quantum key distribution
Counterfactual quantum key distribution
Trojan horse attacks

ABSTRACT

There has been much interest in "counterfactual quantum cryptography" (T.-G. Noh, 2009 [10]). It seems that the counterfactual quantum key distribution protocol without any photon carrier through the quantum channel provides practical security advantages. However, we show that it is easy to break counterfactual quantum key distribution systems in practical situations. We introduce the two types of Trojan horse attacks that are available for the two-way protocol and become possible for practical counterfactual systems with our eavesdropping schemes.

1. Introduction

Quantum key distribution (QKD) assumes absolute security guaranteed by the laws of physics. A so-called quantum channel can effectively protect its quantum systems from possibly probing quantum states. Any eavesdropping on the quantum channel can be detected by the legitimate parties. However, a small difference between physical devices and the theoretical model may introduce flaws that could be open to many attacks, of which the photon-number-splitting attack seems to be the most general threat if there is some loss in the quantum channel and if weak coherent pulses are used. Subsequent to the first QKD protocol, proposed by Bennett and Brassard [1] in 1984, there have been further developments in both the theoretical foundation and experimental demonstration. All these QKD protocols requiring actual photon transmission are vulnerable to such an attack in practice. Although the decoy-state method [2–4] can help to generate secure keys, the security of the final keys is still uncertain because of Trojan horse attacks [5,6] in some protocols. In particular, in a deterministic QKD protocol with a two-way quantum channel [7–9], Eve may attack the information carrier traveling over both the Bob–Alice channel and the Alice–Bob channel. Compared with the one-way protocol, the security analysis of the two-way protocol is complicated and its security has been challenged over time.

In 2009 a novel QKD based on the quantum counterfactual effect (denoted the *N09 protocol*) was proposed by Noh [10]. In his study, a conceptually new approach to accomplish the task of a QKD without any photon carrying secret information through the quantum channel was introduced. The N09 protocol as a method different from classical QKD has two peculiar features: (1) the photon carrying secret information could not have been in the transmission channel; (2) Eve cannot access the entire quantum system of a single photon, but can access only part of the quantum system. It seems that the N09 protocol provides practical security advantages by eliminating the possibility that an eavesdropper can directly access any photon contributing to the sifted keys. Since then, Noh's idea has had a significant effect on research in the QKD field [11–15].

However, we show here that it is easy to break counterfactual QKD systems in practical situations with current technology. We introduce the two types of Trojan horse attacks with which Eve can obtain full information from the practical counterfactual QKD. Trojan horse attacks in the proofs of the N09 protocol, such as the delayed-photon attack [16] or the invisible-photon attack [17], share the same process as in the two-way protocol that Eve extracts the secret information from the quantum states traveling forward and backward through the quantum channel just by probing Alice's or Bob's apparatuses. The security proofs of this protocol cannot be followed by the claim that information can be transferred between Alice and Bob without any photon transmission.

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author at: School of Science, Beijing Jiaotong University, Beijing 100044, China.
** Corresponding author.
E-mail addresses: xqqyang@163.com (X. Yang), shsun@nudt.edu.cn (S. Sun).

http://dx.doi.org/10.1016/j.physleta.2015.09.027
0375-9601/

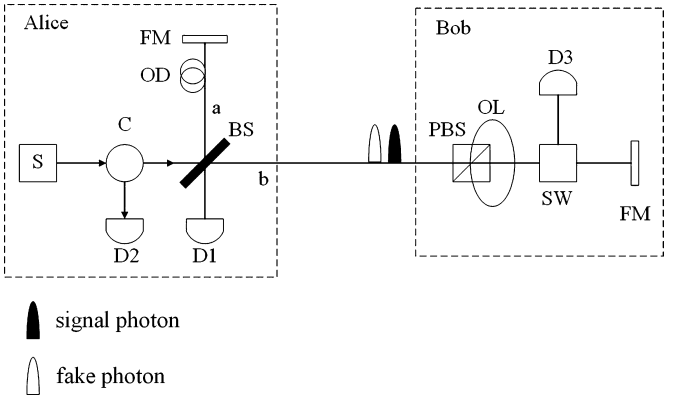

Fig. 1. A counterfactual quantum key distribution system. Alice sends a signal photon split into two pulses by the beam splitter BS. When a split pulse travels along path *b* to Bob, Eve adds a fake photon to the signal pulse with a delay time. For this delayed photon outside the time window Δ_{tw_2}, Bob's detector has practically zero sensitivity. If Eve's and Bob's bit values are different, Eve can capture the fake photon; if the two bit values are equal, Eve cannot separate her photon from the signal photon. The security of the N09 protocol is highly related not only to the error rates in events *D1* and *D3* but also to the probability distribution of events *D1*, *D2*, and *D3*. However, the delayed photon does not contribute to the detection event and is unlikely to be caught. *FM*, Faraday mirror.

2. Security analysis of counterfactual quantum cryptography

We start with a brief description of the N09 protocol. The setup is sketched in Fig. 1. There is a Michelson-type interferometer which is located mostly in Alice's station, but with a portion of one arm in Bob's station. Alice launches a single photon chosen at random either in the horizontally polarized state with bit value 0 or in the vertically polarized state with bit value 1. Bob also randomly chooses one of the two polarizations to block the arm with a detector. If Alice's and Bob's bit values are different, the two split pulses are recombined in the beam splitter and there is constructive interference. If Alice's and Bob's bit values are equal, blocking of the photon pulse in Bob's arm destroys the interference and information about the blocking is obtained without any photon traveling through the quantum channel.

Noh [18] put forward a security advantage that this protocol provides the possibility of hiding the quantum channel in a quantum network, in which several quantum channels separated with spatial, temporal, and spectral filters are available. Alice and Bob randomly choose one of them for communication without the choice being revealed to Eve. If Eve cannot determine the correct quantum channel, she may easily be detected by means of auxiliary detectors monitoring any light through the decoy channels. However, when weak coherent pulses are used, the quantum channel identification problem can also be resolved easily in the N09 protocol. In a passive beam-splitting attack [10], Eve may insert a beam splitter in every quantum channel and identify the correct quantum channel by splitting a photon. Although Eve cannot obtain a copy of the initial quantum state when she succeeds in splitting a photon, she can use Trojan horse attacks against the quantum channel to obtain information without inducing any error.

For instance, a so-called counterfactual attack [19,20] probes Bob's polarization from practical counterfactual systems. However, this attack does not break the security proofs of this protocol when ideal QKD systems with infinite resources are used. In other words, a counterfactual eavesdropping scheme works on practical systems generating only a finite number of keys. Alice and Bob can accurately detect Eve's attack under certain circumstances, provided that the detectors have perfect detection efficiency and dark counter rate. Therefore this attack is available to only imperfect QKD systems in a finite-key scenario.

3. Eavesdropping

The two-way protocol has a feature that Eve has access to a single photon traveling over both the Bob–Alice channel and the Alice–Bob channel. Although in the present protocol no photon actually travels forward and backward over the quantum channel, and there is a possibility for a photon to travel forward and backward, this does not suggest that Eve could never perform the most powerful attacks on the quantum channel. We consider two different eavesdropping strategies, and we find the counterfactual QKD implementation rather than those of two-way systems could not benefit from a higher level of security against Trojan horse attacks.

3.1. Delayed-photon attack

One eavesdropping scheme is the delayed-photon attack (see Fig. 1). Each time Alice sends a pulse to Bob, Eve adds a fake photon either in the horizontally polarized state or in the vertically polarized state with a delay time, outside Bob's detection time window. All events outside the time window are discarded. In this way, Eve's attack is unlikely to be detected because this fake photon is not registered by Bob's detector. After the operation done by Bob, if Eve's and Bob's bit values differ, the fake photon will be reflected by the Faraday mirror and Eve will get a photon back. When Eve receives her photon from Bob, she can obtain information about Bob's operation without making a measurement. Additionally, if Eve's and Bob's bit values are equal, the fake photon is blocked and Eve cannot separate her photon from the signal photon. In both cases, Eve can obtain full information by monitoring the fake photon sent to Bob's site.

In the following, we explain Eve's quantum operation by a delayed photon in more detail. Eve prepares a photon randomly in one of the two states, horizontal polarization and vertical polarization, and sends it to Bob with a delay time. Bob randomly switches the polarization selection through accurate control of the switch timing. If the polarization of the delayed photon is identical to his polarization, Bob can switch the polarization state to detector *D3*. If the delayed photon has a polarization orthogonal to Bob's polarization, it may be reflected by the Faraday mirror and return. If Eve gets a photon back, she can infer that the polarization of her original photon is orthogonal to Bob's polarization. On the other hand, in two cases Eve cannot receive the photon from Bob. One is the fake photon is blocked by Bob. However, this fake photon is not registered by detector *D3* because it is outside Bob's detection time window. The other is the channel loss. This channel loss problem may be overcome if a fake photon with higher light intensity is used. After the detection of a photon has been completed, Alice and Bob publicly announce each event *D1*, *D2*, or *D3*. Only event *D1* is used to generate a key. When event *D1* is announced, Eve extracts the secret information from the fake photon operated without being detected.

Instead of using two orthogonal polarization states, Eve can eavesdrop on the communication using four polarization states as in the protocol of Bennett and Brassard [1] or two nonorthogonal states as in the protocol of Bennett [21]. Eve prepares a fake photon in the state $|+x\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. When Eve and Bob use the same basis, Eve can obtain Bob's encoding information as in the original protocol; with a different basis, Eve obtains Bob's information only if she receives her photon operated. Then Eve performs a polarization measurement on the fake photon using the other measurement basis. If the state of the photon is transformed into horizontal polarization, then Bob's bit value is 0; if the state is vertical polarization, then the bit value is 1. In half of the cases Eve cannot receive the photon from Bob, and she cannot learn Bob's information. Finally, this reduces the eavesdropper's information by 25%, at the expense of the key rate.