Trojan Horse Attack Strategy on Quantum Private Communication

Jinye Peng[1], Guangqiang He[2], Jin Xiong[2], and Guihua Zeng[2]

[1] School of Information Science and Technology, Northwest University, Xi'an, 710069, China
pjy@nwu.edu.cn
[2] Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai 200030, China
{gqhe, jinxiongt, ghzeng}@sjtu.edu.cn

Abstract. Fragility of quantum private communication based on Einstein-Podosky-Rosen (EPR) pair as pre-shared key against trojan horse attack strategy is investigated in detail. To prevent this kind of attack strategy, the EPR pairs employed in the quantum private communication is transferred into non-orthogonal entangled states by employing unitary transformations which are actually rotation operations on the quantum signal. Analysis show that the improved scheme is robust against the trojan horse attack strategy without reducing the security against other kinds of attack strategies.

1 Introduction

In private communication and data security system attackers always try to break the employed confidential system for their benefits. To protect effectively the legitimate information, cryptography has been employed widely to prevent these attack. However, as virtue rises one foot, vice rises ten. To break the private communication and data security system provided by employing cryptographic approach, a concomitant subject called as cryptoanalysis has also been arisen [1]. The so called cryptoanalysis is a science and study of methods of breaking ciphers. Many attack strategies for converting encrypted messages into plaintext without initial knowledge of the key employed in the encryption have been investigated and used in practice. But, success of these strategies completely depends on the drawbacks of the cryptographic system, i.e., cryptosystem. These drawbacks come from two major aspects, i.e., the inappropriate fundamentals, which is employed as a foundation for the scheme, and the imperfection of the cryptosystem's construction. Actually, any improper design will create drawbacks in the cryptosystem, subsequently the attacker may break in principle the scheme by means of these drawbacks.

Trojan horse attack strategy (THAS) may be generated from the drawback of construction of the system (e.g., device, computer program, algorithm or protocol *et al.*). When a trojan horse is hidden without easy detection in a system, attacker can break the system and obtain useful information by employing trojan

K. Chen et al. (Eds.): ISPEC 2006, LNCS 3903, pp. 177–186, 2006.

horses. Unfortunately, this strategy is not only available in classic cryptography but also in the recently proposed quantum cryptography [2, 3, 4, 5, 6]. For the attack of this strategy on the quantum key distribution has been analyzed in [7, 8], and a scheme for preventing this strategy was proposed in [9]. In this paper we study the THAS on the quantum private communication, which employs EPR pair(s) as the symmetrical key [10, 11]. Three aspects will be investigated in this work, including the mechanism, the attack way on the quantum private communication system, and the preventing approach for this attack strategy. Especially an improvement scheme will be investigated in details.

This paper is arranged as follows. In Sec.2, mechanism of the THAS is analyzed firstly. Then in Sec.3 we investigate the fragility of quantum private communication based on EPR pair(s) against the trojan horse attack. An improvement scheme for preventing the THAS is presented in Sec.4. After these a simple remark will be presented in Sec.5. Finally a conclusion is drawn in Sec.6.

2 Mechanism for Trojan Horse Attack Strategy

Let us firstly investigate mechanism of the THAS in cryptography (including classic cryptography and quantum cryptography. Here 'classic' refers to 'quantum'.). In essential, all attack approaches proposed in cryptoanalysis can be categorized mainly as three kinds of attack strategies, i.e., the strategy based on fundamentals drawbacks (SFD), the strategy based on obtained information (SOI), and the strategy based on assistant systems (SAS). In the SFD attacker makes use of fundamentals drawbacks to break the cipher and obtain useful information. As an example, the classic cryptosystem is based on the complexity assumption which has not been proven, thus a fundamentals drawback is usually contained. With the development of mathematics these drawbacks become a means for breaking the cryptosystem [1]. Another example is the attack approaches presented in quantum cryptography, by far most attack strategies such as the individual and collective attacks [8] are based on the fundamentals, i.e., quantum laws. Fortunately all proofs are advantaged to the quantum cryptography but not to the cryptoanalysis. In the SOI attacker makes use of leaked information of the cryptosystem, the ciphertext, and/or the obtained parts of plaintext to break the cryptosystem [1]. We would like to stress here the SAS, which relies on an assistant systems to break the cryptosystem. One of typical approaches in this situation is the THAS.

To study mechanism of the THAS, let us firstly consider what is trojan horse in the field of information protection, since the trojan horse is an important ingredient in the THAS. In data security the trojan horse is defined as a small program inserted by an attacker in a computer system. It performs functions not described in the program specifications, taking advantage of rights belonging to the calling environment to copy, misuse or destroy data not relevant to its stated purpose. For example, a trojan horse in a text editor might copy confidential information in a file being edited to a file accessible to another. More generally, the so-called trojan horse is a 'robot horse' which can become a part of the legitimate users' systems. Then the 'robot horse' can be surreptitiously

exploited the legitimate authorizations of operation (e.g., measurement, detection *et al.*) to the detriment of security. For example, break the system via feeding back information to the attacker (e.g., the dishonest manufacturer or even the adversary) or directly destroying the legitimate data. To the legitimate users' system the trojan horse is actually an additional system with passive effects. Many things, such as devices and small programs inserted in the users' system, probing signals entering users' system through a public channel *et al*, or even the attacker, can become trojan horse. However, we must emphasize that it is impossible for any trojan horse to play the same role as legitimate users since the trojan horse is only a small part of the legitimate system.

There are mainly two kinds of trojan horses, i.e., pre-lurked trojan horse and online trojan horse. The pre-lurked trojan horse is a 'robot horse' which is pre-inserted in legitimate users' system, such as programs, apparatuses or even offices. At an appropriate condition the lurked trojan horse is activated automatically by the legitimate system, and then it feeds back available information to attackers even destroy the users' system. The online trojan horse is actually a probing signal which may enter the confidential system without awareness of legitimate communicators and then back-reflect to the attacker. Both kinds of trojan horses may be classic as well as quantum. In addition, the trojan horse may also be a combination of the 'quantum horse' and 'classic horse' in quantum private communication.

If a trojan horse can be inserted successfully in users' system, the attacker may break the employed cryptosystem and obtain available information by means of the feedback information of the 'robot horse'. This attack strategy is called as THAS. Corresponding to various kinds of the trojan horses there are two kinds THASs, i.e., the strategy relied on a pre-lurked trojan horse and the strategy depended on the probing signal. While the attack ways may be classic approaches or quantum approaches determined by the features of the employed trojan horses. For example, if employing a pointer state of the legitimate system as a trojan horse, or a pre-inserted tiny device as a trojan hose, which is exploited to detect the quantum state of the quantum bits as the key, the attacker can obtain useful messages by analyzing the feedback information of the trojan horse. Consider the case of sending light pulses (probing signal) into the fiber and entering legitimate users's apparatuses, then the attacker may obtain useful information by analyzing the backreflected light [8]. Obviously, the THAS can do nothing without the trojan horse, since the feedback information of the trojan horse is very important in this kind of attack strategy. Obviously this strategy is different from the strategies which always involved in the quantum cryptography, e.g., the intercept/resend attack and the entanglement attack [3, 5, 6], where the attacker can directly obtain the information for attack.

3 Fragility of Quantum Private Communication Based on EPR-Pair Against Trojan Horse Attack

It has been shown that the quantum private communication may be implemented by exploiting a quantum cryptographic key algorithm with EPR pair(s), i.e.,

entangled quantum optical signal, as the pre-shared key. This kind of algorithms is provably secure for the SFD and the SOI. However, they can not circumvent the THAS, which employs pre-lurked trojan horse (in the following we suppose the trojan horse is a tiny device pre-inserted in Alice's or/and Bob's apparatus). To show fragility of the quantum cryptographic algorithm employing EPR pair(s) as key against the THAS, we first give a simple description for this kind of algorithm. Suppose communicators Alice and Bob share n EPR pairs as the key $\mathcal{K} = \{|k_1\rangle, |k_2\rangle, \cdots, |k_n\rangle\}$. Each EPR pair can be expressed as,

$$|k_i\rangle = \frac{1}{\sqrt{2}}\left(|0_a^i 0_b^i\rangle + |1_a^i 1_b^i\rangle\right) = |\Phi_i^+\rangle, \tag{1}$$

where subscripts a, b denote Alice's particle $\mathcal{P}_a$ and Bob's particle $\mathcal{P}_b$ of each EPR pair, $|k_i\rangle$ denotes the i^{th} element in the key $\mathcal{K}$, and $i = 1, 2, \cdots, n$. Denote the plaintext (message) by,

$$|\psi^m\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{2}$$

the corresponding set of particles is expressed by $\mathcal{P}_m$, where $|\alpha|^2 + |\beta|^2 = 1$. Suppose Alice is the sender, then Alice encrypts the qubit $|\psi^m\rangle$ by making use of the quantum controlled-NOT operations on each EPR particle $\mathcal{P}_a$ (key particle) and the message particle $\mathcal{P}_m$. After that, Alice obtains the ciphertext $|\Psi^c\rangle$, which can be denoted as,

$$|\Psi^c\rangle = \mathcal{C}_{mk}^n |k_n\rangle\{\mathcal{C}_{mk}^{n-1}|k_{n-1}\rangle\{\cdots\{\mathcal{C}_{mk}^1|k_1\rangle|\psi^m\rangle\}\}\}, \tag{3}$$

where $\mathcal{C}_{mk}^i$ denotes the i^{th} quantum controlled-NOT gate on $\mathcal{P}_m$ and $\mathcal{P}_a$, the subscript mk denotes the quantum gate operating on the key particle and the message particle, and the controlled-NOT gate $\mathcal{C}$ is defined as,

$$\mathcal{C}|\epsilon_1\rangle|\epsilon_2\rangle = |\epsilon_1\rangle|\epsilon_1 \oplus \epsilon_2\rangle, \quad \epsilon_{1,2} \in \{0, 1\}, \tag{4}$$

in matrix form $\mathcal{C}$ can be denoted as,

$$\mathcal{C} = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & \sigma_x \end{bmatrix}. \tag{5}$$

After encrypting all plaintext elements Alice sends the ciphertext $|\Psi^c\rangle$ to Bob via a quantum channel. Then Bob decrypts the ciphertext by making use of an inverse process controlled under the key, i.e., the $\mathcal{C}_{mb}^{-1}$ on Bob's particles set $\mathcal{P}_b$ and the received particles set $\mathcal{P}_m$. Finally Bob get the message.

Now let us investigate the THAS on the above quantum algorithm. First, we consider the situation of using only one EPR pair as the key. In this case, the key is just the EPR pair, i.e, $|K\rangle = |\Phi^+\rangle$, which can be denoted as,

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|0_a 0_b\rangle + |1_a 1_b\rangle\right). \tag{6}$$

Then ciphertext can be expressed as,

$$|\Psi^c\rangle = C_{ma}|\Phi^+\rangle|\psi^m\rangle = |0_a 0_b\rangle \otimes |\psi^m\rangle + |1_a 1_b\rangle \otimes X_m|\psi^m\rangle. \tag{7}$$

where X_m denotes the quantum X-gate on the particle $\mathcal{P}_m$. Eq. (7) illustrates that when Alice's and Bob's EPR particles are in the states $|0_a 0_b\rangle$ then the message particle is in the state $|\psi^m\rangle$, otherwise, the state of the message particle is in $X_m|\psi^m\rangle$.

Obviously, if Alice's and Bob's EPR particles can not be disturbed by the attacker, above algorithm is secure. However, if the attacker can pre-lurks a trojan horse in Alice's or Bob's apparatus, the legitimate communicators Alice and Bob will be no luck since the attacker can obtain their useful information through the THAS. This can be done very easily. Suppose the attacker puts successfully a trojan hose, Υ, e.g., a set of tiny devices which can distinguish the eigenstates states $|0\rangle$ and $|1\rangle$ (for example a device can recognize the 'bright' and 'dark' pulse) and send feedback information, in Alice's apparatus (this is available since in practice the users are not experts so that they can not easily find the 'robot horse' which is pre-lurked ulteriorly by the dishonest manufacturers), then the key can be written as $|\Phi^+(\Upsilon)\rangle$. Subsequently Alice's encrypting transformation by making use of controlled-NOT yields a ciphertext state, which can be written as,

$$|\Psi_h^c\rangle = |0_a(h_{\shortparallel})0_b\rangle \otimes |\psi^m\rangle + |1_a(h_{\perp})1_b\rangle \otimes X_m|\psi^m\rangle, \tag{8}$$

where $h_{\shortparallel}$ and $h_{\perp}$ are the feedback information of the trojan horse. After Alice has encrypted her message $|\psi^m\rangle$ using the EPR pair, the trojan horse is activated automatically. As an example, if the attacker pre-lurks a measurement bases for the eigenstates states $|0\rangle$ and $|1\rangle$, the trojan horse only needs to distinguish Alice's EPR particle. Now the 'horse' feeds back the result $h_{\shortparallel}$ when the recognized result is $|0\rangle$, otherwise the 'horse' feeds back the result $h_{\perp}$. Then, what the attacker needs to do is to wait Alice's ciphertext $|\Psi^c\rangle$ and the feedback information of the trojan horse. If the attacker can successfully intercept the ciphertext particle $\mathcal{P}_m$ which is sent to Bob, then the attacker can obtain completely the qubit $|\psi_m\rangle$ by making use of the feedback information $h_{\shortparallel}$ and $h_{\perp}$, and the intercepted particle $\mathcal{P}_m$. For example, if the feedback information shows that Bob's key bit is $|0\rangle$, attacker gets $|\psi^m\rangle$. If the feedback information shows that Bob's key bit is $|1\rangle$, attacker gets $X_m|\psi^m\rangle$. By these knowledge, the attacker can completely obtain the plaintext (message).

In the above we have analyzed the trojan horse attack strategy for the situation which makes use only one EPR pair as a key. For the case of making use of two EPR pairs $|\Phi_1^+\rangle$ and $|\Phi_2^+\rangle$ as key, the trojan horse attack strategy can also be successful. In this case the key can be denoted as,

$$|k_1\rangle = |\Phi_1^+\rangle = \frac{1}{\sqrt{2}}\left(|0_a^1 0_b^1\rangle + |1_a^1 1_b^1\rangle\right), \tag{9}$$

and

$$|k_2\rangle = |\Phi_2^+\rangle = \frac{1}{\sqrt{2}}\left(|0_a^2 0_b^2\rangle + |1_a^2 1_b^2\rangle\right). \tag{10}$$

Suppose the attacker pre-lurks successfully two 'horse' Υ_1 and Υ_2 into Alice's or Bob's devices using the similar ways described in above. After Alice's encryption

using controlled-X and controlled-Z gates on the key particle and message particle, the ciphertext state can be written as,

$$
\begin{aligned}
|\Psi_h^c\rangle &= C_{a_2m}^Z\{(C_{a_1m}^X(|\Phi_1^+(\Upsilon_1)\rangle|\psi^m\rangle))|\Phi_2^+(\Upsilon_2)\rangle\} \\
&= \frac{1}{2}|0_a^1 0_b^1(h_{\shortparallel}^1)\rangle\left\{|0_a^2 0_b^2(h_{\shortparallel}^2)\rangle\otimes|\psi^m\rangle + |1_a^2 1_b^2(h_{\perp}^2)\rangle\otimes Z_m|\psi^m\rangle\right\} \\
&+ \frac{1}{2}|1_a^1 1_b^1(h_{\shortparallel}^1)\rangle\left\{|0_a^2 0_b^2(h_{\shortparallel}^2)\rangle\otimes X_m|\psi^m\rangle + |1_a^2 1_b^2(h_{\perp}^2)\rangle\otimes X_m Z_m|\psi^m\rangle\right\} \quad (11)
\end{aligned}
$$

where the superscripts '1' and '2' refer to the particles in the pairs $|\Phi_1^+\rangle$ and $|\Phi_2^+\rangle$, $h_{\shortparallel}^1$ and $h_{\perp}^1$ are feedback information of the trojan horse Υ_1, $h_{\shortparallel}^2$ and $h_{\perp}^2$ are feedback information of the trojan horse Υ_2. Υ_1 and Υ_2 are associated with Bob's particles. It is clear that the attacker can get the message by a similar way as that of employing one EPR pair as key. Therefore, the quantum cryptographic key algorithms based on the EPR pair(s) as keys are fragile against the THAS, although they are provably secure against other attack strategies.

For demonstrating clearly, the THAS on the quantum private communication employing EPR pair(s) as the symmetrical key can be illustrated in Fig. 1. When Alice makes use of her 'machine' to encrypt the message the trojan horse activates automatically a monitor system for obtaining Alice's available information. The obtained information is transmitted automatically to the attacker via certain ways, e.g., classic channel or quantum channel. After eavesdropping ciphertext from the quantum channel, the attacker can obtain the plaintext by the feedback information and the ciphertext.

In summarization, at the situation of communicators' key particles (e.g., Alice's particle $\mathcal{P}_a$ and/or Bob's particle $\mathcal{P}_b$) being orthogonal states, any quantum cryptographic algorithm which employs directly such kind key is not robust against the THAS. Because in such situation a trojan horse can recognize the possible states of the key particle. For example, while Alice and Bob employ the EPR pair as the key then Alice's or Bob's key particle takes the state $|0\rangle$ or $|1\rangle$. Then a proper trojan horse, e.g., a device which can distinguish the eigenstates $|0\rangle$ and $|1\rangle$, can recognize exactly the state of the key particle as described in above. Thus available feedback information can be obtained by the attacker. It is obvious that the teleportation can not circumvent the THAS. Therefore,

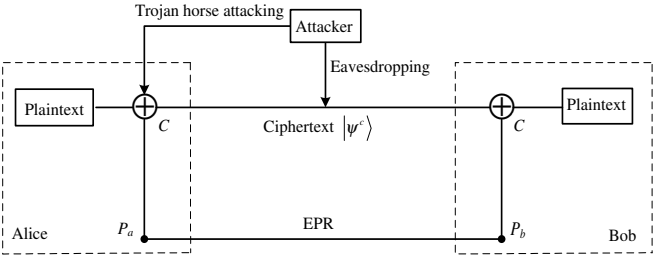

Fig. 1. Diagram of the trojan horse attack strategy on quantum cryptographic algorithm

to prevent the THAS one should use non-orthogonal states as a sharing key or employing operations to change orthogonal state in the quantum private communication based on symmetrical quantum cryptographic key algorithm.

4 Preventing Trojan Horse Attack in the Quantum Private Communication

In this section we show that the THAS can be prevented by transferring the EPR pair(s) into non-orthogonal entanglement state. The process is as follows. The legitimate users Alice and Bob create a set of EPR pairs or pre-share EPR pair(s) as key, each pair can be denoted as,

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0_a 0_b\rangle + |1_a 1_b\rangle) = \frac{1}{\sqrt{2}}(|+_a +_b\rangle + |-_a -_b\rangle), \tag{12}$$

where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$. Before encryption, Alice or Bob randomly choose an operator $\mathcal{U}$ from $\{\mathcal{I}, H\}$ to apply on her (his) EPR particles until all EPR pairs have been operated, where $\mathcal{I}$ and H are respectively the unit operation and the Hadamard gate. This operation yields,

$$|\psi_1\rangle = \mathcal{I}|\Phi^+\rangle = |\Phi^+\rangle, \tag{13}$$

and

$$|\psi_2\rangle = \mathcal{H}|\Phi^+\rangle. \tag{14}$$

Employing bases $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$, $|\psi_2\rangle$ can be expressed as,

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}(|1_a +_b\rangle + |0_a -_b\rangle) = \frac{1}{\sqrt{2}}(|+_a 1_b\rangle + |-_a 0_b\rangle). \tag{15}$$

After these operation, Alice and Bob obtain a random sequence which is consist of $\{|\psi_1\rangle, |\psi_2\rangle\}$. Finally Alice and Bob take this sequence as key. Since $|\langle\psi_1|\psi_2\rangle|^2 \neq 0$ which means that the states $|\psi_1\rangle$ and $|\psi_2\rangle$ are non-orthogonal, any quantum attack strategies can not be available [12]. This property is guaranteed by the well-known no-cloning theorem [13]. In the following we will show this property can also be employed to prevent the trojan horse attack strategy.

First we demonstrate how to encrypt and decrypt the plaintext based on the above transformation. Supposed only Alice operates randomly the particles of n sharing EPR pairs by using of $\mathcal{U}$, then Alice knows results of operation. Of course Bob as well as any attacker don't know Alice's operation results. Define a new Controlled-NOT gate $\mathcal{D}$,

$$\mathcal{D} = \begin{bmatrix} \sigma_x & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix}. \tag{16}$$

where σ_x and I are x-component of Pauli matrix and unit matrix, respectively. One can prove easily that the gate $\mathcal{D}$ is a unitary matrix. The controlled-Not gate $\mathcal{D}$ generates the following transformation,

$$\mathcal{D}|\lambda\rangle|\epsilon\rangle = |\lambda\rangle|\epsilon \oplus \delta_{+,\lambda}\rangle. \tag{17}$$

where $\epsilon \in \{0,1\}, \lambda \in \{+,-\}$ and $\delta_{+,+} = 1, \delta_{+,-} = 0$.

Applying the quantum logic gates $\mathcal{C}$ and $\mathcal{D}$ which is controlled by the operation result of $\mathcal{U}$, Alice encrypts the plaintext $|\psi^m\rangle$. It is noted that the gate $\mathcal{C}$ corresponds to the key element $|\psi_1\rangle$ and $\mathcal{D}$ corresponds to the key element $|\psi_2\rangle$. In other words if the key element is the Bell state, Alice encrypts the plaintext by employing $\mathcal{C}$, otherwise, Alice encrypts the plaintext by employing $\mathcal{D}$. In Bob's side, Bob decrypts the ciphertext always uses the quantum logic gate $\mathcal{C}$ on his particle and the ciphertext particle.

To the attacker, the key $|K\rangle$ is a mix state of $|\psi_1\rangle$ and $|\psi_2\rangle$, i.e.,

$$|K\rangle = \{|\psi_1\rangle, |\psi_2\rangle\}. \tag{18}$$

For convenience, we provide that the probabilities of $|\psi_1\rangle$ and $|\psi_1\rangle$ are the same. Under the control of the key $|K\rangle$ with the quantum logic gates $\mathcal{C}$ and $\mathcal{D}$, the proposed algorithm generates the following ciphertext state, which can be written as,

$$|\Psi_e^c\rangle_{\{|K\rangle=|\psi_1\rangle\}} = \mathcal{C}_{am}|\psi_1\rangle|\psi^m\rangle = \alpha|0_a 0_b\rangle \otimes |\psi^m\rangle + \beta|+_a 1_b\rangle \otimes X_m|\psi^m\rangle, \tag{19}$$

or

$$|\Psi_e^c\rangle_{\{|K\rangle=|\psi_2\rangle\}} = \mathcal{D}_{am}|\psi_2\rangle|\psi^m\rangle = \beta|-_a 0_b\rangle \otimes |\psi^m\rangle + \alpha|1_a 1_b\rangle \otimes X_m|\psi^m\rangle, \tag{20}$$

After the encrypting transformation, Alice obtains the ciphertext

$$|\Psi_e^c\rangle = \{|\Psi_e^c\rangle_{\{|K\rangle=|\psi_1\rangle\}}, |\Psi_e^c\rangle_{\{|K\rangle=|\psi_2\rangle\}}\}. \tag{21}$$

After that, Alice sends the particle $\mathcal{P}_m$ to Bob.

The encryption and decryption processes is shown in Fig. 2. Q can be viewed as a 'quantum switch', which can be implemented by using optical switch in fiber communication on the quantum optical signal. When the key element is $|\psi_1\rangle$, Alice connects '1' with '2', otherwise Alice connects '1' with '3'. It is noted that Alice and Bob pre-share EPR pair in the proposed scheme as the quantum cryptographic algorithm described in section III. The difference between our

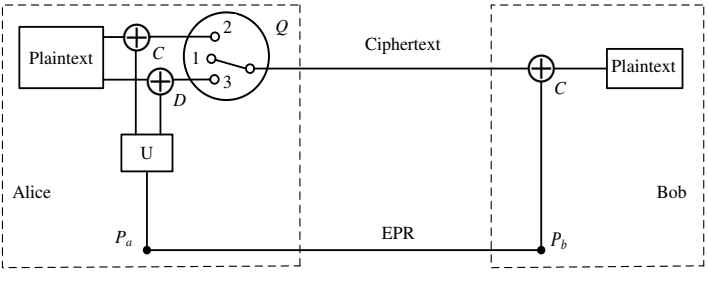

Fig. 2. Diagram of the proposed cryptographic algorithm

algorithm and previous algorithm is the employment of a random operation $\mathcal{U}$ on the EPR pair which is transferred into one of non-orthogonal entangled states $\{|\psi_1\rangle, |\psi_2\rangle\}$.

Now let us show how to prevent the trojan horse attack strategy in the improvement scheme. Suppose the attacker lurks successfully a 'horse', Υ, in Bob's apparatus, then the ciphertext state takes,

$$|\Psi_e^c\rangle(\Upsilon) = \{|\Psi_e^c\rangle(\Upsilon)_{\{|K\rangle=|\psi_1\rangle\}}), |\Psi_e^c\rangle(\Upsilon)_{\{|K\rangle=|\psi_2\rangle\}}\}. \tag{22}$$

where

$$|\Psi_e^c\rangle(\Upsilon)_{\{|K\rangle=|\psi_1\rangle\}} = \alpha|0_a 0_b(h_{\shortparallel})\rangle \otimes |\psi^m\rangle + \beta|_{+a}\, 1_b(h'_?)\rangle \otimes X_m|\psi^m\rangle, \tag{23}$$

and

$$|\Psi_e^c\rangle(\Upsilon)_{\{|K\rangle=|\psi_2\rangle\}} = \beta|_{-a}\, 0_b(h_?)\rangle \otimes |\psi^m\rangle + \alpha|1_a 1_b(h_\perp)\rangle \otimes X_m|\psi^m\rangle, \tag{24}$$

where $h_?$ and $h'_?$ denote the inconclusive feedback information. Although the key is a mix state (see Eq.(18)), Alice and Bob only choose one state from $\{|\psi_1\rangle, |\psi_2\rangle\}$ as the key element in each encrypting operation. Accordingly, if the attacker prelurk one trojan horse, e.g., Υ_1 (for $\{|0\rangle, |1\rangle\}$), in Bob's apparatus, then another states, i.e., $\{|+\rangle, |-\rangle\}$ can not be recognized exactly. If the attacker employs two trojan horses, e.g., 'robot horse' Υ_1 and 'robot horse' Υ_2 (for $\{|+\rangle, |-\rangle\}$), the attacker can also be impossible to get the useful feedback information. Because Alice and Bob's choices for the key is completely random, this leads the impossibility for the trojan horses Υ_1 and Υ_2 to follow completely the changes of the key elements. By other terms, because there are two pairs random bases, i.e., $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$ in Alice's and Bob's apparatuses, it is impossible for the the attacker's 'horse' to recognize these bases. Subsequently, the 'horses' are 'blind' and can not gives correct feedback information. The security is the same as the BB84 protocol [3].

5 Remarks

In above we have analyzed the fragility of quantum private communication based on quantum cryptographic key algorithm against the trojan horse attack strategy, where the EPR pair(s) are employed as a key. However, we would like to stress that the quantum key distribution protocols which are implemented by making use of the EPR pair(s) do not suffer this kind of drawbacks. Since the EPR pair carries initially no information in the quantum key distribution. Especially the users's measurement for obtaining the final key is random. This features leads the trojan horse employed in the above section to be unavailable so that the THAS is impossible in the EPR key distribution protocol [4].

6 Conclusion

In this work, the fragility of the THAS on the quantum private communication based on quantum cryptographic key algorithm implemented by the EPR pairs as

the key has been analyzed in details. It is found that any quantum cryptographic key algorithm exploiting a set of orthogonal states as the symmetrical key can not circumvent the THAS. To prevent this kind attack strategy we proposed a new approach which makes use of the non-orthogonal entangled states. The improvement scheme is robust to the THAS. In addition, the mechanism for the THAS on the quantum cryptography as well as the classic cryptography is also investigated.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (Grant no. 60472018).

References

1. Schneier, B.: *Applied Cryptography:protocols, algorithms, and source code in C* (John Wiley & Sons, Inc., 1994)
2. Wiesner, S., Conjugate coding. Sigact News. **15** (1983), 78-98.
3. Bennett, C. H., and Brassard, G.: Advances in Cryptology: Proceedings of Crypto'84, August 1984, Springer-Verlag, 475 (1984)
4. Ekert, A. K.: Quantum cryptography bases on Bell's theorem. Phys. Rev. Lett. **67** (1991), 661-664.
5. Bennett, C. H., Bessette, F., Brassard,G., Salvail, L. and Smolin, J.: Experimental quantum cryptography. J. Cryptology, **5** (1992), 3-28.
6. Hillery, M., Buzek,V., and Berthiaume,A.: Quantum secret sharing. Phys. Rev. A, **59**(1999), 1829-1834.
7. Lo, H.-K. and Chau,H.F.: Unconditional Security of Quantum Key Distribution over Arbitrarily Long Distances. Science, **283** (1999), 2050-2057.
8. Gisin, N., Ribordy,G., Tittel, W. and Zbinden, H.: Quantum cryptography. Reviews of Modern Physics, **74**(2002), 145-195.
9. Larsson, J.: http://xxx.lanl.gov/quant-ph/0111073, 13 Nov 2001.
10. Y. Zhang, C. Li and G. Guo, Phys. Rev. A **64**, 024302 (2001).
11. Leung, D. W.: Quantun vernam cipher. Quantum information and computation, **2** (2002), 14-30.
12. Karlsson, A., Koashi,M., and Imoto,N.: Quantum entanglement for secret sharing and secret splitting. Phys. Rev. A, **59** (1999), 162-168.
13. Wootters, W. and Zurek, W.: A single quantum cannot be cloned. Nature, **299** (1982), 802-803.