特洛依木马攻击下的量子密码安全性[*]

曾贵华[+]

(上海交通大学 电子工程系,上海 200030)

Security of Quantum Cryptography Against Trojan Horse Attacking

ZENG Gui-Hua[+]

(Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

+ Corresponding author: Phn: +86-21-62933189, E-mail: ghzeng@sjtu.edu.cn

Received 2002-11-19; Accepted 2003-07-02

Zeng GH. Security of quantum cryptography against Trojan horse attacking. *Journal of Software*, 2004,15(8): 1259~1264.
http://www.jos.org.cn/1000-9825/15/1259.htm

Abstract: Trojan horse attacking strategy on quantum cryptography is investigated. First, the fragility of the quantum cryptographic algorithm employing EPR (Einstein-Podosky-Rosen) pairs as a key against the Trojan horse attacking strategy is analyzed. To prevent the Trojan horse attacking, an improved scheme which makes use of the non-orthogonal entangled states is proposed. This scheme is robust to the Trojan horse attacking, without reducing the security on other kinds of attacking strategies.

Key words: Trojan horse attacking strategy; quantum cryptographic algorithm; quantum cryptography; quantum cryptoanalysis

摘 要: 研究了特洛伊木马对量子密码算法的攻击.首先分析了以 EPR 纠缠量子比特为密钥的量子密码算法在特洛伊木马攻击下的脆弱性.在此基础上,基于非正交纠缠量子比特提出了一个改进方案.该方案能有效地防止特洛伊木马的攻击.

关键词: 特洛依木马攻击策略;量子密码算法;量子密码;量子密码分析

中图法分类号: TP309 **文献标识码**: A

1 Introduction

Trojan horse attacking strategy (THAS)[1] has arisen from the drawback of construction of the system e.g. device, computer program, algorithm or protocol *et al*. When a Trojan horse can be hidden without easy detection in a system, attacker can make use of this kind of strategy to break the system and then obtain useful information.

[*] Supported by the National Natural Science Foundation of China under Grant No.60102001 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2002AA145091 (国家高技术研究发展计划(863))

ZENG Gui-Hua was born in 1966. He is a professor and doctorial supervisor at Shanghai Jiaotong University. His current research areas include quantum cryptography, cryptography and information security.

Unfortunately, this strategy is available not only in the classic cryptography but also in the recently proposed quantum cryptography[2~5]. This strategy on the quantum key distribution has been analyzed in Refs.[3,4,6], and a scheme for preventing this strategy was proposed in Ref.[6].

In this paper we consider the THAS on the quantum cryptographic algorithm, which employs EPR pairs as the symmetrical key. Three aspects will be investigated, including the mechanism, the attacking way on the quantum cryptographic algorithm, and the preventing approach for this attacking strategy. Especially the improvement scheme will be investigated in detail.

This paper is arranged as follows. In Section 2, we investigate the THAS on the quantum cryptographic algorithm which employs EPR pairs as the key. An improvement scheme for preventing the THAS is presented in Section 3. After these a simple remark is presented in Section 4. Finally, conclusions are drawn in Section 5.

2 Trojan Horse Attacking Strategy on Quantum Cryptographic Algorithm

Recently, two interesting quantum vernam algorithms based on EPR pairs have been proposed. They employ EPR pairs as the symmetrical keys of the algorithms. In Ref.[7] the message is encrypted by means of a quantum controlled-NOT, with the employment of a symmetrical key which consists of one EPR pair and one bilateral rotation. In Ref.[8] the message is encrypted with a key which consists of two EPR pairs. A common feature of the above quantum vernam algorithms is that EPR pairs are applied as a shared key between the two legitimate users, Alice and Bob. However, the algorithms can not circumvent the THAS. In the following, we investigate the fragility of these algorithms against the THAS which employs pre-lurked Trojan horse (in this section and the following section we suppose the Trojan horse is a tiny device pre-inserted in Alice's or Bob's apparatus).

To show fragility of the quantum cryptographic algorithm employing EPR pair(s) as the key against the THAS, we first give a simple description for this kind of algorithm. In general, this kind of algorithm can be summarized generally as follows. Suppose Alice and Bob sharing n EPR pairs as the key $K=\{|k_1\rangle,|k_2\rangle,...,|k_n\rangle\}$. Each key element is associated with an EPR pair which can be expressed as

$$|k_i\rangle=|\Phi_i^+\rangle=\frac{1}{\sqrt{2}}\left(|0_a^i0_b^i\rangle+|1_a^i1_b^i\rangle\right) \tag{1}$$

where subscripts *a*,*b* denote respectively Alice's particles P_a and Bob's P_b of each EPR pair, $|k_i\rangle$ denotes the i^{th} key element, and i=1,2,…,n. Denote the plaintext (message) by

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

the corresponding particle is expressed as 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 her each EPR particle P_a (key particle) and the message particle P_m. After that, Alice obtains the ciphertext $|\Psi^c\rangle$, which can be described as:

$$|\Psi^c\rangle=C_{mk}^n|k_n\rangle C_{mk}^{n-1}|k_{n-1}\rangle...C_{mk}^1|k_1\rangle|\psi^m\rangle \tag{3}$$

where C_{mk}^i represents the i^{th} quantum controlled-NOT gate on P_m and P_a, the subscript *mk* denotes the quantum gate operating on the message particle and the key particle. Then Alice sends the ciphertext to Bob via a quantum channel. After received the ciphertext $|\Psi^c\rangle$, Bob decrypts the ciphertext by making use of an inverse process controlled by the key. Finally, Bob gets the message.

Now let us investigate 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, $\left|K\right\rangle = \left|\Phi^{+}\right\rangle$, which can be written as:

$$\left|\Phi^{+}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0_a 0_b\right\rangle + \left|1_a 1_b\right\rangle\right) \tag{4}$$

then the ciphertext can be expressed as:

$$\left|\Psi^{c}\right\rangle = C_{mk}\left|\Phi^{+}\right\rangle\left|\psi^{m}\right\rangle = \left|0_a 0_b\right\rangle \otimes \left|\psi^{m}\right\rangle + \left|1_a 1_b\right\rangle \otimes X_m\left|\psi^{m}\right\rangle \tag{5}$$

where X_m is the quantum X-gate on the particle P_m. Eq.(5) illustrates that when Alice's and Bob's EPR particles are in the states $\left|0_a 0_b\right\rangle$, then the message particle is in the state $\left|\psi^{m}\right\rangle$; otherwise, the state of the message particle is in $X_m\left|\psi^{m}\right\rangle$.

Obviously, if Alice's and Bob's EPR particles can not be disturbed by the attacker, the 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 not be lucky 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 $\left|0\right\rangle$ and $\left|1\right\rangle$ (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 ulteriorly pre-lurked by the dishonest manufacturers), then the key can be written as $\left|\Phi^{+}(\Upsilon)\right\rangle$. Subsequently Alice's encrypting transformation by making use of controlled-NOT yields a ciphertext state, which can be written as:

$$\left|\Psi^{c}\right\rangle = \left|0_a(h_{\parallel})0_b\right\rangle \otimes \left|\psi^{m}\right\rangle + \left|1_a(h_{\perp})1_b\right\rangle \otimes X_m\left|\psi^{m}\right\rangle \tag{6}$$

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

In the above we have analyzed the Trojan horse attacking strategy for the situation which makes use of only one EPR pair as a key. For the case of two EPR pairs $\left|\Phi_1^{+}\right\rangle$ and $\left|\Phi_2^{+}\right\rangle$ as the key (see Ref.[8]), the Trojan horse attacking strategy can also be successful. In this case the key can be expressed as:

$$\left|k_1\right\rangle = \left|\Phi_1^{+}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0_a^1 0_b^1\right\rangle + \left|1_a^1 1_b^1\right\rangle\right),\ \left|k_2\right\rangle = \left|\Phi_2^{+}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0_a^2 0_b^2\right\rangle + \left|1_a^2 1_b^2\right\rangle\right) \tag{7}$$

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

$$\begin{aligned}\left|\Psi_h^c\right\rangle &= C_{a_2m}^Z\left\{C_{a_1m}^X\left|\Phi_1^+\left(\Upsilon_1\right)\right\rangle\left|\psi^m\right\rangle\left|\Phi_2^+\left(\Upsilon_2\right)\right\rangle\right\}\\ &=\frac{1}{\sqrt{2}}\left|0_a^1 0_b^1\left(h_{\|}^1\right)\right\rangle\left(\left|0_a^2 0_b^2\left(h_{\|}^2\right)\right\rangle\otimes\left|\psi^m\right\rangle+\left|1_a^2 1_b^2\left(h_{\perp}^2\right)\right\rangle\otimes Z_m\left|\psi^m\right\rangle\right)\\ &\quad+\frac{1}{\sqrt{2}}\left|1_a^1 1_b^1\left(h_{\perp}^1\right)\right\rangle\left(\left|0_a^2 0_b^2\left(h_{\|}^2\right)\right\rangle\otimes X_m\left|\psi^m\right\rangle+\left|1_a^2 1_b^2\left(h_{\perp}^2\right)\right\rangle\otimes X_m Z_m\left|\psi^m\right\rangle\right)\end{aligned} \tag{8}$$

where the superscripts 1 and 2 refer to the particles in the pairs $\left|\Phi_1^+\right\rangle$ and $\left|\Phi_2^+\right\rangle$, $h_{\|}^1$ and $h_{\perp}^1$ are the feedback information of the Trojan horse Υ_1, $h_{\|}^2$ and $h_{\perp}^2$ are the 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 to that of employing one EPR pair as the key. Therefore, the quantum cryptographic algorithms based on the EPR pairs as keys are fragile against the THAS, although they are secure against the other attacking strategies.

Actually, if the possible states of Alice's key particles P_a (or Bob’s key particle P_b) are orthogonal states, any quantum cryptographic algorithm which employs directly such a kind of key is not robust to the THAS. Because, in such a situation, the successful Trojan horse can recognize the possible states of the key particle. For example, while Alice and Bob employ the EPR pair as the key, Alice’s or Bob’s key particle takes the state $\left|0\right\rangle$ or $\left|1\right\rangle$. Then a proper Trojan horse, e.g. a device which can distinguish the eigenstates $\left|0\right\rangle$ or $\left|1\right\rangle$, can recognize exactly the state of the key particle as described above. Thus feedback information available can be obtained by the attacker. Therefore, to prevent the THAS one should use the non-orthogonal states as a shared key in the symmetrical quantum cryptographic algorithm.

3 Prevention of Trojan Horse Attacking Strategy

In this section we will show that the mentioned THAS can be prevented by making use of the non-orthogonal entanglement state as the key. The process is as follows. The legitimate users Alice and Bob create a set of EPR pairs, each pair can be expressed as:

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

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

$$\left|\psi_1\right\rangle=I\left|\Phi^+\right\rangle=\left|\Phi^+\right\rangle \tag{10}$$

and

$$\left|\psi_2\right\rangle=H\left|\Phi^+\right\rangle \tag{11}$$

In bases $\left|0\right\rangle$, $\left|1\right\rangle$, $\left|+\right\rangle$, and $\left|-\right\rangle$, $\left|\psi_2\right\rangle$ can be expressed as:

$$\left|\psi_2\right\rangle=\frac{1}{\sqrt{2}}\left(\left|1_a+_b\right\rangle+\left|0_a-_b\right\rangle\right)=\frac{1}{\sqrt{2}}\left(\left|+_a1_b\right\rangle+\left|-_a0_b\right\rangle\right) \tag{12}$$

After these operations, Alice and Bob obtain a random sequence which consists of $\{|\psi_1\rangle,|\psi_2\rangle\}$. Finally, Alice and Bob take this sequence as the key. Since $\langle\psi_1|\psi_2\rangle\neq 0$, which means that the states $|\psi_1\rangle$ and $|\psi_2\rangle$ are non-orthogonal, any quantum attacking strategies can not be available. This point is guaranteed by the no-cloning theorem. In the following we will show this property can also be employed to prevent the Trojan horse attacking strategy.

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

$$|K\rangle=c_1|\psi_1\rangle+c_2|\psi_2\rangle \tag{13}$$

then the ciphertext state can be written as:

$$\begin{aligned}\left|\Psi_e^c\right\rangle&=C_{mk}|K\rangle\left|\psi^m\right\rangle\\&=\left(\tilde{\alpha}_k^c\left|0_a0_b\right\rangle+\tilde{\beta}_k^c\left|0_a-_b\right\rangle\right)\otimes\left|\psi^m\right\rangle\\&\quad+\left(\tilde{\alpha}_k^c\left|1_a1_b\right\rangle+\tilde{\beta}_k^c\left|1_a+_b\right\rangle\right)\otimes X_m\left|\psi^m\right\rangle\end{aligned} \tag{14}$$

where $\tilde{\alpha}_k^c=c_1/\sqrt{2}$, $\tilde{\beta}_k^c=c_2/\sqrt{2}$. After the encrypting transformation, Alice obtains the ciphertext, i.e. Eq.(14). Then Alice sends the particle P_m to Bob.

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

$$\begin{aligned}\left|\Psi_e^c(\Upsilon)\right\rangle&=C_{am}\left\{c_1\left|\psi_1(\Upsilon)\right\rangle+c_2\left|\psi_2(\Upsilon)\right\rangle\right\}\left|\psi^m\right\rangle\\&=\left(\tilde{\alpha}_k^c\left|0_a0_b\left(h_{\|}\right)\right\rangle+\tilde{\beta}_k^c\left|0_a-_b\left(h_?\right)\right\rangle\right)\otimes\left|\psi^m\right\rangle\\&\quad+\left(\tilde{\alpha}_k^c\left|1_a1_b\left(h_{\perp}\right)\right\rangle+\tilde{\beta}_k^c\left|1_a+_b\left(h_?'\right)\right\rangle\right)\otimes X_m\left|\psi^m\right\rangle\end{aligned} \tag{15}$$

where $h_?$ and $h_?'$ denote the inconclusive feedback information. Although the key is a superposition state (see Eq.(13)), for each encrypting operation Alice and Bob only choose one state from $\{|\psi_1\rangle,|\psi_2\rangle\}$ as the key element. Accordingly, if the attacker pre-lurks one Trojan horse, e.g. Υ_1 (for $\{|0\rangle,|1\rangle\}$), in Bob's apparatus, then another state, 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 is also impossible to get the useful feedback information. Because Alice and Bob's choices for the key are completely random, this leads the impossibility for the Trojan horses Υ_1 and Υ_2 to follow completely the changes of the key elements. In other words, because there are two pairs of the random bases, i.e. $\{|0\rangle,|1\rangle\}$ and $\{|+\rangle,|-\rangle\}$ in Alice's and Bob's apparatuses, it is impossible for the attacker's 'horse' to recognize these bases. Subsequently, the 'horses' are blind and can not give correct feedback information. The security is the same as that of the BB84 protocol[2].

4 Remark

In the above we have analyzed the fragility of the quantum cryptographic algorithm against the Trojan horse strategy, where the EPR pairs are employed as a key. However, we here would like to stress that the quantum key distribution protocols which are implemented by making use of the EPR pairs do not suffer this kind of drawbacks, because in the quantum key distribution the EPR pair initially carries no information, especially the users's measurement for obtaining the final key is completely random. This random feature leads to the Trojan horse employed in the above section to be of no use.

5 Conclusions

In this paper, the fragility of the THAS on the quantum cryptographic algorithm implemented by the EPR pairs as the key has been analyzed in detail. It is found that any quantum cryptographic algorithm exploiting a set of orthogonal states as the symmetrical key can not circumvent the THAS. To prevent this kind of attacking strategy we propose 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. In any THAS, the Trojan horse is very important.

References:

[1] Schneier B. Applied Cryptography: Protocols, Algorithms, and Source Code in C. John Wiley & Sons, Inc., 1994.

[2] Bennett CH, Brassard G. An update on quantum cryptography. In: Advances in Cryptology: Proc. of Crypto'84. Springer-Verlag, 1984. 475~478.

[3] Lo H-K, Chau HF. Unconditional security of quantum key distribution over arbitrary long distances. Science, 1999,283:2050~2056.

[4] Gisin N, Ribordy G, Tittel W, Zbinden H. Quantum cryptography. Review Modern Physics, 2002,74:145~195.

[5] Zeng G, Zhang W. Identity verification in quantum cryptography. Physical Review A, 2000,61:032303/1-5.

[6] Larsson J. A practical Trojan horse for bell-inequality-based quantum cryptography. arXiv: quant-ph/0111073, 13 Nov 2001.

[7] Zhang Y, Li C, Guo G. Quantum key distribution via quantum encryption. Physical Review A, 2001,64:024302/1-4.

[8] Leung DW. Quantum vernam cipher. arXiv: quant-ph/0012077, newest version, 2001.