EXPERIMENTAL ATTACK ON QUANTUM MONEY SCHEME USING MACHINE LEARNING

Jiráková K.[1], Bartkiewicz K.[2, 1], Černoch A.[3], Lemr K.[1]

[1]RCPTM, Joint Laboratory of Optics of Palacký University and Institute of Physics of Academy of Sciences of the Czech Republic, 17. listopadu 12, 771 46 Olomouc, Czech Republic
[2]Faculty of Physics, Adam Mickiewicz University, PL-61-614 Poznań, Poland
[3]Institute of Physics of Academy of Sciences of the Czech Republic, Joint Laboratory of Optics of PU and IP AS CR, 17. listopadu 50A, 772 07 Olomouc, Czech Republic

The concept of quantum money has been originally suggested by S. Wiesner [1]. Its main advantage is that every attempt to copy quantum banknotes leaves the quantum states changed which provides a mark that the money has been counterfeited. According to the no-cloning theorem the quantum states cannot be in general perfectly cloned (copied). However, an imperfect cloning is still possible. Practical implementation of quantum banking is highly dependent on the safety of quantum communication between the bank and a payment terminal. This aspect of security has been already addressed experimentally [2].

We present a version of an eavesdropping attack on the protocol proposed by Bozzio *et al.* [3]. The aim of our research is to demonstrate that cloning implemented even rarely enough that it is indistinguishable from noise is fully sufficient to acquire useful information to counterfeit quantum banknotes. We exploit the fact that completely random encoding of quantum banknotes is computationally impractical [4] and that the bank needs to select a non-random but secret encoding algorithm. Machine learning allows the attacker to predict future banknotes from partial information gained when previous banknotes were cloned.

The scheme of the attack is depicted in the Fig. 1. We consider a credit card containing banknotes encoded using some secret function derived from the public serial number of the banknote. As proposed in [3], quantum banknotes were encoded into sequences of photon pairs choosing one of the 8 encodings for each pair:

$$S = \{|DR\rangle, |DL\rangle, |AR\rangle, |AL\rangle, |RD\rangle, |LD\rangle, |RA\rangle, |LA\rangle\},$$

where $|D\rangle$, $|A\rangle$, $|R\rangle$ and $|L\rangle$ stand for polarisation states on the equator of Poincaré sphere, i.e. diagonal, anti diagonal, right-handed and left-handed polarisation. The vendor's terminal is supposed to perform measurement on these pairs as requested by the bank (randomly choosing D/A and R/L basis for the entire pair). In our case, however, the terminal controlled by the attacker performs quantum phase covariant cloning on a fraction of the pairs and then subjects all copies to the above mentioned measurement. Information gained by this procedure feeds a subsequent machine learning algorithm trying to predict bank's encoding secret.

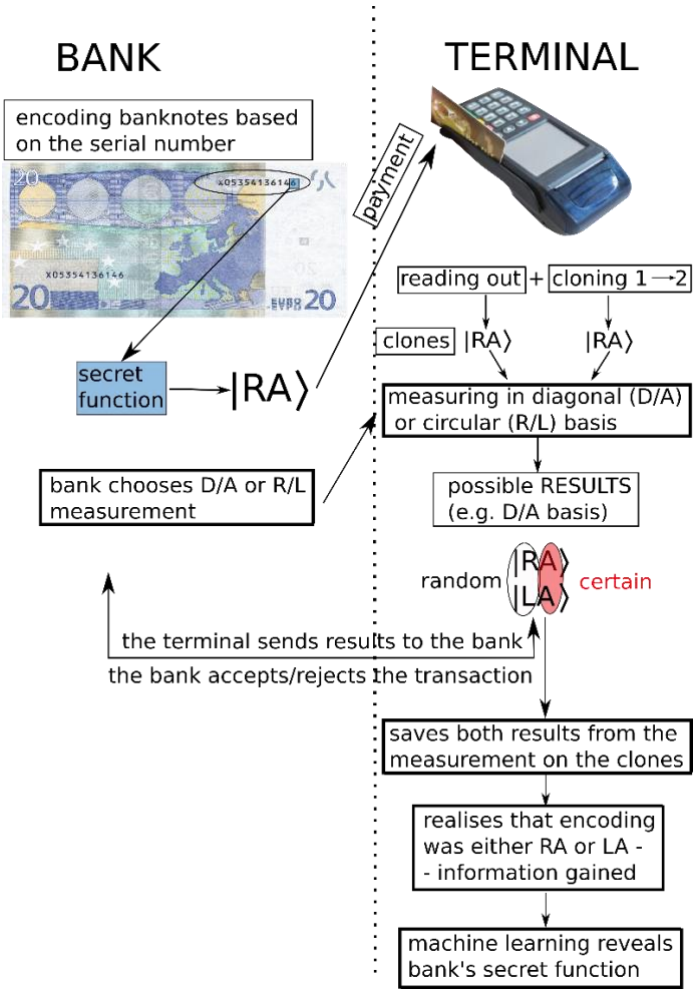

Figure 1: Scheme of the attack. By performing phase covariant cloning the attacker acquires some information about the encoding of the current banknotes. Subsequently this information can be used to counterfeit so far unused banknotes.

Acknowledgement: This research is supported by the Czech Science Foundation under the project No. 17-10003S and by an internal grant of Palacký University IGA-Prf-2018-009.

References: [1] S. Wiesner, Conjugate coding. ACM SIGACT News **15**, 7888 (1983). Original manuscript written circa 1970.
[2] K. Bartkiewicz, A. ernoch, G. Chimczak, K. Lemr, A. Miranowicz and F. Nori, NPJ Quant. Inf. **3**, 7 (2017).
[3] M. Bozzio, A. Orieux, L. Trigo Vidarte, I. Zaquine, I. Kerenidis and E. Diamanti, NPJ Quant. Inf. **4**, (2018).

[4] S. Aaronson, and P. Christiano, arXiv, 1203.4740 (2012).