Quantum hacking with realistic Trojan-horse attacks

Nitin Jain,[1,2,*] Elena Anisimova,[3] Imran Khan,[1,2] Vadim Makarov,[3] Christoph Marquardt,[1,2] and Gerd Leuchs[1,2,4]

[1]*Max Planck Institute for the Science of Light, Erlangen, Germany*
[2]*Institut für Optik, Information und Photonik, University of Erlangen-Nürnberg, Germany*
[3]*Institute for Quantum Computing, University of Waterloo, Ontario, Canada*
[4]*Department of Physics, University of Ottawa, Canada*

Quantum key distribution (QKD) is the most promising application of quantum information technology as of today [1]. In theory, any eavesdropping on the key exchange between two legitimate parties Alice and Bob introduces errors thus disclosing the presence of their adversary Eve. In practice however, the theoretical model may not be properly implemented or may fail to provide a complete description of the implementation. Such theory-practice deviations may arise due to technical imperfections in the hardware or inadvertant leakage of information in side channels and may lead to security loopholes. Indeed, the last decade has witnessed many proof-of-principle "quantum hacking" attacks [2] devised and performed on practical QKD systems that led to improvements of the implementations.

In general, the objective of any eavesdropping strategy is to yield Eve a bitstring that is (at least partially) correlated to the raw key of Alice and Bob. Using privacy amplification, Alice and Bob then obtain a shorter secret key, *ideally* known to Eve only with a negligible probability. However, if the attack enables Eve to learn this secret key with a non-negligible probability and without being discovered, then the security of the QKD system is breached. If q is the quantum bit error rate (QBER) observed by Alice and Bob, γ is the detection rate, and I_E quantifies Eve's correlations with the raw key, the previous statement may be formalized (given Alice and Bob distill a *positive* secret key at the conclusion of a QKD protocol) with the following conditions:

1. the QBER does not cross the hardcoded abort threshold in the QKD system ($q \leq q_{\text{abort}}$),

2. the deviation of the observed detection rate from the expected value, given by $\delta = \left|1 - \frac{\gamma^{\text{obs}}}{\gamma^{\text{exp}}}\right|$, is within tolerable limits, and

3. Eve's actual correlations with the raw key surpass whatever Alice and Bob estimate based on the security proof ($I_E^{\text{act}} > I_E^{\text{est}}$)

In this work, we discuss how Trojan-horse like attacks [3, 4] may satisfy the above conditions. In such attacks, Eve probes the (secret) settings of the QKD system by sending in bright optical pulses from the quantum channel and analyzing the back-reflections [5]. We recently performed a proof-of-principle Trojan-horse attack [6] on Clavis2, the practical QKD system from ID Quantique [7]. The objective of the attack was to read Bob's phase modulator to acquire knowledge of his basis choice – this information suffices for constructing the entire raw key in the Scarani-Acín-Ribordy-Gisin 2004 protocol [8]. We used homodyne detection to discriminate between Bob's phase modulation; see Fig. 1.

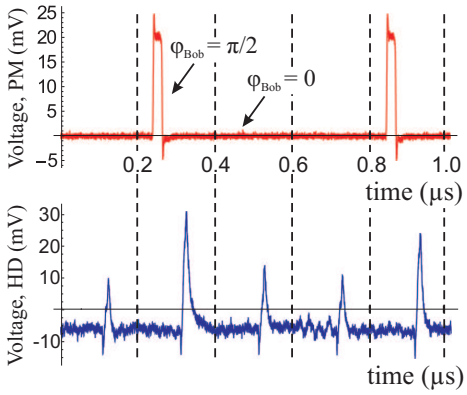

FIG. 1. *Experimental results of phase readout*. Bob's randomly-chosen phase modulation (in red) and output of Eve's homodyne detector (in blue) for a sequence of 5 arbitrarily chosen slots.

Although the phase readout succeeds with a very high accuracy, as may be observed in Fig. 1, the chosen intensity of Eve's pulses cause a strong afterpulsing in Bob's detectors. This has a severe impact on the QBER due to the massive increase in the dark count rate. We describe an elaborate strategy devised to overcome these problems and satisfy the above conditions for a regime of operating parameters. We also discuss the applicability of the attack strategy to other type of QKD systems along with countermeasures.

* nitin.jain@mpl.mpg.de

[1] N. Gisin *et al.*, Rev. Mod. Phys. **74**, 145195 (2002); V. Scarani *et al.*, Rev. Mod. Phys. **81**, 1301 (2009).
[2] A. Lamas-Linares and C. Kurtsiefer, Opt. Express **15**, 9388 (2007); Y. Zhao *et al.*, Phys. Rev. A **78**, 042333 (2008); S. Nauerth *et al.*, New J. Phys. **6**, 065001 (2009); L. Lydersen *et al.*, Nat. Photonics **4**, 686 (2010); N. Jain *et al.*, Phys. Rev. Lett. **107**, 110501 (2011).
[3] N. Gisin *et al.*, Phys. Rev. A **73**, 022320 (2006).
[4] A. Vakhitov *et al.*, J. Mod. Opt. **48**, 2023 (2001).
[5] That arise from different optical interfaces in the system.
[6] N. Jain *et al.*, *Manuscript in preparation*.
[7] ID Quantique website: `www.idquantique.com`
[8] V. Scarani, A. Acin, G. Ribordy and N. Gisin, Phys. Rev. Lett. **92**, 057901 (2004).