Invisible Trojan-horse attack

Shihan Sajeed,[1, 2, *] Carter Minshull,[3, 1] Nitin Jain,[4] and Vadim Makarov[3, 1, 2]

[1] *Institute for Quantum Computing, University of Waterloo, Waterloo, ON, N2L 3G1 Canada*
[2] *Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1 Canada*
[3] *Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, N2L 3G1 Canada*
[4] *Department of Physics, Technical University of Denmark, Fysikvej, Kongens Lyngby 2800, Denmark*

Quantum key distribution (QKD) allows two remote parties, Alice and Bob, to obtain symmetric keys – random but correlated sequence of bits – by exchanging quantum states [1]. The security relies on the fact that an adversary cannot eavesdrop without introducing noticeable errors. However, due to gaps existing between theory and practice, it is often possible for the eavesdropper Eve to implement some attacks to extract information about the keys without introducing noticeable errors. Such gaps can arise due to imperfections in the physical devices and/or incorrect assumptions in the theoretical security proofs [2, 3]. Hence, before trying to promote and commercialize QKD, it is of utmost importance to identify these gaps and evaluate their effects on the security.

This is where *security evaluation of QKD* comes in. It involves investigating practical QKD implementations to identify such theory-practice deviations, demonstrate the resultant vulnerability, and propose countermeasures to protect Alice and Bob from Eve. In this work, we have done such a security evaluation of the practical QKD system Clavis2 [4] running the Scarani-Acín-Ribordy-Gisin QKD protocol [5]. We evaluated its security against the so-called Trojan-horse attack (THA) [6] (introduced as 'large pulse attack' a few years before [7]).

In a THA, Eve can probe the properties of a component inside Alice or Bob by sending in a bright pulse into the system and analyzing a suitable back-reflected portion of it. This attack was previously attempted [8] at telecom wavelength with the intention to breach the security of a similar system and was successful in remotely reading the phase modulator settings in Bob via a homodyne measurement. However, the resultant increase in afterpulsing [9] at Bob's InGaAs single-photon detectors (SPDs) led to an elevated quantum bit error rate (QBER) which disclosed the presence of Eve.

In this work [10], we have tested the experimental feasibility of a THA that remains nearly invisible and evaluated the security of Clavis2 system against it. The invisibility has been achieved by using a wavelength where detectors are expected to be less sensitive. To quantify the decrease in afterpulsing probabilities, we have compared the afterpulsing at $\lambda_l = 1924$ nm with that at $\lambda_s = 1536$ nm as shown in Fig. 1. From Eve's point of view, the benefit of reduced afterpulsing at λ_l comes at the expense of a much higher attenuation inside Bob. Additionally, the degree of modulation received at λ_l differs from that at λ_s substantially. We quantify the increased optical attenuation and the suboptimal modulator response by means of further experimental measurements. Taking all these factors into account as well as optimizing the attack path, we calculate, through numerical simulation, Bob's actual QBER Q and Eve's actual knowledge of the secret key I_E^{act} for different combinations of attack parameters (see Ref. 10 for details of the attack parameters). For each combination, we check whether $I_E^{\text{act}} > I_E^{\text{est}}$ and $Q \leq Q_{\text{abort}}$ can be simultaneously satisfied. Here I_E^{est} and Q_{abort} are the estimated security bound on Eve's knowledge and maximum QBER permitted by the system respectively. For attacks at λ_s, no such combination was found [8]. However, for the current attack at λ_l, we have found several such combinations. For example, one of the attack combinations lead to $I_E^{\text{act}} = 0.515 > I_E^{\text{est}} = 0.506$ and $Q = 7.8\% < Q_{\text{abort}} \approx 8\%$ which clearly showed that a Trojan-horse attack at λ_l can extract some key information from the system at the same time being nearly invisible to Alice and Bob.

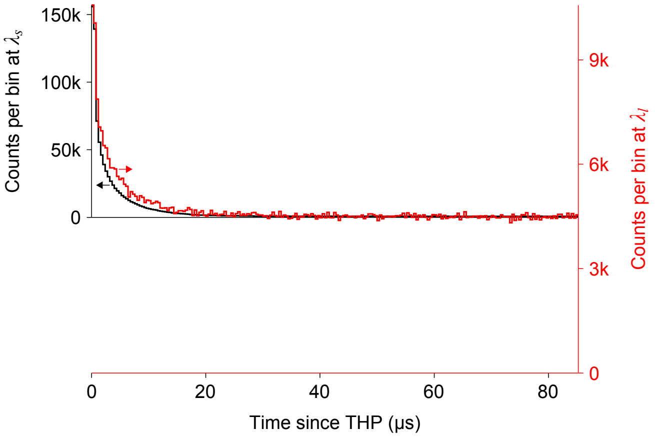

FIG. 1. Afterpulse profiles at $\lambda_s = 1536$ nm and $\lambda_l = 1924$ nm. The histograms are rescaled such that their peak counts and dark count rates match in the plot, making visual comparison of decay curves easy. The decay curves are similar but not identical. A total of 10^6 counts were histogrammed at each wavelength.

This work underscores that even when an attack seems to hopelessly fail at the first attempt [8], Eve can persist by tinkering with attack parameters until she suc-

[*] Corresponding author: shihan.sajeed@gmail.com

ceeds [10]. It poses a threat to the security of practical QKD if proper countermeasures are not adopted.

[1] C. H. Bennett and G. Brassard, in *Proc. IEEE International Conference on Computers, Systems, and Signal Processing (Bangalore, India)* (IEEE Press, New York, 1984) pp. 175–179.

[2] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, Rev. Mod. Phys. **81**, 1301 (2009).

[3] N. Jain, B. Stiller, I. Khan, D. Elser, C. Marquardt, and G. Leuchs, Contemp. Phys. **57**, 366 (2016).

[4] Clavis2 specification sheet, `http://www.idquantique.com/images/stories/PDF/clavis2-quantum-key-distribution/clavis2-specs.pdf`, visited 16 Apr 2017.

[5] V. Scarani, A. Acín, G. Ribordy, and N. Gisin, Phys. Rev. Lett. **92**, 057901 (2004).

[6] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, Phys. Rev. A **73**, 022320 (2006).

[7] A. Vakhitov, V. Makarov, and D. R. Hjelme, J. Mod. Opt. **48**, 2023 (2001).

[8] N. Jain, E. Anisimova, I. Khan, V. Makarov, C. Marquardt, and G. Leuchs, New J. Phys. **16**, 123030 (2014).

[9] C. Wiechers, L. Lydersen, C. Wittmann, D. Elser, J. Skaar, C. Marquardt, V. Makarov, and G. Leuchs, New J. Phys. **13**, 013043 (2011).

[10] S. Sajeed, C. Minshull, N. Jain, and V. Makarov, arXiv:1704.07749 [quant-ph].