Securing two-way quantum communication: the monitoring detector and its flaws

Shihan Sajeed,[1,*] Igor Radchenko,[2] Sarah Kaiser,[1] Jean-Philippe Bourgoin,[1]
Laurent Monat,[3] Matthieu Legré,[3] and Vadim Makarov[1]

[1] *Institute for Quantum Computing, University of Waterloo, Waterloo, ON, N2L 3G1 Canada*
[2] *General Physics Institute, Russian Academy of Sciences, Moscow, 119991 Russia*
[3] *ID Quantique SA, Chemin de la Marbrerie 3, 1227 Carouge, Geneva, Switzerland*

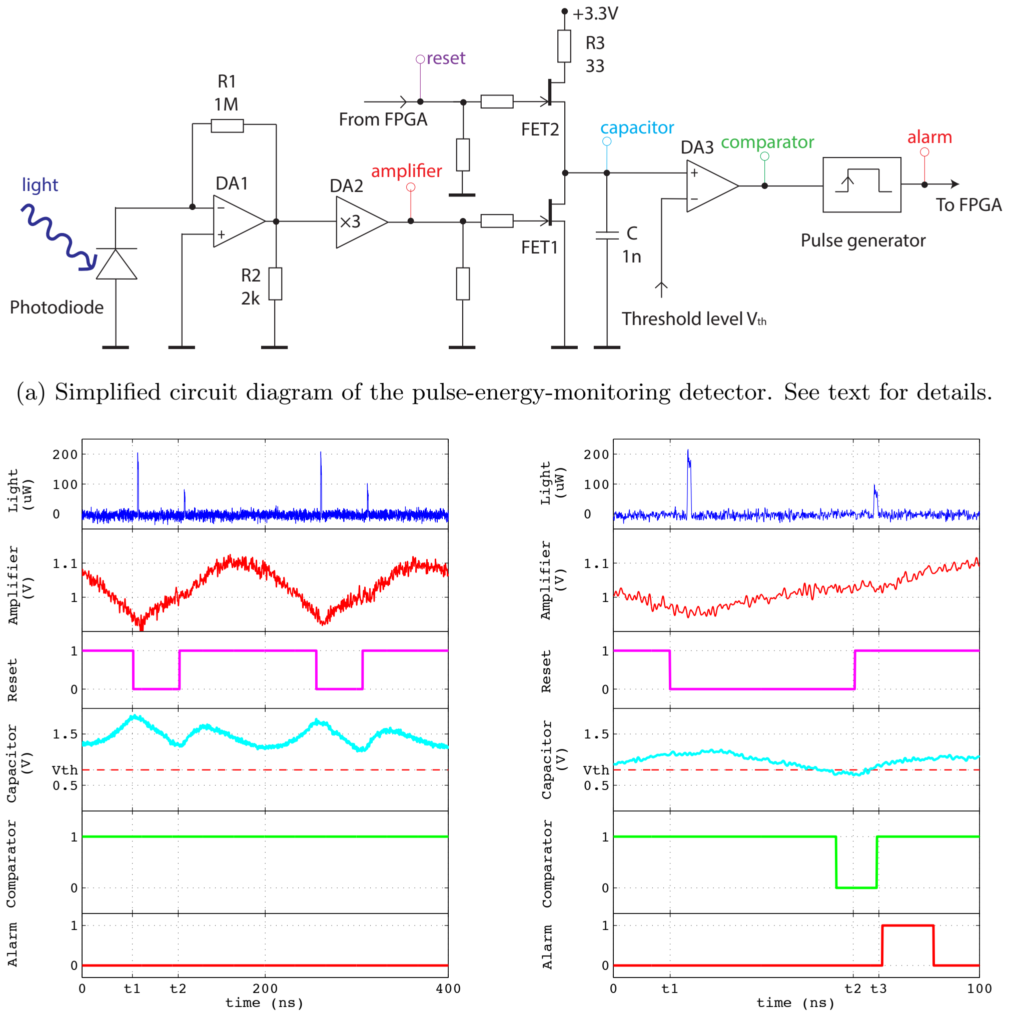
(Dated: May 5, 2014)

Monitoring incoming pulse energy is obligatory for any two-way system that sends bright pulses from Bob to Alice, such as plug-and-play and relativistic quantum cryptography schemes. Implementation of this monitoring detector has largely been ignored in experimental realizations so far. However, ID Quantique has implemented the hardware and associated software routines in their commercial system Clavis2. We scrutinize this implementation for security problems, and show that designing a hack-proof pulse-energy-measuring detector is far from trivial. Indeed the first implementation has three serious flaws (confirmed experimentally), each of which may be exploited in a cleverly constructed Trojan-horse attack. We model attack performance. We also discuss requirements for a loophole-free monitoring detector.

Two-pass optical schemes have significant practical advantages and are widely used today, e.g., in plug-and-play quantum key distribution [1] and relativistic quantum cryptography [2]. In any two-pass scheme, it is necessary for security to monitor the light coming to Alice from Bob. Otherwise, Eve could substitute a brighter pulse and check the reflected signal to estimate the bit value sent by Alice [3, 4]. The first implementation of such monitoring detector was done in ID Quantique's commercial QKD device Clavis2 [5]. In this study, we investigate the implemented detector and show that the current implementation is incapable of being perfectly secure [6]. We demonstrate three flaws in the implementation and show experimentally that each of the flaws can be exploited to breach the security. We model both an ideal attack and a practically implementable attack, and show that even the latter breaks security of the current implementation completely. We then discuss how to redesign the detector in a secure way.

Detector implementation. ID Quantique's monitoring detector receives $\approx 80\%$ of Alice's input light via fiber-optic beamsplitters at Alice's input. A fiber-pigtailed p-i-n photodiode is used to detect this light. Its photocurrent is processed by an electronic circuit shown in Fig. 1a, while signals at test points marked in the circuit are shown in Figures 1b and 1c. At the front-end there is a two-stage transimpedance amplifier, converting photocurrent into voltage signal. Due to insufficient bandwidth of the amplifier first stage (opamp DA1; Texas Instruments OPA380), it outputs slow-rising electrical pulses that extend to the next few bit slots and interfere with the signals from those slots. The amplifier output acts as a gate pulse for an N-channel field-effect-transistor (FET1 in Fig. 1a). The gate pulse for FET2 (reset signal) is applied by the field-programmable-gate-array (FPGA) system controller. This reset signal is normally high, keeping FET2 in a conductive state such that current flows through it to an integrating capacitor C. At time t_1, the reset signal switches FET2 into high-impedance state for 50 ns, and the capacitor starts to discharge through FET1 (see capacitor signal). The amount of discharge is higher when the power of incoming light is higher. At time t_2, reset signal switches FET2 into conductive state again and stops the discharging. This happens in each bit slot, and a negative spike proportional to the incoming light energy is generated at the capacitor. The negative spike is compared to a predefined threshold level V_{th}, whose value is calibrated at the factory in such a way that during normal operation the negative spike amplitude is very close but almost never goes below V_{th}. However, when there is an extra light, this voltage crosses the threshold causing the output of comparator DA3 to go low. This signal is fed to a pulse generator that produces fixed-width pulse on the low-to-high logic transition. This is the alarm signal fed to the FPGA that indicates the excess of incoming light. The system firmware discards a packet of ~ 1000 qubits (so-called frame) if one or more pulses inside the frame have triggered alarm. Thus any attempt by Eve to inject brighter pulses in a frame should lead to that frame being dropped from QKD.

Hacking. In this work we built our own Eve to take advantage of three different loopholes in the the above-mentioned implementation of Clavis2 [6]. The first loophole we exploited was the low bandwidth of the front-end transimpedance amplifier DA1, spreading its pulse response over several bit slots. For every N light pulses, we suppressed the first $N-1$ pulses and injected a brighter pulse at the Nth slot. Due to the blocked pulses, the DC average level at the output of the amplifier was lower than normal. When the much brighter pulse arrived at the Nth slot, the capacitor voltage was already higher than its calibrated value, and a much larger negative spike was needed to generate the alarm. We were able to inject up to 8.5 times (at N=4) more light at the Nth pulse than the calibrated signal. In an attack, Eve would attack



(a) Simplified circuit diagram of the pulse-energy-monitoring detector. See text for details.

(b) Signals at test points of Fig. 1a during normal operation.

(c) Generation of an alert signal when excess light is injected. The optical pulse has slightly higher energy than in Fig. 1b.

FIG. 1: Pulse-energy-monitoring circuit with oscillograms illustrating normal operation.

these pulses, while the blocked ones would be written off by Bob as lost to normal transmission losses.

Our second attack exploited the long recovery time of the front-end amplifier from negative saturation regime. In Clavis2 during the idle period between the frames (when no light is coming in), the opamp DA1 is driven into negative saturation. Once pulses appear, recovery from this saturation state takes ≈ 123 µs. For this reason, Alice starts applying phase modulation only from 140 µs or 701 pulses into the frame. In order to exploit this loophole, we removed pulses 501–700, which forced the amplifier to re-enter negative saturation. Then at 140 µs, we applied a series of bright pulses with varying spacing. These pulses could be up to 39.1 times brighter (at largest spacing) without triggering alarm.

The third attack exploited edge-triggering of the pulse generator. As seen in Fig. 1, the pulse generator generates the alarm pulse on low-to-high transition of its input. What if Eve sends light in such a way that there is never a low-to-high transition at the pulse generator? This is precisely what we did (see Fig. 2). We were able to inject pulses with energy up to 7.15 pJ or 97 times more than the normal calibrated signal pulse (the experimental limit was our source laser). At such high pulse energy, classical 100% extraction of bit value information should be easy in practice.

Modeling complete attacks. We modeled an optimal photon-number-splitting attack to estimate the information leaked to Eve when we increase the mean photon number μ emitted by Alice by a certain multiplication factor x. In order to grasp a practical scenario, we also modeled a practical beam-splitting attack using only existing components. Results for both BB84 and SARG protocols are shown in Figures 3 and 4 [6]. While

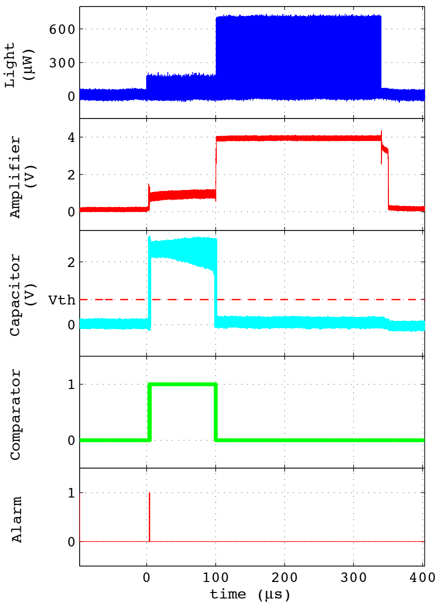


FIG. 2: Exploiting the edge-triggered alarm monitoring. At the start of the frame, normal pulses were sent to maintain Alice's synchronization. At 100 μs, bright pulses were injected which pulled the capacitor voltage far below $V_{\rm th}$. As a result the comparator output was constant low and never provided the low-to-high transition required to generate alarm.

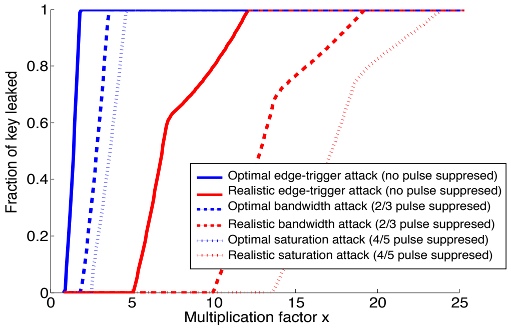


FIG. 3: Fraction of secret key leaked to Eve in BB84 protocol. The attacks that can increase μ in all pulses allow Eve to gain more information than the attacks that require suppression of pulses.

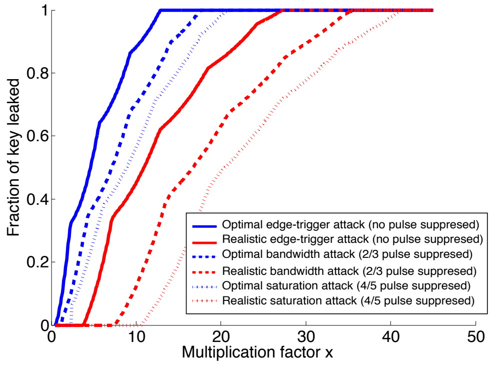


FIG. 4: Fraction of secret key leaked to Eve in SARG04 protocol. As with BB84, the attack that does not require Eve to suppress pulses (i.e., our last attack exploiting edge-triggering) performs better than our other two attacks that require pulse suppression.

all three attacks presented in this study work in the ideal scenario, only the last edge-triggering attack works in the practical scenario. We hope to improve on this, though.

Countermeasures. Although the considered strategy of the pulse-energy-monitoring module is generally correct, the circuit realization should be revised dramatically. Edge-triggering of the alarm is not needed in this circuit at all. Instead, a simple level triggering can be used. The front-end opamp should be replaced by a wideband amplifier that has neither bandwidth problem nor saturation behavior. Furthermore, in the present circuit the source current of FET1 nonlinearly depends on its gate voltage, which makes an imperfect integrator. For a perfect implementation of the integrator, the capacitor should be charged by a current that is linearly dependent on the photocurrent.

Conclusion. In this work, we point out the risk to security that exists when the communicating parties do not have an exact estimate of the systems security parameters (μ in this case). Even if they have the complete knowledge of the security parameters at the beginning, there is no guarantee that the parameters will remain the same as new classes of attacks are being reported that have the ability to change the properties of the system [7]. This poses the risk that an attack that seems implausible or impossible might become realistic. This emphasizes the need to examine implementation details, no matter how little, more closely now than ever before.

Acknowledgments. We thank N. Lütkenhaus and N. Jain for useful discussions.

* ssajeed@uwaterloo.ca

[1] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, New J. Phys. **4**, 41 (2002).

[2] I. V. Radchenko, K. S. Kravtsov, S. P. Kulik, and S. N. Molotkov, Laser. Phys. Lett. **11**, 065203 (2014).

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

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

[5] Clavis2 specification sheet, http://www.idquantique.com.

[6] S. Sajeed, I. Radchenko, S. Kaiser, J.-P. Bourgoin, M. Legré, and V. Makarov, manuscript in preparation.

[7] A. N. Bugge, S. Sauge, A. M. M. Ghazali, J. Skaar, L. Lydersen, and V. Makarov, Phys. Rev. Lett. **112**, 070503 (2014).