Gap between industrial and academic solutions to implementation loopholes: testing random-gate-removal countermeasure in commercial QKD system

Anqi Huang,[1,2] Shihan Sajeed,[1,2] Poompong Chaiwongkhot,[1,3]
Mathilde Soucarros,[4] Matthieu Legré,[4] and Vadim Makarov[1,3,2]

[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] *ID Quantique SA, Chemin de la Marbrerie 3, 1227 Carouge, Geneva, Switzerland*

Reviewers kindly note that this abstract contains currently confidential information that may not be publicly disclosed before the conference.

Practical security issues in quantum key distribution (QKD) systems have attracted an increasing attention in the last decade. To bridge the gap between theoretical security and practical imperfect implementations, one possible approach is introduction of side-channel-free QKD protocols, and another is patching loopholes in existing protocols. However, in the process of solving the security problems, another gap has emerged between the academic countermeasures and industrial realizations. The excellent academic protocols (e.g., measurement-device-independent QKD [1]) are not adopted by the industry yet, because their implementation in a customer-friendly product with stable performance is too challenging. Aside of the technical difficulty, a more critical factor is the mindset. Ideally, the industry should be able to implement robust fixes to the security problems in commercial products, and get them tested independently to verify the quality of implementations. However the industry is not quite there yet. A few companies (such as ID Quantique and SeQureNet) are trying to break the mold, and are letting third-party labs examine company's solutions to previously discovered loopholes. Meanwhile, it seems that most of the industry has their heads firmly planted into sand, and does not invite independent testers to examine their "countermeasure implementations" hands-on.

This work examines ID Quantique's attempted countermeasure to the earlier discovered detector control attacks [2] that were demonstrated 5 years ago on ID Quantique's and MagiQ Technologies' QKD products. The present work indicates, unfortunately, that the first countermeasure implementation against this attack in ID Quantique's system is ineffective.

The timeline of this security problem is shown in Fig. 1. In 2009, we found that Clavis2 QKD system was vulnerable to detector blinding attack and submitted a confidential report about this loophole to ID Quantique (the work was published shortly afterwards [2]). After this, ID Quantique has been trying to figure out an experimental countermeasure against such kind of attack. In 2010, ID Quantique proposed a countermeasure that randomizes the efficiency of a gated avalanche photodiode (APD) by randomly choosing one out of two different gate voltages, and applied for a patent [3]. In this way, Eve does not know the exact efficiency of Bob in every slot. At the sifting phase, if the observed detection rates differ from the expected values, Alice and Bob would be aware of the presence of Eve and discard the raw keys. In 2014, Lim *et al.* have published a paper to propose a specific protocol to realize this countermeasure [4], which takes the blinding attack into account and analyses the security mathematically. This solution intends to introduce an information gap between Eve and Bob, for Eve has no information about Bob's random efficiency choice. In 2014, ID Quantique has implemented the countermea-

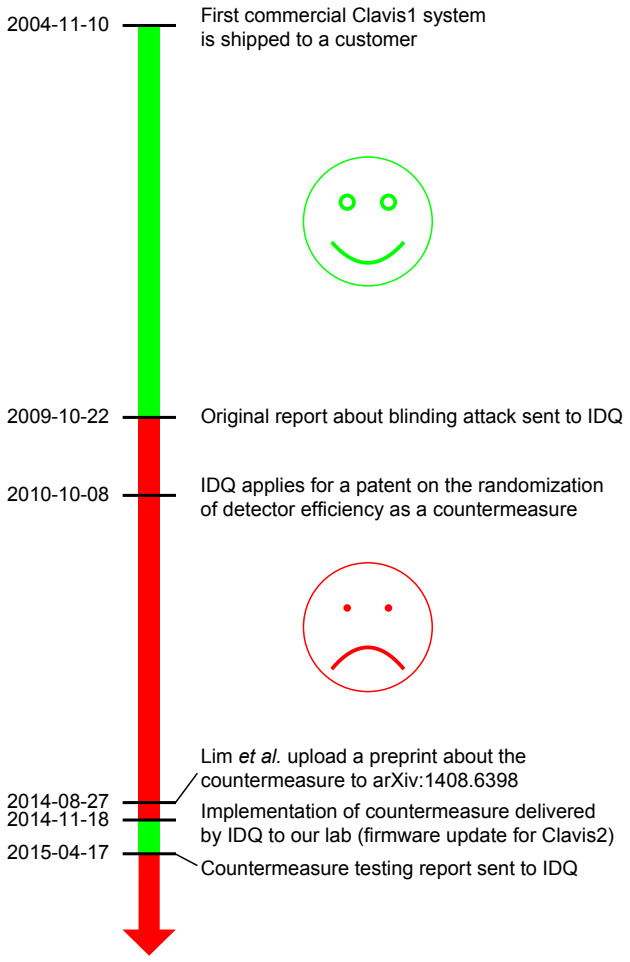

FIG. 1. Timeline of hacking-countermeasure-hacking for the bright-light detector control attack.

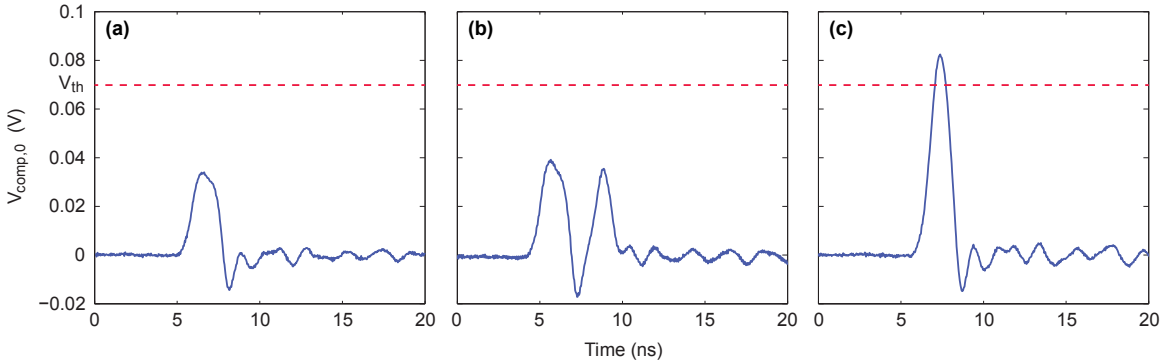

FIG. 2. Signal at the comparator input in a detector blinded with 0.64 mW c.w. illumination. (a) no trigger pulse is applied. (b) 0.61 pJ trigger pulse is applied 5 ns after the gate. (c) 0.61 pJ trigger pulse is applied in the gate.

sure as a firmware patch. Since the hardware available in Clavis2 is not directly capable of generating two nonzero efficiency levels switched randomly between adjacent detector gates, implementation is in a simplified form in which the gate is suppressed randomly with 2% probability. The suppressed gates represent zero efficiency, while the rest of the gates represent calibrated efficiency η. Although it was noted in Ref. 4 that such reduced implementation might be insecure, it was the only one easily implementable in Clavis2.

According to ID Quantique's idea behind this countermeasure implementation, Eve cannot know the slots of gate absence and cannot avoid causing clicks during these slots. A single click without a gate reveals Eve's attack. Thus she cannot fully control Bob's detectors and obtain all secret keys. Unfortunately, here we show a way to hack this improved system. Our approach is similar to the original blinding attack [2]. Firstly, bright-light continuous-wave (c.w.) laser light is sent to blind Bob's detectors. Then Eve sends a trigger pulse on top of the c.w. illumination *during the detector gate,* as opposed to sending the trigger pulse slightly after the gate as in the original attacks [2]. A click is produced in one of Bob's two detectors only if Bob's basis matches Eve's and he applies his gate; otherwise there is no click.

To explain why this attack succeeds, let's first elaborate the operating principle of an avalanche photodiode (APD). The detectors in Clavis2 are gated APDs. When no gate signal is applied, the APD is biased slightly below its breakdown voltage $V_{\rm br}$ by a high-voltage supply. To bring the APD into Geiger mode in which APD is sensitive to single photon, an additional 3 V high gate pulse is applied across APD to make the bias voltage greater than $V_{\rm br}$. When single photon comes during the gated time, an avalanche happens and APD generates a large current. The avalanche current is sensed by a small load resistor and AC-coupled fast comparator (the AC coupling is via a capacitor, in order to prevent the high DC bias voltage from damaging the low-voltage comparator). If the peak current of the avalanche pulse causes a larger voltage across the load resistor than the comparator threshold $V_{\rm th} = 70$ mV, the comparator produces a logic output signal indicating a photon detection (a click).

Under bright c.w. illumination, the APD produces constant photocurrent that overloads the high-voltage supply and lowers the APD voltage. Then, even in the gate the voltage across APD does not reach $V_{\rm br}$, and the APD remains in the linear mode as a classical photodetector (with some finite internal gain owing to a limited amount of avalanche multiplication). It is no longer sensitive to single photons.

Under the blinding attack, Fig. 2 shows the signal at the comparator input when no trigger pulse is applied, and when it is applied either after or in the gate. Since in the linear mode the APD gain depends on the voltage across it, increased bias voltage when gate is applied corresponds to a larger gain that assists the APD in generating larger current under c.w. illumination. Therefore the gate signal causes a positive pulse at the comparator input, shown in Fig. 2(a). The trigger pulse applied after the gate produces a second pulse, but the peak voltage of neither pulse exceeds $V_{\rm th}$ [Fig. 2(b)]. However, when the trigger pulse is shifted inside the gate, the two pulse amplitudes add and $V_{\rm th}$ is reached, producing the detector click [Fig. 2(c)]. Thus we can make a click conditionally on the gate, defeating the countermeasure.

The key technology of our attack is controlling the energy of trigger pulse. Different amount of energy will result in different detection probabilities with and without the gate. If trigger pulse energy $E \leq E^{\rm gate}_{{\rm never},i}$ (where $i \in \{0,1\}$ is detector number), the detector never clicks whether there is a gate or not. If $E \geq E^{\rm gate}_{{\rm always},i}$, the detector always clicks when the gate is applied. If $E \leq E^{\rm no\,gate}_{{\rm never},i}$, the detector never clicks when the gate is not applied.

To test practical detector controllability in Clavis2, we use 1 ns wide trigger pulse coinciding in time with the gate. The click thresholds for a range of c.w. blinding powers are shown in Fig. 3. As can be seen, for any given blinding power, $E^{\rm no\,gate}_{{\rm never},i}$ is much higher than the other click thresholds. This easily allows the original detector control attack [2] to proceed undetected by the countermeasure. A more formal analysis follows.

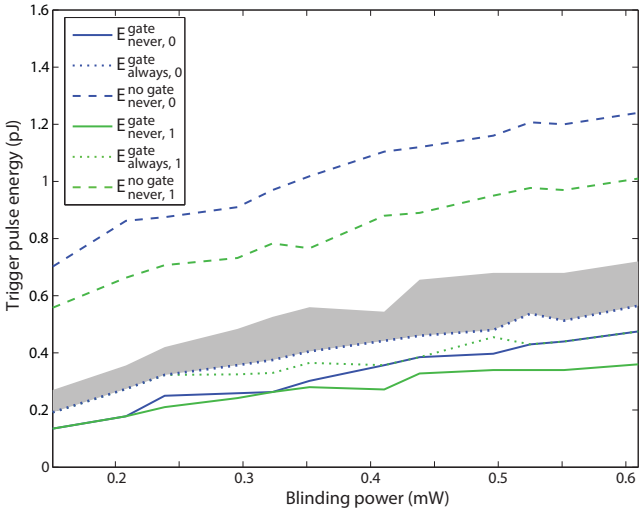

FIG. 3. Click thresholds versus c.w. blinding power. Shaded area shows the range of trigger pulse energies of the perfect attack.

After blinding the APDs, Eve may eavesdrop the secret key with a faked-state attack [5]. Eve intercepts the quantum states from Alice measuring them in a random basis. She resends her detection results to Bob via a bright trigger pulse of a certain energy, superimposed on her blinding c.w. illumination. For a perfect attack, Eve triggers a click in Bob's detector with 100% probability if their bases match and gate is applied, and 0% probability if their bases do not match or gate is absent. The need to avoid causing a click when the gate is absent requires that

$$E^{\text{no gate}}_{\text{never},i} > E^{\text{gate}}_{\text{always},i} > E^{\text{gate}}_{\text{never},i}. \qquad (1)$$

If Eve and Bob select opposite bases, half of the energy of trigger pulse goes to each Bob's detector. In this case, none of the detectors should click despite the gate presence. This is achieved if [2]

$$\max_i \left\{ E^{\text{gate}}_{\text{always},i} \right\} < 2 \left(\min_i \left\{ E^{\text{gate}}_{\text{never},i} \right\} \right). \qquad (2)$$

Satisfying inequalities (1) and (2) represents the perfect attack conditions and guarantees the same performance as in Ref. 2.

In the entire tested range of 0.1–0.6 mW blinding power, the thresholds satisfy the conditions for the perfect attack with a wide margin. The shaded area in Fig. 3 represents the range of trigger pulse energies suitable for the perfect attack.

Note that an attack may still be possible even if the perfect attack conditions are not met [6]. The random gate removal countermeasure only requires that $E^{\text{no gate}}_{\text{never},i} > E^{\text{gate}}_{\text{never},i}$, which means Eve should be able to at least sometimes cause a click in the gate while never causing a click without the gate (lest she is discovered). This is a much weaker condition. It is likely to always be satisfied because of inherent APD characteristics.

Conclusion. We have demonstrated that the gated single-photon detectors used in Clavis2 system with the first implementation of countermeasure are still fully controllable by Eve via bright light, without getting detected by the countermeasure. Our attack method is similar to the previous detector blinding attack, with the difference that the trigger pulse needs to be time-aligned to coincide with the detector gate instead of following it. A detailed confidential technical report, on which this submission is based, has already been sent to ID Quantique.

Our work is an example that shows some gap between the academic community and industry. To close it, industrial engineers can try to realize advanced secure protocols developed in the academia, and academic researchers can try to propose more implementation-friendly countermeasures. It also hints that these two parties should cooperate to figure out countermeasures, implement them and test them. In terms of this specific countermeasure against the detector blinding attack, the performance of a full proposed implementation with more than one non-zero efficiency level [4] is still to be tested.

The Waterloo testing team also concludes that, first, there is not enough money in the QKD market. They would expect a larger market to support a faster and more intense patching and testing process than shown in Fig. 1. Second, addressing practical vulnerabilities at the design stage of a QKD system is both cheaper and less messy than trying to retrofit patches on an existing deployed solution. Addressing security at the design stage should be the goal whenever possible. *Both the testing team and ID Quantique* agree that these design-stage goals should be formulated as a security target in a certification framework as with any other security equipment. The standardization of a security target will require to take into account the existing security targets for crypto and key management hardware devices, as well as to include new potential vulnerabilities arising from the quantum aspects which are not yet documented into a certification framework. Such QKD implementation vulnerabilities need to be well characterized in order to standardize the process and produce systematic and replicable certification criteria. Without such a certification standard for QKD, the market will remain limited.

[1] H.-K. Lo, M. Curty, and B. Qi, Phys. Rev. Lett. **108**, 130503 (2012).
[2] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nat. Photonics **4**, 686 (2010).
[3] M. Legre and G. Ribordy, international patent appl. WO 2012/046135 A2 (filed 2010-10-10, published 2012-04-12).
[4] C. C. W. Lim, N. Walenta, M. Legré, N. Gisin, and H. Zbinden, IEEE J. Sel. Top. Quantum Electron. **21**, 6601305 (2015).
[5] V. Makarov and D. R. Hjelme, J. Mod. Opt. **52**, 691 (2005).
[6] L. Lydersen, N. Jain, C. Wittmann, Ø. Marøy, J. Skaar, C. Marquardt, V. Makarov, and G. Leuchs, Phys. Rev. A **84**, 032320 (2011).