Certification of a commercial quantum key distribution system against implementation loopholes

Vadim Makarov,[1, 2, 3, *] Alexey Abrikosov,[1, 3] Poompong Chaiwongkhot,[4, 5, 6, 7] Aleksey K. Fedorov,[1, 8]
Anqi Huang,[9] Evgeny Kiktenko,[1, 3, 10] Mikhail Petrov,[2, 1, 11, 3] Anastasiya Ponosova,[1, 3]
Daria Ruzhitskaya,[1, 3] Andrey Tayduganov,[3, 8] Daniil Trefilov,[2, 1, 11, 3, 12, 13] and Konstantin Zaitsev[2, 1, 11, 3, 12]

[1] *Russian Quantum Center, Skolkovo, Moscow 121205, Russia*
[2] *Vigo Quantum Communication Center, University of Vigo, Vigo E-36310, Spain*
[3] *NTI Center for Quantum Communications, National University of Science and Technology MISiS, Moscow 119049, Russia*
[4] *Department of Physics, Faculty of Science, Mahidol University, Bangkok, 10400 Thailand*
[5] *Institute for Quantum Computing, University of Waterloo, Waterloo, ON, N2L 3G1 Canada*
[6] *Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, N2L 3G1 Canada*
[7] *Quantum technology foundation (Thailand), Bangkok, 10110 Thailand*
[8] *QRate, Skolkovo, Moscow 143026, Russia*
[9] *Institute for Quantum Information & State Key Laboratory of High Performance Computing, College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, People's Republic of China*
[10] *Steklov Mathematical Institute, Russian Academy of Sciences, Moscow 119991, Russia*
[11] *atlanTTic Research Center, University of Vigo, Vigo E-36310, Spain*
[12] *School of Telecommunication Engineering, Department of Signal Theory and Communications, University of Vigo, Vigo E-36310, Spain*
[13] *National Research University Higher School of Economics, Moscow 101000, Russia*

We report recent advances in the development of certification for quantum key distribution (QKD) systems. We give an example of a commercial QKD system that we have analysed for possible loopholes, improved to close the vulnerabilities identified, and designed a set of tests for that can be used by a certification lab [1]. We explain some of the testbenches in this lab, such as an ultrawide spectral characterisation testbench [2], automated detector testing [3], and laser damage testbench that verifies the quality of a power limiter [4]. This work is in line with the requirements of the ISO standard for QKD [5] and paves the way for the creation of certification services.

Cryptographic systems must undergo a formal certification in order to enable their wide deployment. This certification includes the system's resistance to known vulnerabilities. It is now being created for QKD, with an international standard [5] and a catalog of attacks [6] recently published, and a 16 M€ development program launched by the European Commission [7].

Here we report the work we've done in this direction primarily for Russian domestic certification of QKD. However, our methodology is open and readily applicable to the international certification and systems from different vendors. Over the past few years, we have focused on a fiber-optic QKD system from QRate [1]. It has a prepare-and-measure scheme and uses a decoy-state BB84 protocol with polarisation-encoded states at approximately 1550 nm wavelength and 312.5 MHz clock rate. Its optical scheme is shown in Fig. 1.

Preparing a QKD system for certification involves (i) documenting the system in sufficient detail for it to be analysed, (ii) analysing it, (iii) patching the security loopholes found [8], and (iv) proposing the requirements for future certification tests. These four steps should be completed by the developer of the QKD system and possibly involve an external security analysis team. We perform them for this system, utilising the latest developments in vulnerabilities, countermeasures, and security proofs. This is to be followed by (v) the actual implementation of certification, for which we prototype testbenches and testing methodology.

Preliminaries. At the documentation analysis stage, it is useful to rank the vulnerabilities by their risk, in order to prioritise vendor's work on eliminating them. The standard scale for the difficulty of exploit [5] is poorly applicable to our vulnerabilities, thus we are using our own scale. We ask three questions about each vulnerability: "Do we think the vulnerability likely exists?"; "Is it exploitable with present-day technology?"; and "Does it give the attacker a full or nearly-full information about the secret key?" If all three answers are positive, the vulnerability is high-risk (**H**). If only two are positive, it's medium-risk (**M**); one or zero, low-risk (**L**). The vendor then patches the high-risk issues first. The remaining issues could be addressed as well or they could be verified in the course of the formal certification before taking any action.

Several vulnerabilities are addressable by quantifying a partial information leakage and eliminating it with privacy amplification. Unfortunately, no security proof exists that accounts for all such vulnerabilities simultaneously. We have thus empirically decided to reduce the individual leakage for each vulnerability to a negligibly low amount, by strengthening countermeasures.

[*] makarov@vad1.com

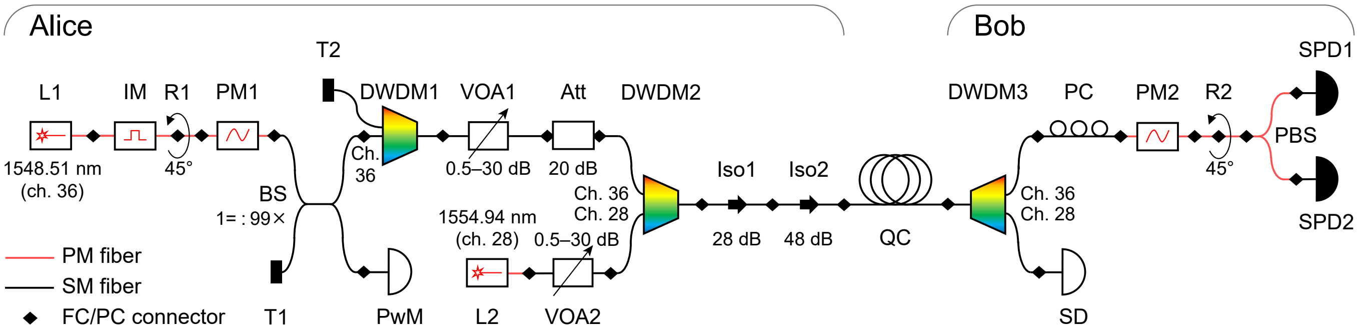

FIG. 1. Optical scheme of the QKD system under evaluation [1].

TABLE I. Summary of potential security issues in the system under evaluation. Q, implementation layers (see [8]).

Potential security issue	Q	Target component	Action recommended to the vendor or patch applied	Risk evaluation
Choice of QKD protocol	Q5	Protocol	None needed.	Solved
Superlinear detector control	Q1–5,7	SPDs	Develop photocurrent-measurement countermeasure.	**H**[a]
Detector efficiency mismatch	Q1–5	SPDs, Bob's PM	Patched by four-state Bob.	**H**[a]
Detector deadtime	Q1,2,5	SPDs	Add simultaneous deadtime in post-processing.	**H**[a]
Trojan-horse	Q1,2	Alice's optics	Characterise Alice's components in a wide spectral range. Install additional isolators.	**L**
Laser seeding	Q1,2	Laser	None needed.	Solved
Light injection into Alice's power meter	Q1–3	IM	Characterise Alice's components in a wide spectral range. Install additional isolators.	**L**
Induced photorefraction	Q1–3	Alice's IM and PM	Characterise Alice's components in a wide spectral range. Characterise the effect in modulators.	**M**
Laser damage	Q1	Alice's & Bob's optics	Install a power limiter at Alice's exit.	**M** **M**
Backflash from avalanche photodiodes	Q1,2	SPDs	Measure backflash photon emission probability. Characterise Bob's components in a wide spectral range.	**M**
Intersymbol interference	Q1–3	Alice's active components	Characterise state-preparation imperfections.	**L**
Imperfect state preparation	Q1–3,5	Alice's optics	Characterise state-preparation imperfections.	**L**
Calibration via channel Alice–Bob	Q1–5	SPDs, IM, PM	Alice's calibrations made local. Bob's remain but are patched by four-state Bob.	**H**[a]
Non-quantum random number generator	Q5	Protocol	Use a physical quantum random number generator.	**L**
Compromised supply chain	All	Any	Learn mitigation strategies from the national cryptography licensing authority.	**M**

[a] All the high-risk issues identified have been addressed by QRate before publication of this report.

Security analysis and countermeasures applied. Our initial analysis yielded 15 potential vulnerabilities, summarised in Table I. Let's discuss them briefly [1].

The decoy-state BB84 *protocol* this system uses has a well-scrutinised general security proof and also many proofs that account for imperfections. This is very good.

Bob uses sinusoidally-gated single-photon detectors based on avalanche photodiodes, which allow *superlinear control.* These detectors are fully blindable and controllable with bright light. The photocurrent monitor intended to alarm against this attack could originally be circumvented by pulsed blinding [3]. QRate has widened the monitor circuit bandwidth to patch the latter. Meanwhile, testing for non-blinding after-gate and edge attacks will be part of certification.

Detector efficiency mismatch, definitely a vulnerability, has been solved by the implementation of four-state Bob. I.e., Bob now randomly applies one of four phase shifts at his phase modulator that not only choose his measurement basis, but also randomly flip his detectors' bit-value assignment.

Simultaneous *detector deadtime* is implemented by an electrical cross-link between Bob's detectors. We have found that it needs to be supplemented with post-processing, to avoid efficiency mismatch at the edges of the hardware deadtime.

To prevent a *Trojan-horse attack,* a sufficiently high optical isolation is needed between Alice's line exit and her modulators. This isolations needs to be maintained at all wavelengths that fiber can carry, because Eve is free to inject any wavelength. This necessitates characterising the insertion loss of every component in the path in the wide spectral range. A special testbench has been created for this.

Luckily, the existing isolation at 1550 nm prevents an effective *seeding of Alice's laser,* with a large margin.

Another point of vulnerability to light injection is *Alice's internal power meter* (PwM), which is used to maintain the working point of her intensity modulator. The wideband characterisation of isolation is also needed here.

Induced photorefraction is a recently discovered attack that tampers with the modulators via a short-wavelength light injection. The characterisation of isolation is necessary here, too.

To prevent *laser damage,* we add a power-limiting component at Alice's exit. A certain type of fiber-optic isolator seems to be a good fit for this job [4].

Light emission from Bob's detectors *(backflash)* needs to be characterised, both in photon emission probability and spectrum. Then, spectral filtering at Bob's entrance and wideband characterisation of his components should reduce the information leakage to a negligible level.

State preparation flaws of the source (both *average imperfect state preparation* and *intersymbol interference*) should be characterised, for the intensity states and polarisation states. Two special testbenches and a data processing methodology have been created for this [9, 10].

To the analysis team's surprise (and horror), the system originally performed several internal *calibrations by using photons sent over the quantum channel* and public data from Bob's detectors. This immediately opens several vulnerabilities, because Eve can set any of the parameters being calibrated by tampering with these photons. The parameters are the timing of modulation pulses at both Alice's modulators, at Bob's modulator, and the timing of Bob's detector gates. QRate has eliminated the former two calibrations (making them internal within Alice) and applied the four-state Bob patch to prevent attacks exploiting Bob's efficiency mismatch.

We had to remind QRate to implement *quantum random number generators* in Alice and Bob, as required by the security proof. The prototype we analysed lacked this hardware.

Finally, we can't help noticing that almost all the system components are sourced from third-party suppliers. The system is thus in principle vulnerable to *intentional compromise* of any of them by a malicious supplier, not unlike any classical cryptography hardware. This is a known risk and mitigating it should be part of certification.

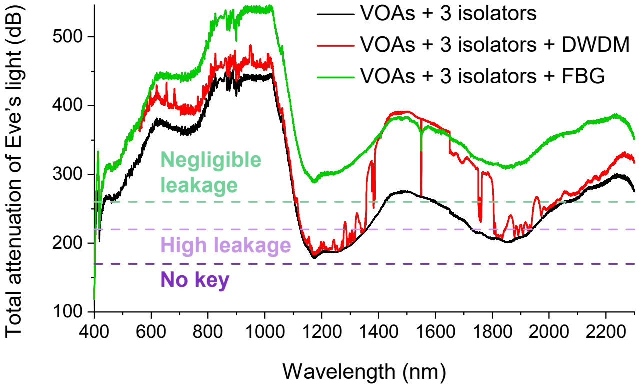

FIG. 2. Spectral characterisation of three source configurations against the Trojan-horse attack (adapted from [2]).

Structure of certification lab. Five testbenches are needed to cover all the 'quantum' issues in this system.

1. Wideband spectral characterisation of components.
2. Characterisation of detector controllability.
3. Characterisation of state preparation flaws.
4. Characterisation of light emission from SPDs.
5. Laser damage.

Our *spectral characterisation testbench* uses a bright supercontinuum white-light fiber source (NKT Photonics) and spectrum analysers (Yokogawa) that together cover 400–2400 nm range with 30–65 dB dynamic range [2]. This is sufficient to characterise insertion loss for all the above attacks that need it. An example characterisation of different source configurations against the Trojan-horse attack is shown in Fig. 2.

Our *detector testbench* checks the detector's controllability in both continuous-wave and pulsed blinding attack, as well as the performance of its integrated countermeasure [3]. The test process is automated and outputs a detailed report that states whether the detector is controllable. The data in the report also allows the operator to conclude whether the countermeasure performs satisfactorily and catches both attacks.

Our *laser damage testbench* verifies the performance of the power-limiting component at Alice's exit [4]. It subjects the component to 1550 nm cw light of up to 6.7 W and monitors its insertion loss in both directions. The test is designed to verify that this component always attenuates Eve's light to a safe level, either reversibly or not (acting as a sacrificial fuse that breaks the communication line).

Experimental results from the other two testbenches are currently being analysed [10, 11].

In summary, we have prepared a QKD system of a commonly used type for certification. We have developed and prototyped certification methodology against the known vulnerabilities. The application of our results in the upcoming certification process is straightforward.

[1] Vadim Makarov, Alexey Abrikosov, Poompong Chaiwongkhot, Aleksey K. Fedorov, Anqi Huang, Evgeny Kiktenko, Mikhail Petrov, Anastasiya Ponosova, Daria Ruzhitskaya, Andrey Tayduganov, Daniil Trefilov, and Konstantin Zaitsev, "Preparing a commercial quantum key distribution system for certification against implementation loopholes," arXiv:2310.20107 [quant-ph].

[2] Hao Tan, Mikhail Petrov, Weiyang Zhang, Liying Han, Anqi Huang, Sheng-Kai Liao, Vadim Makarov, and Feihu Xu, "Wide-spectrum security against attacks in quantum key distribution," (2024), unpublished.

[3] Polina Acheva, Konstantin Zaitsev, Vladimir Zavodilenko, Anton Losev, Anqi Huang, and Vadim Makarov, "Automated verification of countermeasure against detector-control attack in quantum key distribution," EPJ Quantum Technol. **10**, 22 (2023).

[4] Anastasiya Ponosova, Daria Ruzhitskaya, Poompong Chaiwongkhot, Vladimir Egorov, Vadim Makarov, and Anqi Huang, "Protecting fiber-optic quantum key distribution sources against light-injection attacks," PRX Quantum **3**, 040307 (2022).

[5] "ISO/IEC 23837-2:2023(en). Information security — Security requirements, test and evaluation methods for quantum key distribution — Part 2: Evaluation and testing methods," `https://www.iso.org/obp/ui/en/#iso:std:iso-iec:23837:-2:ed-1:v1:en`, visited 22 Nov 2023.

[6] Christoph Marquardt, Ulrich Seyfarth, Sven Bettendorf, Martin Bohmann, Alexander Buchner, Marcos Curty, Dominique Elser, Silas Eul, Tobias Gehring, Nitin Jain, Thomas Klocke, Marie Reinecke, Nico Sieber, Rupert Ursin, Marc Wehling, and Henning Weier, "Implementation attacks against QKD systems," BSI technical report, `https://www.bsi.bund.de/EN/Service-Navi/Publikationen/Studien/QKD-Systems/Implementation_Attacks_QKD_Systems_node.html`, visited 14 Feb 2024.

[7] "Testing and evaluation infrastructure for EuroQCI," European Commission's tender call closed 29 Aug 2023, `https://etendering.ted.europa.eu/cft/cft-display.html?cftId=14339`, visited 3 Apr 2024.

[8] Shihan Sajeed, Poompong Chaiwongkhot, Anqi Huang, Hao Qin, Vladimir Egorov, Anton Kozubov, Andrei Gaidash, Vladimir Chistiakov, Artur Vasiliev, Artur Gleim, and Vadim Makarov, "An approach for security evaluation and certification of a complete quantum communication system," Sci. Rep. **11**, 5110 (2021).

[9] Anqi Huang, Akihiro Mizutani, Hoi-Kwong Lo, Vadim Makarov, and Kiyoshi Tamaki, "Characterization of state-preparation uncertainty in quantum key distribution," Phys. Rev. Appl. **19**, 014048 (2023).

[10] Daniil Trefilov, Xoel Sixto, Víctor Zapatero, Anqi Huang, Marcos Curty, and Vadim Makarov, "Intensity correlations in decoy-state BB84 QKD systems," (2024), unpublished.

[11] Konstantin Zaitsev, Dmitriy Kuzmin, and Vadim Makarov, "Energy-time attack on detectors in quantum key distribution," (2024), unpublished.