Polarization-randomized gateway against detector-blinding hacks of quantum key distribution

Salem F. Hegazy*
National Institute of Laser Enhanced Sciences, Cairo University, 12613 Giza, Egypt
shegazy@zewailcity.edu.eg

Salah S. A. Obayya
Centre for Photonics and Smart Materials, Zewail City of Science and Technology, 12578 Giza, Egypt
sobayya@zewailcity.edu.eg

Bahaa E. A. Saleh*
CREOL, The College of Optics & Photonics, University of Central Florida, Orlando, FL 32816, USA
besaleh@creol.ucf.edu

***Abstract*—A quantum key distribution system –employing a *key* time-bin qubit and a *security-pass* polarization-randomized qubit– is shown to overcome a wide class of intercept-resend attacks adopting the use of faked-state light; including attacks based on blinding of single-photon avalanche detectors (SPADs).**

***Index Terms*—single-photon detector control, faked state light, optical fiber communication, bidirectional quantum key distribution.**

I. Introduction

Although the unconditional security of quantum key distribution (QKD) has been established in principle and verified under idealized conditions, implementations using available components offer vulnerabilities that have opened the door for several practical schemes of quantum hacking. Most notable is a well-known class of attacks adopting the use of faked-state photons. This includes detector-control attacks and, in a more general form, the intercept-resend strategies [1]. Here, we put forward a QKD scheme that renders such type of attacks impossible.

II. QKD system

The QKD system is shown in Fig. 1. The legitimate users, Bob and Alice, exchange the *quantum key* encoded in a time-bin qubit together with a *security pass* via an ancillary polarization qubit in a roundtrip transmission from Bob to Alice and back. Bob employs a reciprocal polarization randomizer at his gateway which distorts the outgoing polarization state. A Faraday mirror (FM) at Alice's site transforms the incoming polarization state to its orthogonal, which compensates for any slowly varying random birefringence when back-tracing the fiber link. Thanks to the FM, upon back passage through the reciprocal randomizer, the the outgoing polarization state is restored. However, the polarization state of a faked photon from an intruder, Eve, sent directly to Bob is randomized and hence directed to a SPAD in a different path, whereupon it triggers an alert.

The system operates as follows. Bob sends a single-photon pulse in polarization-path state $\frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)|1_s\rangle$ through a polarizing Mach-Zehnder interferometer (PMZI) that copies the polarization state into a time-bin state, and the path state into a polarization state. The outcome is a time-bin-polarization state $\frac{1}{\sqrt{2}}(|t_1\rangle + |t_2\rangle)|H\rangle$, which is transmitted through the randomizer and undergoes a random reciprocal rotation P, known only to Bob. The randomization P is kept fixed during the photon roundtrip.

Alice encodes the leading time-bin by a random phase ϕ_A taking one of the values: $\{0, \pi\}$; or $\{\pi/2, 3\pi/2\}$. After Faraday-mirror reflection, the photon is sent back to Bob, passing again through the randomizer, This produces the state $\frac{1}{\sqrt{2}}(e^{i\phi_A}|t_1\rangle + |t_2\rangle)|V\rangle$. Bob's SPADs are gated to only measure the interfered possibilities after short-long or long-short double-pass in the PMZI. The PMZI then copies back the polarization state of the received photon into the path state, and the time-bin state into a polarization state. This yields the state $\frac{1}{\sqrt{2}}(|H\rangle + e^{i\phi_A}|V\rangle)|2_s\rangle$ which Bob measures in either diagonal or circular polarization basis. While this random basis choice is made passively in path 2_s, it is performed actively in path 1_s by means of a polarization switch (PS) controlled by a quantum random number generator.

Bob's receiver is therefore configured such that the legitimate photon is detected in path 2_s, called the secure path, while photons produced by an intercept-resend attacker, Eve, who does not know the randomizer state P, will be randomly detected in path 1_s, triggering an alert. Within the session, Bob may switch the polarization randomizer to send the genuine photon intentionally to the alert SPADs to verify their proper operation. After the quantum transfer session completes, Bob checks the count of alert events, if within the range of error tolerance, Alice and Bob follow the usual steps of the BB84 protocol. If not, the session aborts.

Now, let us consider that Eve intercepts the legitimate photon, measures its state in polarization and in time, and then sends to Bob faked-state photon(s) of a time-bin state $\frac{1}{\sqrt{2}}(e^{i\phi_E}|t_1\rangle + |t_2\rangle)$ and a polarization state ρ. It can be derived that after the randomized gateway P, the detection probability in the alert path is given by $\langle H|P\rho P^+|H\rangle$. This alert probability cannot be zero, unless Eve knows P. Not knowing about P, it can be shown that the probability of an alert is $\geq 25\%$ (given by averaging over the continuum of randomization states P) with a minimum of 25% probability given if Eve sends pure states. This assumes that Bob's SPADs

This work is supported by Academy of Scientific Research and Technology (ASRT), JESOR program (ID 5283), Egypt.

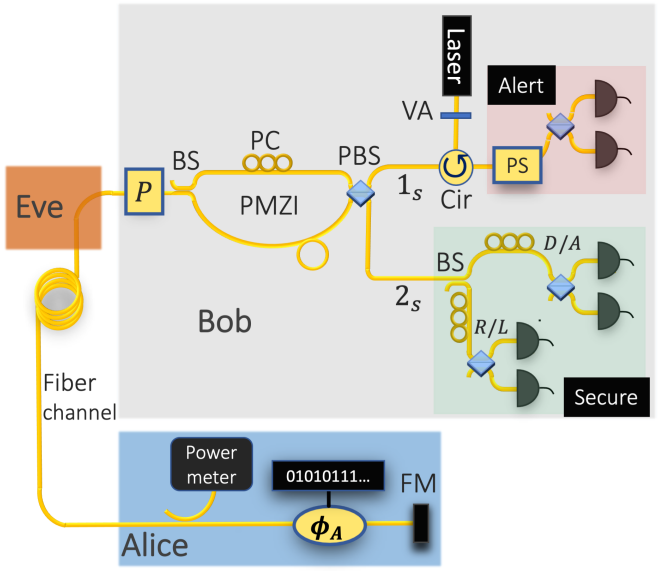

Fig. 1. Schematic of the QKD system [2]. PBS: polarization beam splitter; BS: beam splitter; PMZI: polarizing Mach-Zehnder interferometer; PS: polarization switch; PC: polarization controller; VA: variable attenuator; Cir: optical circulator.

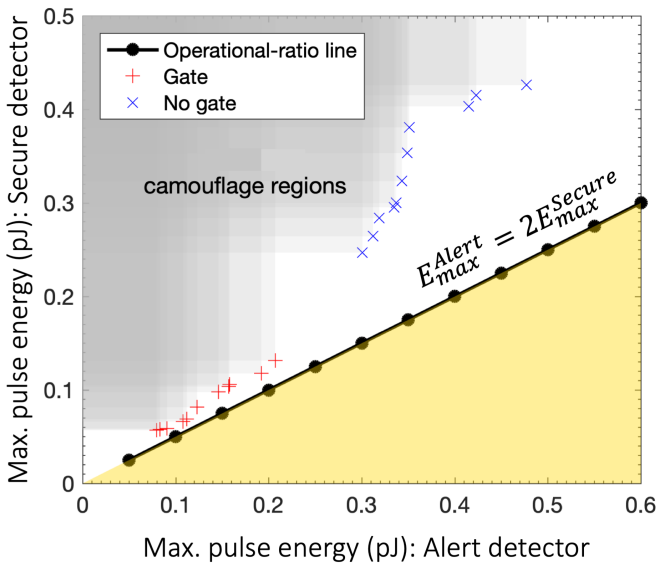

Fig. 2. The operational-ratio $E_{max}^{Alert} = 2E_{max}^{secure}$ dictates that, over all random settings of P, the maximum trigger pulse energy that may strike an alert SPAD is double that for a secure SPAD. The operation of Bob's receiver is restricted to the operational-ratio line as long as Eve does not know about the randomization P. The markers are intersection points of E_{never} thresholds for the alert SPAD (vertical threshold) and secure SPAD (horizontal threshold). These thresholds are experimentally measured in Ref. [4] at different blinding powers for the two SPADs of commercial QKD system Clavis2 (ID Quantique) in the presence (+) and the absence (×) of the SPAD gate. The created camouflage region (grey area) for each intersection point defines Eve's unnoticeable operation space.

are in Geiger mode, and Eve sends single photons. If Eve sends bright control pulses to trigger one of the SPADs or to blind all but one of them, the alert probability can be deduced accordingly based on the specific attack.

Eve might also send trigger pulses in the presence of a blinding light which turns Bob's SPADs to the linear mode. In this mode, the SPAD can be ticked only if the trigger pulse energy is greater than a threshold E_{never}. Below this threshold, the triggering probability is zero. The arrangement of Bob's system dictates that the triggering pulse energy and the blinding light power received by a SPAD in the alert-path will on average be twice that of the secure-path SPAD.

To launch an unnoticeable detector-control attack, Eve's trigger pulse energy would have to be less than E_{never} of the alert SPAD in order to avoid triggering an alert. It should also be greater than E_{never} of the secure SPADs to be able to *remote-control* them. This defines a camouflage region for Eve's successful attack in Fig. 2 bordered by these two crossing thresholds. In Fig. 2, we consider two SPADs included in the Clavis2 system (by ID Quantique). To disable Eve's attack, the SPADs of higher sensitivity are a assigned to the alert path. The higher sensitivity in the linear mode is determined by the lower profile of the threshold E_{never} as function of the blinding power. The threshold data are extracted from a reported experiment by Huang et al. [4] for several levels of blinding powers.

The portions of the trigger pulse energy that may reach Bob's SPADs on each path are randomized by the transformation P. Over all random settings of P, the maximum trigger pulse energy reaching the alert SPAD is double that for the secure SPAD. The doubling of the maximum trigger pulse energy (and also the minimum blinding power) is introduced by a beam splitter (BS) in path 2_s. This gives the operational-ratio line $E_{max}^{Alert} = 2E_{max}^{secure}$; defining the ratio between the ceilings of pulse energy portions received by secure and alert SPADs –as constrained by the system.

The impossibility of launching an unnoticeable attack is then verified in Fig. 2 by the non-crossing of the operational-ratio line with any of Eve's camouflage regions [3]. It follows that by assigning SPADs of higher sensitivity to the alert path 1_s, the operational-ratio line $E_{max}^{Alert} = 2E_{max}^{secure}$ does not cross any camouflage region, which disables Eve's faked state attack, no matter what faked state of light she uses.

References

[1] F. Xu, X. Ma, Q. Zhang, K. H. Lo, and J. W. Pan, "Secure quantum key distribution with realistic devices," Rev. Mod. Phys., **92**, 025002 (2020). DOI: https://doi.org/10.1103/RevModPhys.92.025002

[2] S. F. Hegazy, S. S. A. Obayya, B. E. A. Saleh, "Quantum key distribution system to overcome intercept-resend and detector-control quantum hacking," US Patent application 63/296,711.

[3] S. F. Hegazy, S. S. A. Obayya, B. E. A. Saleh, "Randomized ancillary qubit overcomes detector-control and intercept-resend hacking of quantum key distribution," Journal of Lightwave Technology, **40**, 6500 (2022). DOI: https://doi.org/10.1109/JLT.2022.3198108

[4] A. Huang, S. Sajeed, P. Chaiwongkhot, M. Soucarros, M. Legré, and V. Makarov, "Testing random-detector-efficiency countermeasure in a commercial system reveals a breakable unrealistic assumption," IEEE J. Quantum Elect., **52**, 8000211 (2016). DOI: https://doi.org/10.1109/JQE.2016.2611443