Revise and Resubmit

The journal has requested that this manuscript be revised and resubmitted.

Submit a revised manuscript (/submissions/quantum-journal/new?previous_version_id=1716759&proxy_submission=false)

Publication Decision from *Quantum*

"Countering detector manipulation attacks in quantum communication through detector self-testing"

Decision made on October 20th, 2022

Editorial board's determination

Revise and resubmit

Comments from the editor

Dear Christian,

We have received the referee reports for your paper “Countering detector manipulation attacks in quantum communication through detector self-testing”. You may find them enclosed.

Based on their recommendation, we are likely to accept the paper only after you carefully **revise and resubmit** your manuscript, addressing all the points raised by the two referees.

Of special importance are:

- additional details for an example protocol (such as the BB84 as requested by Ref 1)
- clarification of the two technical points raised as contradictions by Ref 2

Please upload the **revised version** and a **response letter file** to the points raised by the referees through Scholastica. It is in your own interest to ensure that the Referees and Editor can easily **retrace your revisions** (e.g., by attaching a detailed list of changes or a version of the manuscript with the changes highlighted). There is no need to update the arXiv version at this stage. If you have **confidential information** for the editor, please upload it as a **separate file**.

Best regards,
Krishnakumar

Krishnakumar Sabapathy
Editor at Quantum

The editorial policy (http://quantum-journal.org/editorial-policies/) and code of conduct (https://quantum-journal.org/about/terms-and-conditions/#coc) apply. By submitting your work to Quantum you have agreed to its terms and conditions (http://quantum-journal.org/terms-and-conditions/).

Reviewer 1

Open response questions

Summary: what are the main questions posed by the manuscript and how does it answer them?

Quantum Key Distribution (QKD) is a quantum technology concept trying to guarantee the availability of secret keys between communicating parties in networks.
While the original idea is very clear and accurate, implementations suffer from two problems, where the first is a model deviation and the second are implementation inaccuracies.
Model deviation from the original idea occurs for example when loss is considered. They can be discussed on a theoretical level.
Implementation inaccuracies are to some degree also model deviations, where the device implementing a certain functionality brings with it an environment that has not been accounted for in the theoretical analysis prior to the development.
These inaccuracies can be exploited by operating on these additional degrees of freedom. In this category are detector blinding attacks.
The severity of such inaccuracies is very high. Their discovery can stop research funding and deployment.

The idea proposed by the authors is to prevent detector blinding via self-testing.

What is your assessment of the paper? If you recommend acceptance, make a case that this work does indeed make a significant contribution to scholarship.

The idea to prevent detector blinding via self-testing is really very simple, and such simplicity is sought for to solve implementation problems.
While not all problems may be solved by this approach, it is nontheless pathleading. Therefore I highly recommend this work should be published given some more accurate protocol description was added.

To what extent have you checked the technical correctness of the paper?

To arrive at my statement, I have only considered the idea as such and *not* the particular implementation details.
Unfortunately, I can also not judge whether the experiments could be reproduced.

Comment on the presentation of the paper. Is it well written? Are the main results clearly laid out? Does the manuscript clearly describe assumptions and limitations? Is the literature review adequate?

The paper is well written. I argue that in order to be accepted it should include a theoretical analysis on a simple protocol level.
This analysis should cover the prototypical BB84 protocol as well as the particular implementations described in the manuscript.
It can be kept simple - in a sense, only the most necessary degrees of freedom need to be added to the protocol description and the detector blinding attack needs to be written as a finite-dimensional quantum channel which can be used by the eavesdropper to replace the identity map.

If the submission includes numerical or physical experiments, does it provide sufficient details such that they could be reproduced by readers? This includes for example source code, documentation, experimental data, experimental setup specifications, etc.

I regard the presented information as sufficient, but am not an expert on the particular implementation details. Therefore I cannot answer this question.

Suggested changes, corrections, and general comments.

Modelling inaccuracy is a wide-spread error in security system design. I argue that in order to be accepted the manuscript should include a theoretical analysis on a simple protocol level, to increase accuracy of corresponding scientific debate.
This analysis should cover the prototypical BB84 protocol as well as the particular implementations described in the manuscript.
It can be kept simple - in a sense, only the most necessary degrees of freedom need to be added to the protocol description and the detector blinding attack needs to be written as a finite-dimensional quantum channel which can be used by the eavesdropper to replace the identity map. A complete model of all system components including their respective environmental system is clearly not realistic to ask for. However it might be nice if the authors could point out possibilities where possible.

Reviewer 2

Open response questions

Summary: what are the main questions posed by the manuscript and how does it answer them?

See below.

What is your assessment of the paper? If you recommend acceptance, make a case that this work does indeed make a significant contribution to scholarship.

-

To what extent have you checked the technical correctness of the paper?

-

Comment on the presentation of the paper. Is it well written? Are the main results clearly laid out? Does the manuscript clearly describe assumptions and limitations? Is the literature review adequate?

-

If the submission includes numerical or physical experiments, does it provide sufficient details such that they could be reproduced by readers? This includes for example source code, documentation, experimental data, experimental setup specifications, etc.

-

Suggested changes, corrections, and general comments.

The authors claim to have proposed a universal approach that tests the single photon detector against manipulation as a black box. Meanwhile, there are a number of detector types and manipulation attacks. The first contradiction appears in the hypothesis:
"When the single photon detector is under a blinding attack, it is insensitive to low intensity light fields used for quantum key distribution." This does

not apply to the aftergate attack, which is one of the well-known detector manipulation attack.

The second problem is the test of the blinding approach, which claims, “However, any positive detector manipulation will overrule the local blinding, and cause a false detection event.”

In QKD certification, Eve is assumed to know all parameters about the setup of Alice and Bob, including the amplitude of the test blinding field. When we consider a blinding attack, the detector enters the linear range due to the blinding, but if we apply too much energy, it becomes saturated. In this case, Eve can set the blinding field so that the Eve blinding+manipulation signal triggers a click, but the Eve blinding+test blinding+manipulation signal is in the saturation region and cannot pass the comparator in the detector. This means that this countermeasure cannot be considered reliable.

In the meantime, I have heard many discussions about self-tests of the single photon response. Therefore, careful consideration of this problem is important. I propose to reject this paper, revise it, and resubmit it.

The new paper should include a brief overview of detector manipulation attacks and detector types. Then I recommend that the authors consider what types of attacks and detectors they want to address, and finally establish quantitative criteria for additional quantum key distribution privacy amplification due to the measured results.