Response Letter for manuscript #32012 20 June 2020

Dear Editor,

We went through the reviewers' comments and found them to be very helpful. As such, we have modified the manuscript in accordance with the comments. We hope that it satisfies the reviewer's concerns with the improved flow, readability, and the applicability of the paper.

In the revised version, we improved the figure and the explanation of the BB84 protocol, and we added a caption to guide readers through the figure (Reviewer 2 comments). We also removed a lot of the technical discussions in section 3 and put it on the footnote (Reviewer 3 comments). We have also acted on the more specific comments raised by all the reviewers - refer to the point-by-point responses in the attachment below.

We address the claim by Reviewer 2 about the originality of the idea in the reply to Reviewer 2 below. Although Ref 15 performed a large fraction of the key distribution procedure, we proceeded with the key sifting portion, and used the sifted key to encode and distribute messages. Furthermore, we provide Eve with the tools to intercept and reconstruct the intercepted message, a technique not demonstrated in Ref 15. We believe that our approach helps the participants to experience the practical consequences of distributing the key with macroscopic beams, instead of using single-photons, thus helping the participants remember the lesson better. Subsequently, we demonstrated this technique with student participants, inviting them to rate its efficacy later.

We have also reduced the length of our paper to keep within the editorial guidelines (< 5000 words, excluding acknowledgement, references and footnotes), with technical details relegated to footnotes. However, as our paper examines our demonstration method from several angles: theory, experimental implementation, pedagogy, teaching objectives, workshop conduct, and participant response with accompanying diagrams and pictures, we were unable to reduce the content to half its original length as suggested by Reviewer 3.

Thank you very much.

Best regards,
Adrian N. Utama
Jianwei Lee
Mathias A. Seidler

Replies to Reviewer 1:

1. **“It is a very remarkable effort, well documented, with online resources made available. I do recommend it for publication.”**
   We thank Reviewer 1 for the feedback.
2. **Point (a) “... BB84 could be made secure with laser sources, provided they are attenuated well below 1 photon/coherence time, which requires operating photodiodes in the photon counting regime. It would be good to state this very clearly, as most commercial implementations of QKD use lasers: you don't want the students to get away with the idea that all those companies are cheating.”**
   We thank the reviewer for the observation. We modified the fourth paragraph to include a discussion regarding the use of laser sources and their security, along with comments on the commercial implementation of QKD (underlined below):

   The security of BB84 relies on the fact that a single quantum bit (qubit) cannot be copied. When multiple qubits of the same state, e.g. multiple photons with the same polarization, are distributed, security is compromised since a fraction of the qubits can be intercepted and measured by an adversary (a form of side-channel attack [Ref 21]). In this workshop, we demonstrate this attack by using macroscopic laser intensities consisting of millions of photonic qubits per coherence time, creating an exploitable security loophole. This security loophole can be addressed by attenuating macroscopic sources to a mean photon number well below one per coherence time [Ref 22]. In commercial implementations, more sophisticated decoy-state protocols have been adopted to allow higher photon numbers per pulse – useful for transmitting over long distances [Ref 23]. However, we intentionally leave this loophole open to allow the students to revisit the ‘no-cloning’ theorem and its role in quantum cryptography – we task one group of students to implement a side-channel attack, retrieve the key, and decode secret messages.
3. **Point (b) “The references for device-independent QKD are not well chosen… "**
   We thank the reviewer for pointing out better alternatives. We have followed his advice and included Phys. Rev. Lett. 98, 230501 (2007) and SIAM Journal on Computing 48.1 (2019): 181-225.
4. **Point (c) “ Maybe the authors could suggest existing resources for upgrading: serious YouTube videos, books aimed at the same audience…”**
   We add a recommendation to the book “Six Quantum Pieces” in section 4, which we also distributed to the students during the workshop.

Replies to Reviewer 2:

1. **“As a whole, the paper is extremely unclear, the explanations are confusing and obscure and does not follow an adequate didactic sequence.”**
   We thank Reviewer 2 for his/her views on the writing style of our paper. We aimed at a brief description of the protocol with the intention of focusing more on our unique implementation and the learning experience afforded by a practical hands-on experience. We have also taken steps to improve the clarity of our paper as described in subsequent points.

2. **"... the figures are over-polluted and too technical, do not respect any standardisation, and explanations on both captions and on the main text are missing in many of them."**
   We improved the clarity of some figures. We improved Figure 1 and its caption, so that the reader can more easily appreciate the protocol sequence. We also simplified the caption in Figure 2 to improve readability. We retained the level of detail for Figures 4 and 6 in the interest of readers who might wish to recreate the setup for themselves.
3. **".. the idea is not original in its essence, since a very similar paper has already been published in Reference 15 ...."**
   We appreciate Reviewer 2 for highlighting that the essence of our paper could be perceived as being similar to Ref 15 (Ref 16 in the revised version). Although both our work uses macroscopic light to improve the accessibility of the setup, however, Ref 16 focuses on the underlying physics, the no-cloning theorem, and its use for quantifiably identifying an intercept-resend attack. Whereas our work focuses on the full deployment of the BB84 protocol which includes applying the generated key for transmitting an encrypted message - the consequences of the no-cloning theorem is demonstrated in the apparent failure of the protocol when Eve successfully deciphers the intercepted message. In other words, we provide students a real-world implementation of cryptography and its failings. To highlight the difference in our contributions, we changed on page 13:

   > *However, by incorporating Eve's hacking attempt, our approach provides students with the opportunity to investigate the consequences of using classical resources instead of quantum resources for key distribution.*

   to

   > *While the effects on the transmitted "qubits" during an intercept-resend attack has been explored [Ref 16], we focus instead on providing the students with the hands-on experience of decrypting the intercepted photons. This we think, provides a visceral experience for the students who witness the retrieval of secret transmission using a purportedly quantum-secure protocol executed with classical resources.*
4. **Point 1 "I have noticed some specific expressions which, to my knowledge, should not be assumed to be familiar in advance."**
   We improved on some expressions:
   a. Quarter-wave plate: We explain its working principle in a footnote, as it is not really important for the readers to understand the mechanism of a quarter-wave plate in the main text.
   b. Mutually-unbiased bases: We give the mathematical definition right after the term **(page 5)**.
   c. One time pad: We provide the definition right after the term, along with a reference **(page 11)**.
   d. DI QKD protocol: A description of DI QKD is beyond the scope of the paper. However, a reference is provided.
   e. Symmetric key: We give examples of distributed symmetric keys in paragraph 1.

f. Beam splitter: We describe its role in the setup in Section 3.D. and in the caption of Figure 2.
g. XOR: We point out that it is a bitwise exclusive or operation in the main text **(page 11)**.

5. **Point 2 "... there should be added that the security of the protocol is based both on the no cloning theorem and on the capability of identifying the presence of an eavesdropper."**
   We thank Reviewer 2 for his comment. However, we would like to focus on the underlying physical principle, as the capability of identifying the presence of an eavesdropper is derived from the no-cloning theorem; the theorem describes how Eve is prohibited from intercepting and resending the same quantum state. If she attempts to, she inadvertently introduces errors in the quantum channel, revealing her presence.
6. **Point 2 "... there should be mentioned some of the companies which today make quantum cryptography products available for the market."**
   In the same spirit of focusing on the underlying physical principle, we also do not list companies that manufacture QKD systems. Moreover, this list will change over time as the commercial realizations of the technology are still in their infancy.
7. **Point 3 "By the end of the second paragraph of the Introduction, one can find the pair of sentences … The second sentence may induce one to think that encryption belongs to the protocol and, in doing so, also uses quantum resources."**
   Indeed the wording was vague. We have modified the second sentence to:

   > *However, a full demonstration that involves both the use of QKD to generate the encryption key, and subsequently applying the key for encrypting a secret message, will help learners to appreciate how both relatively non-trivial procedures are integrated to realize a fully-functional quantum cryptography system.*

   We hope that our goal, of going beyond just demonstrating the underlying physical principle of QKD, to actually using the key generated from QKD to encrypt messages, is clearer now.
8. **Point 4 "The usage of infrared pulses, instead of visible light, for the classical channel should be justified."**
   We added the following sentence on **(page 8):**

   > *We chose to establish the classical channel with infrared LEDs due to their wide-spread availability in many consumer-electronic devices -- a hands-on experience assembling and operating the electronic circuits equips students to <u>embark</u> on their own projects after the workshop.*

9. **Point 5 "The BB84 protocol's description on Section II is not as clear ... Fig.1 should have been explained in details in the main text ... A simple title in the caption is not of much use."**
   We improved the consistencies of the notations. We improved Figure 1 and expanded the caption to include a brief description of the protocol (also refer to point 2).
10. **Point 6 "... when the basis does not match, one measures medium values for the intensity. How is this case categorized? How is the threshold established given**

the presence of noise to really simulate measures at random when bases do not match? …"

We expanded the statement on **(page 7)** to

> *For each "qubit'", the light intensity measured after the polarizer is compared with a predetermined threshold set at half the expected maximum light intensity. When Alice and Bob transmit and receive using the same basis, Bob observes two distinct intensity classes: high and low, allowing his measurements to be encoded as a binary number. Measuring the entire sequence generates a binary string B. Intensities measured when both parties use different bases are discarded in subsequent steps.*

We hope that the above statement clarifies that when the basis does not match, the categorization results are discarded during the subsequent key sifting procedure, and such results are therefore inconsequential.

11. **Point 6 "... I suggest the last sentence to be altered to read "Second, to learn about the state transmitted to Bob, Eve would have to perform a measurement which inadvertently disturbs the state *and may reveal herself*'. It is worth to mention that in the original BB84 quantum protocol a fraction of qubits are publicly compared for disagreement (and though not used for the secret key) in order for them to check the presence of an eavesdropper."**

We thank the reviewer for this observation and have changed

> *First, a single photon is a quantum state which cannot be cloned. . Second, to learn about the state transmitted to Bob, Eve would have to perform a measurement which inadvertently disturbs the state.*

to

> *First, to learn about the state transmitted from Alice, Eve would have to intercept the photon, and perform a measurement which inadvertently disturbs its state. Second, if Eve resends the photon she measured to Bob, she cannot create a perfect copy due to the quantum no-cloning theorem. Consequently, Alice and Bob will be able to identify the presence of Eve by checking for inconsistencies for a subset of the transmitted bits that were prepared and measured in the same basis. The rest of the bits are then used to generate the final encryption key.*

We included a clarification in Section 3.B and footnote to highlight where the public comparison of qubits may be implemented in future setups.

> *Typically, a subset of K is checked for inconsistencies to reveal the presence of adversary Eve, when she performs an intercept-resend attack. As our implementation of Eve does not resend "qubits" to Bob, we omitted this step to simplify the setup [footnote].*
>
> *Footnote: We aim to perform the key comparison step in future iterations of the setup, as it allows students to verify that their key was not corrupted by Eve using an intercept-resend attack. In our workshop, Eve intercepts, but does not resend "qubits", rendering her eavesdropping attempt immune to detection by the key comparison step. Consequently, the impact of her success will be greater on the students who may have considered the key to be private given that they have*

performed key comparison but might have forgotten about the role that single quanta plays in the security of the protocol.

12. **Point 7 “Figure 5 is not referenced in any part of the main text.” and Point 9 “Figure 9 is also not referenced in any part of the main text.”**
    We thank the reviewer for pointing this out. We now provide reference to Figures 5 and 9 in the main text.
13. **Point 8 “Figure 8 and its corresponding explanation text are incomprehensible for those who do not know the “K-means clustering algorithm".”**
    We thank the reviewer for his comments for improving the clarity of our explanation for our clustering technique used to identify the state of the intercepted photon. We have since included an explanatory statement for the k-means algorithm:

    > The algorithm *iteratively computes the positions of the four groups so as to minimize the least-squared distance between each data point and the mean position of its assigned group.*

    and included a reference for interested readers.
14. **Point 8 “... which are photodiodes 1 and 2? In Fig 2, since at least one of the photodiodes on Eve’s apparatus is right after a beam splitter, it must measure a constant intensity from the incoming beam. How can one measure different intensity values as suggested by Fig. 8?”**
    We have identified photodiodes 1 and 2 in the caption by including the following statement:

    > *Four identified clusters of signal voltages measured by Eve’s dual photodiodes 1 and 2 that measure light intensity after projecting the incoming polarization in two different basis (see Figure 2)*

    We thank the reviewer for correctly pointing out the missing rotatable polarizer (RP) after Eve’s BS, which we have inserted in Fig 2.
15. **Point 8 “[in Figure 8] What do mean the different colours? What does Eve extract from the graph of Figure 8 and why does she need such a brutal force method to derive the key?**
    The different colors in Fig 8 correspond to the four distinct polarizations that are being intercepted by Eve. The colors provide a visual aid to distinguish the different marginal distributions plotted at the edge of the figure. This allows readers to appreciate that the ability to distinguish the four distinct distributions is poor when measuring only with Photodiode 1, whereas measuring with the additional Photodiode 2 increases distinguishability. We describe in the main text with the statement:

    > *As Eve’s basis choice is a priori not aligned to Alice and Bob’s, she may not be able to distinguish between polarization states optimally. However, by measuring in more than one basis choice simultaneously, she improves her ability to identify distinct polarization states even in the presence of laser intensity noise.”*

    When Eve is able to distinguish the four groups, she is able to permute the polarization assignment to the four categories, and derive 6 possible keys. These possibilities include the key capable of decoding encrypted messages, transmitted in the classical channel, into a legible message.

Replies to Reviewer 3:

1. **“However, it seems to me that, considering the subject matter, the paper as it stands is too long and includes too many technical details.”**
   We agree with Reviewer 3 that our paper includes several technical details. However, we think that these details are helpful to educators who wish to implement the demonstration themselves. Nevertheless, to improve readability, we have, after including further details requested by the reviewers, also reduced its length from 5100 words (excluding acknowledgement, references and footnotes), to within the editorial guidelines (< 5000 words), relegating technicalities to footnotes [Ref 25-35] that may detract from the presentation of the main material.
2. **“... the authors of this new paper are using the standard BB84 polarizations and the standard BB84 measurements. Anyone familiar with the BB84 protocol would be able to imagine something like the authors’ set-up ... It does not seem appropriate to me to use precious pages of AJP to explain the technical details of the apparatus.”**
   We thank the reviewer for pointing out that a physicist familiar with the BB84 will be able to imagine our setup. However, we would like to keep within the AJP editorial policy which states that

   > *“Such manuscripts should not be review articles, but rather self-contained articles that describe a particular piece of research in such a way that it is accessible to as many physicists as possible.”*

   Consequently, keeping in mind that not all physics educators interested in exposing their students to quantum cryptography would be *a priori* acquainted with BB84, we decided to include sufficient details of the original protocol, and its implementation, to allow our article to be self-contained.

   The level of technical detail, as explained in our response to the previous point, is necessary to enable the interested educator to fully implement the setup. We did this bearing in mind that not all educators would *a priori* be sufficiently proficient with deploying custom electronics, and other optomechanics, for realizing a fully functional cryptography demonstration. This is particularly true for educators who may not have the benefit of being trained as an experimental physicist. Even if a reader is familiar with a particular implementation of BB84, our setup includes a specific realization of the classical channel with IR devices, and our unique approach to hacking the quantum channel. These parts of the setup were designed to be understood and assembled with widely-available components, and at the skill-level appropriate to our target audience. Consequently, we believe that a sufficiently detailed description is helpful for readers wishing to recreate the same learning experience. We have, in fact, excluded tedious details from the paper and relegated them onto a Github repository.
3. **“... Of course the BB84 protocol needs to be explained briefly, and the basic scheme of the apparatus needs to be presented (also briefly), the primary focus should be the *pedagogical method* and the students’ experiences: ...”**

While we understand that brevity in describing the implementation can be useful for an audience already familiar with BB84, we decided to explain in sufficient detail our specific implementation, which was designed to increase the likelihood that pre-university students were able to understand and assemble the setup on their own - the use of easily operable Arduino microcontroller, the adoption of the ubiquitous IR transceivers for the classical channel, etc. In addition, all the components used are easily sourced, commercially available, and relatively cheap. The careful selection of components, and their implementation served the pedagogical purpose: to ensure that as many educators were able to provide their students an experience in building a functioning BB84 demonstration with a relatively low cost and knowledge barrier.

4. **Point 1 "I think the authors should point out that no physical principle prevents the eavesdropper from participating in this calibration procedure."**
   We have added a footnote on the sentence:

   As Eve's basis choice is a priori not aligned to Alice and Bob's, she may not be able to distinguish between polarization states optimally [footnote].

   Footnote: Eve could have performed the polarization calibration procedure along with Alice and Bob, which allows her to select the best polarization measurement for distinguishing and identifying the intercepted polarization states. However, given the limited time for the workshop, it was difficult to prepare the students in team Eve in time for the calibration procedure as they would then have to learn calibration, principles of the quantum channel, using two polarization measurements to extract the polarization from the intercepted photons, and the clustering algorithm - the latter two concepts that Alice and Bob did not have to learn.
5. **Point 2 " Perhaps the authors could instead use "simulated qubit," or could explain clearly that they will be using the word "qubit" in a loose sense, to refer to a simulated qubit."**
   We used "qubit" with the quotation marks. We added the definition of ``qubit'' at the end of Section 2.
6. **Point 3 "In citing the no-cloning theorem, along with the Wootters-Zurek paper, I would recommend also citing D. Dieks, "Communication by EPR Devices," Phys. Lett. A 92, 271 (1982), which presents essentially the same no-cloning argument."**
   We added the reference, thanks for the suggestion!

List of other changes to improve flow and readability:

1. The second author changed his name presentation to Jianwei Lee.
2. We clarify the focus of other QKD demonstrations in paragraph 2 of introduction, from:

   There now exist several QKD demonstrations for non-experts which focus on the key distribution step of the protocol [Ref 15-20].

   to

   There now exist several QKD demonstrations for non-experts which focus on teaching the underlying physics of QKD [Ref 15-20].

3. We highlighted more explicitly the relation between qubits and photons, and defined the side-channel attack and provided references in paragraph 4 of the introduction (additional content is underlined):

   When multiple qubits of the same state, e.g. multiple photons with the same polarization, are distributed, security is compromised since a fraction of the qubits can be intercepted and measured by an adversary (a form of side-channel attack [Ref 21]).
4. In Section 2, we change all the variables notations, i.e. *A, B, X, Y*, to capital letters to avoid confusion with the notations in Figure 1.
5. We added a paragraph and footnote to elaborate on our implementation of the key comparison step in Section 3.B.:

   Typically, a subset of K is checked for inconsistencies to reveal the presence of Eve when she performs an intercept-resend attack. However, we chose to omit this step as our implementation of Eve does not perform this attack. [footnote]

   Footnote: We aim to perform the key comparison step in future iterations of the setup, as it allows students to verify that their key was not corrupted by Eve, prior to sending their secret message. When Eve performs the side-channel attack, the impact of her success will be greater on the students who may have considered the key to be secure given that they have performed key comparison but might have forgotten about the role that single quanta plays in the security of the protocol.
6. We moved the sentence "*The spatial mode of the IR LED has a relatively large solid angle*." from the main text in Section 3.D. to the caption of Figure 7.
7. We combined paragraphs 2 and 3 in Section 4.A., expanding on the hands-on approach, and deferred the discussion about "making abstract concepts concrete" to Section 4.B., which addresses our learning objectives (see also point no. 8):

   Another powerful motivation for simplifying the setup is to allow students the opportunity to build a working QKD system from the ground up -- a hands-on approach generally increases students' engagement compared to didactic methods [Ref 39]. A simple setup could be built within the time-constraints expected for an extra-curricular activity. Furthermore, our approach allows students to demonstrate competence, which is an essential intrinsic motivating factor in any learning task [Ref 40].
8. We elaborated on "more concrete understanding of the cryptography protocol" in the learning objective in Section 4.B:

   We use these processes as pedagogical tools to help students understand the cryptography protocol more concretely. Two examples are as follows: (i) establishing a common polarization basis between Alice and Bob allows students to physically implement qubits in the polarization degree-of-freedom, (ii) implementing the classical channel with IR pulses provides, for most students, a first encounter of how pulse sequences are used to transmit information.
9. We added the following sentence at the end of Section 4.B: "*Facilitators may also take this opportunity to explore other ingenious hacks that exploit the vulnerabilities of practical QKD systems. [Ref 45]* ". Ref 45 elaborates several important, yet often overlooked security loopholes that surface during practical QKD implementation.

10. We modified the presentation of the student's feedback (Section 4.D.) to reduce its length, transferring most of the details into Table 1.
11. In Appendix A, we relegate the technical discussion about the key "leaks" and brute force attack to the footnotes (Ref. 49 and 53) to improve readability.