Practical Security of Semi-Quantum Key Distribution*

Walter O. Krawec
University of Connecticut
walter.krawec@uconn.edu

Abstract

Unconditionally secure key distribution is impossible using classical communication only. However, by providing Alice and Bob with quantum capable hardware the task becomes possible. How quantum does a protocol need to be, though, in order to gain this advantage? In 2007, Boyer et al., proposed "semi-quantum key distribution" where only Alice need be quantum while Bob need only limited "classical" capabilities. Several protocols were proposed and proven secure in the "perfect qubit scenario" but not necessarily against realistic attacks (with one exception being recently published in (PRA 96 062335)). In this paper, we devise a new SQKD protocol and analyze its security against certain practical attacks.

1 INTRODUCTION

Quantum key distribution (QKD) allows for the establishment, between two users Alice (A) and Bob (B), of a shared secret key the security of which is guaranteed even when faced with an all-powerful adversary Eve (E). Such a task is impossible using classical communication alone (where security of key distribution always necessarily depends on unproven computational assumptions being placed on the adversary). However by careful use of a quantum communication channel, and an authenticated (but not secret) classical channel, this task is provably secure in a variety of security models. The reader is referred to [1] for a general survey of QKD protocols and their security proofs.

Recently, in [2], a new class of QKD protocol was proposed whereby severe restrictions on the quantum capabilities of the user B are placed. In fact, B may only operate in a "classical" manner by either completely ignoring any incoming quantum signal from A, or he may only measure and send qubits in a single, fixed, publicly known basis (generally the computational Z basis $\{|0\rangle, |1\rangle\}$). Protocols operating in this model are known as *semi-quantum* key distribution (SQKD) protocols.

In more detail, such protocols operate over a two-way quantum communication channel, allowing A (who has no resource restrictions placed on her) to send qubits to B. These

*This paper is a pre-print of one published in Proc. SPIE Defense 2018

qubits may be prepared in arbitrary ways. B, the semi-quantum or "classical" user has a choice to either:

1. `Measure and Resend`: If B chooses this operation, he will subject the incoming qubit to a computational Z basis measurement. He is allowed to output a Z basis qubit.

2. `Reflect`: If B chooses this operation, he will simply "disconnect" from the quantum channel and reflect any incoming state back to A without disturbing it (or learning anything about it).

Regardless of B's choice, a quantum state may arrive back at A who is free to perform any quantum operation on it (e.g., measure in an arbitrary basis).

Numerous SQKD protocols have been proposed since the model's initial introduction in 2007 (see, for instance,[3, 4, 5, 6, 7]). Some have been provided with complete information theoretic proofs of security [8, 9]. Most of these protocols, however, require the assumption that "perfect" qubits are traveling on the quantum channel. Indeed, if this assumption were not taken, then certain attacks such as the photon tagging attack described in [10, 11] are possible completely breaking the security of most SQKD protocols.

The situation, however, has improved recently. In [12], M. Boyer, M. Katz, R. Liss, and T. Mor have devised a new protocol based on the use of "mirrors." Importantly, this protocol did not require B to prepare fresh qubits. Security against practical (and theoretical) attacks was proven in terms of *robustness* - namely, any attack against the protocol which causes an adversary to potentially learn information on A and B's key (including attacks involving an adversary sending multiple qubits or vacuum states) can be detected with non-zero probability. This protocol is also immune from the photon-tagging attacks mentioned earlier.

In this work, we propose a different SQKD protocol also designed to counter these so-called "practical" attacks against it (e.g., multi-photon attacks, photon tagging, or photon losses). Our protocol is a generalization of one we first proposed in [13] which may be considered the semi-quantum version of Extended-B92 [14]. It's security *assuming perfect qubits* was recently proven in [15].

There are several contributions made in this work. We discuss a new protocol, modified from work we did initially in [13], adapted to work in more practical scenarios. The protocol we discuss in this paper does not require B to prepare fresh qubits. Furthermore, we design an appropriate mechanism to model its security and propose a novel potential optical implementation of the protocol. Finally, we use our security model to perform an information theoretic security analysis against certain practical attacks developing a framework from which future researchers in this area may benefit from.

Naturally, considering the security of any QKD protocol is a far more difficult problem when "real-world" implementations are used. This problem seems exacerbated in the case of semi-quantum protocols due to their reliance on a two-way quantum channel and B's inability to gather accurate statistics in all measurement bases. While this paper does not claim to solve all issues pertaining to the difficulties in implementing "practical" SQKD

protocols, we do attempt to address some of the more challenging problems, namely multi-photon attacks and tolerance to photon loss. A complete security analysis of this protocol remains an open problem. As does a comparison to the security properties of the mirror protocol introduced in [12].

1.1 Notation

We use $H(X)$ to represent the Shannon entropy of random variable X. By $h(x)$ we mean the binary entropy function $h(x) = -x\log(x) - (1-x)\log(1-x)$, where all logarithms in this paper are base two. $H(A|B)$ is the conditional Shannon entropy.

Let ρ_{AB} be a density operator acting on Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. Then we write ρ_A to mean the result of tracing out the B portion; i.e., $\rho_A = tr_B \rho_{AB}$. Given an element $|\psi\rangle$, we write $[\psi]$ to mean $|\psi\rangle\langle\psi|$. Similarly, for $i, j \in \mathbb{N}$, we write $[\mathbf{i}, \mathbf{j}]$ to mean $|i\rangle\langle i| \otimes |j\rangle\langle j|$.

If ρ_A is a density operator acting on Hilbert space $\mathcal{H}_A$, then we denote by $S(A)_\rho$ to mean the von Neumann entropy of ρ_A. We also write $S(A|B)_\rho$ to mean the conditional von Neumann entropy, namely $S(A|B)_\rho = S(AB)_\rho - S(B)_\rho = S(\rho_{AB}) - S(\rho_B)$. When the context is clear, we will forgo writing the subscript "ρ." Finally, we define $I(A:B)_\rho$ to be the quantum mutual information.

Given a bit string $q \in \{0,1\}^n$, we write $w(q)$ to be the Hamming weight of q, namely, the number of 1's in the bitstring q. We write 1^N to mean $11\cdots1$ (N times).

1.2 General QKD Security

A (S)QKD protocol begins with a *quantum communication stage* resulting in the distillation of a *raw key*. This is a string of N classical bits (one string for A and one for B) which are partially correlated, and partially secret. An error correcting protocol followed by privacy amplification results in a secret key of size $\ell(N)$. Assuming collective attacks (i.e., attacks whereby E treats each signal independently and identically, but is free to postpone measuring her ancilla to any future point in time and, furthermore, is free to perform any theoretically optimal, measurement of her ancilla) then the Devetak-Winter key-rate expression applies. Namely:

$$r := \lim_{N\to\infty} \frac{\ell(N)}{N} = S(B|E) - H(B|A)$$

While we consider the asymptotic scenario here, the computations we do in this paper to bound the von Neumann entropy $S(B|E)$ can also be used in the finite key setting using techniques from [16] (though, there, one must be careful of imperfect parameter estimation - in this paper we will assume our parameter estimates are arbitrarily accurate leaving this more complete analysis as future work).

Figure 1: Our security proof will assume the existence of a box as depicted here, where ρ_1 is a quantum input; ρ_2 is a quantum output; c_1 is a classical input; and c_2 is a classical output. How the box operates is described in the text.

2 THE PROTOCOL

The protocol is a generalization of one proposed by us first in [13]; the greatest difference between the two is that we do not require B to prepare fresh qubits. Instead, we assume B is equipped with a box, depicted in Figure 1, which has one quantum input ρ_1, one quantum output ρ_2, one classical input c_1, and one classical output c_2. This box is capable of implementing the following functionality:

1. If $c_1 = 0$ (i.e., `Reflect`), then $\rho_2 = \rho_1$ and $c_2 = 0$.

2. If $c_1 = 1$ (i.e., `Measure and Resend`), then with probability P_{NC}, it holds that $c_2 = 0$ and:

$$\rho_2 = \frac{1}{P_{NC}} \sum_{n \geq 0} q_n [\mathbf{1}]^{\otimes n}. \tag{1}$$

   Otherwise, with probability $p_c = 1 - P_{NC}$, it holds that $c_2 = 1$ and:

$$\rho_2 = \frac{1}{1 - P_{NC}} \sum_{n \geq 0} p_n [\mathbf{1}]^{\otimes n}. \tag{2}$$

   We do not require that $p_n = q_n$ only that they may be characterized if a known input to the box is given.

Notice that, if B chooses to `Measure and Resend`, the box will only output states of the form $|1\rangle^{\otimes n}$ (i.e., it may output multiple qubits, but each qubit is in the state $|1\rangle$). How the box is actually implemented is irrelevant to our security proof in this paper so long as the values of p_n and q_n can be characterized given a known input state. A possible optical implementation of such a box is shown in Figure 2. Note that if the detector does not click, it should be that if n photons were inputted into the box in ρ_1, there should be n photons leaving but all in a state of $|1\rangle$. Detector efficiency and dark counts will affect the various probabilities.

The protocol, then, operates as follows:

1. A prepares and sends a qubit in the state $|+\rangle$.

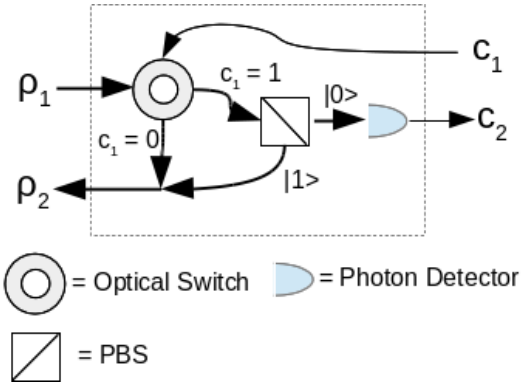

Figure 2: A potential implementation of the semi-quantum box for our protocol. While the security analysis does not require this particular implementation, our evaluations later will assume this particular implementation.

2. B will pick a random bit k_B to be his potential raw key bit for this iteration. He sets the input c_1 of his box to be this bit (i.e., $c_1 = k_B$). He will save the classical output of his box c_2 for use later.

3. A will measure the incoming qubit in either the Z basis or the X basis, choosing randomly. If she chooses to measure in the Z basis and observes a $|0\rangle$, then she will guess that B's key bit k_B is 0; otherwise, if she measures in the X basis and observes a $|-\rangle$, she will guess that B's value of k_B is 1. Other events are inconclusive.

4. If B's value of c_2 is 1, then he will inform A to discard the iteration. Likewise, if A's measurement was inconclusive, A will tell B to discard the iteration. This communication is done via the authenticated classical channel (and, of course, may be performed at the conclusion of the quantum communication stage). Thus, only when A receives a conclusive result and only when B's box outputs $c_2 = 0$ will the iteration contribute to the raw key.

It is easy to see the similarities to the B92 [17] protocol (in essence, if A's source is pure and the forward channel is not attacked, B is sending $|+\rangle$ or $|1\rangle$). Of course, the forward channel may be attacked, a likely scenario that we will need to consider in the security proof (and is not necessary when proving security of B92). This protocol is also a more general one than we proposed in [13]. Indeed, that protocol can be considered a particular instance of this one, where the "box" is such that the dimension of the input and output is 2 (i.e., $\dim \rho_1 = \dim \rho_2 = 2$; it is a perfect qubit), and where $p_n = 1$.

3 SECURITY ANALYSIS

The state leaving A's lab is publicly known. We assume it is produced by an attenuated laser pulse, thus the state leaving her lab is of the form:

$$\rho_A = \sum_{n\geq 0} a_n [+]^{\otimes n}, \tag{3}$$

where a_n is the probability her source emits an n photon pulse, namely:

$$a_n = e^{-\mu}\frac{\mu^n}{n!},$$

(i.e., a Poissonian distribution with mean photon number μ).

Eve will attack in the forward direction (i.e., when the signal travels from A to B). Since we assume she may count the number of photons in the pulse, she may adapt her strategy based on this count. Namely, she may apply an attack evolving the state in Equation 3 to the following:

$$\mathcal{E}\left(\sum_{n\geq 0} a_n [+]^{\otimes n}\right) = \sum_n q_n [\mathbf{n}]_R \otimes [\mathbf{E_n}],$$

where $[\mathbf{n}]_R$ is an internal register in E's memory storing the result of her photon counting attack and where $|E_n\rangle$ is the (without loss of generality pure) state leaving E's lab in this case. Due to the concavity of conditional entropy, the result of $S(B|E)$ following the protocol will be the average over all possible counts of n photons appearing in the above (i.e., the average of the case had E sent $|E_1\rangle$ or $|E_2\rangle$ and so on). Thus, it is to E's advantage to simply choose a *fixed* N that optimizes her entire attack. Therefore, we will assume that the state leaving E's lab after her first attack on the forward channel is simply an N qubit state of the form:

$$|e\rangle := |E_N\rangle = \sum_{x\in\{0,1\}^N} \alpha_x |x\rangle_T \otimes |e_x\rangle_E ,$$

where the T portion of the state will travel to B's box (on input ρ_1) and the E portion will remain private to E. This latter represents E's quantum memory which will play a part in her subsequent attack when the qubit returns. Note that if there is no adversary and if A's source is perfect (i.e., $p_1 = 1$), then $|e\rangle = |+\rangle$. Of course, E is free to prepare whatever she likes at this stage of the protocol.

The state $[\mathbf{e}]$ is input into Bob's box. In the event he chooses `Reflect`, leaving his lab will be $[\mathbf{e}]$ and he will save, in an internal register, that his key-bit is 0. If he chooses `Measure and Resend`, the state leaving B's lab is the mixture:

$$\rho_2 = [\mathbf{0}]_{c_2} \otimes \left(\sum_n q_n [\mathbf{1}]^{\otimes n}\right) + [\mathbf{1}]_{c_2} \otimes \left(\sum_n p_n [\mathbf{1}]^{\otimes n}\right),$$

as described in Equations 1 and 2 and where we introduce a new register to store the result of the box's classical output bit c_2 (which he saves). In this event, he will set his key-bit to

be 1 (though, will later discard the result if $c_2 = 1$). Ultimately, the joint system may, at this point, be modeled using the density operator:

$$\frac{1}{2}[\mathbf{0}]_B \otimes [\mathbf{e}]_T + \frac{1}{2}[\mathbf{1}]_B \otimes \left[[\mathbf{0}]_{c_2} \otimes \left(\sum_n q_n [\mathbf{1}]_T^{\otimes n} \otimes \sigma_n^E \right) + [\mathbf{1}]_{c_2} \otimes \left(\sum_n p_n [\mathbf{1}]_T^{\otimes n} \otimes \sigma_n'^E \right) \right] \quad (4)$$

It is only to E's advantage to assume that σ_n^E is pure. For our proposed implementation shown in Figure 2, σ_N will indeed be pure (it is not difficult to show that $\sigma_N = [\mathbf{e_{11\cdots 1}}]$) while the others may be mixed states. Thus, this assumption will decrease the overall key-rate compared to a real-world implementation.

At this point, E is free to attack again. Note that, again, E may perform a quantum non-demolition measurement of the photon number count. If $q_N = 1$, then nothing useful can be learned from this (as the number of photons leaving B's lab when his key-bit is 0 is identical to when his key-bit is 1 - i.e., N photons in either case). However, if q_N is smaller than 1, a photon count leaving B's lab less than N betrays his choice of operation. Indeed, it is clear that an attack exists causing Eve to gain full information on the key-bit whenever the photon number count is less than N (the number of photons entering B's box). With this in mind, it is clear that this box must be designed so that, when B chooses `Measure and Resend`, either the number of photons leaving his lab is equal to the number entering *or* the box outputs $c_2 = 1$ (in which case it doesn't matter as the users will with certainty, discard the iteration so there is nothing to learn anyway - note that E would still learn B's operation, but she cannot change the fact that he will later tell A to discard the iteration as this information is transported over an authenticated channel). In our proposed implementation, the value of q_N depends heavily upon the efficiency of the detector used. We will discuss this again momentarily. However, while our security proof does not depend on the actual implementation, the final key-rate result does indeed, depend on this value q_N.

We will assume that E's attack is chosen so that she only forwards one photon to A, while "absorbing" the additional qubits from B into her ancilla (allowing her to use them later to learn B's operation). We model the second attack operation as a unitary operator acting on the quantum channel, and E's private ancilla. Let U denote this attack; without loss of generality, we may write its action as:

$$\begin{aligned} U|E_N\rangle &= |+, f_0\rangle + |-, f_1\rangle + |v, f_v\rangle \\ U|1\rangle^{\otimes N} \otimes |\sigma_N^E\rangle &= |0, e_0\rangle + |1, e_1\rangle + |v, e_v\rangle \\ U|1\rangle^{\otimes n} \otimes |\sigma_n^E\rangle &= |0, g_0^n\rangle + |1, g_1^n\rangle + |v, g_v^n\rangle\,, \text{ for } n < N, \end{aligned} \quad (5)$$

where $|v\rangle$ is the "vacuum" state. Note that we abused notation above and wrote $|\sigma_n\rangle$; however, since we are assuming the σ_n appearing in Equation 4 are pure, such elements exist. We also abused notation by writing a single attack operator U. Technically, E will perform a non-demolition measurement of the photon number and apply an operator U_n based on this attack. Clearly, above, the only line that would change would be the third line (since the first two must be the same $U = U_N$).

Following this attack on the state shown in Equation 4, the (now single) qubit travels to A who performs a measurement leading to her "guess" of B's key. Finally, both A and B will inform one-another whether to discard or accept this iteration. The final state, conditioning on it not being discarded, is easily found to be:

$$\rho_{ABE} = \frac{1}{M}[\mathbf{00}]_{BA} \otimes [\mathbf{F}] + \frac{1}{M}[\mathbf{01}]_{BA} \otimes [\mathbf{f_1}] + \frac{q_N}{M}[\mathbf{10}]_{BA} \otimes [\mathbf{e_0}] + \frac{q_N}{M}[\mathbf{11}]_{BA} \otimes [\mathbf{E}] + \sum_{n<N} \frac{q_n}{M} \left([\mathbf{10}]_{BA} \otimes [\mathbf{g_0^n}] + [\mathbf{11}]_{BA} \otimes [\mathbf{G^n}] \right),$$

where we define:

$$|E\rangle = \frac{1}{\sqrt{2}}(|e_0\rangle - |e_1\rangle) \qquad |F\rangle = \frac{1}{\sqrt{2}}(|f_0\rangle + |f_1\rangle) \qquad |G^n\rangle = \frac{1}{\sqrt{2}}(|g_0^n\rangle - |g_1^n\rangle),$$

and M is a normalization term. In particular, let $\texttt{PKey}_{i,j}$ be the following (observable) probabilities:

$$\begin{aligned}
\texttt{PKey}_{0,0} &= \langle F|F\rangle = Pr(A \text{ observes } 0 \mid A \text{ chooses } Z \text{ and } B \text{ chooses } \texttt{Reflect}) \\
\texttt{PKey}_{0,1} &= \langle f_1|f_1\rangle = Pr(A \text{ observes } - \mid A \text{ chooses } X \text{ and } B \text{ chooses } \texttt{Reflect}) \\
\texttt{PKey}_{1,0} &= Pr(A \text{ observes } 0 \mid A \text{ chooses } Z \text{ and } B \text{ chooses } \texttt{Measure and Resend}) \\
&= \sum_n q_n Pr(A \text{ observes } 0 \mid A \text{ chooses } Z \text{ and } B \text{ chooses } \texttt{Measure and Resend} \text{ and } n \text{ photons leave} \\
\texttt{PKey}_{1,1} &= Pr(A \text{ observes } - \mid A \text{ chooses } X \text{ and } B \text{ chooses } \texttt{Measure and Resend}) \\
&= \sum_n q_n Pr(A \text{ observes } - \mid A \text{ chooses } X \text{ and } B \text{ chooses } \texttt{Measure and Resend} \text{ and } n \text{ photons lea}
\end{aligned}$$

Then, M is simply the sum $M = \sum_{i,j} \texttt{PKey}_{i,j}$. Note that from the above, $\langle F|F\rangle$ and $\langle f_1|f_1\rangle$ are directly observable.

We now break ρ_{ABE} into a "good" case and a "bad" case. The good case occurs when the number of qubits leaving B's box is equal to the number entering (which, if B's box is built with high efficiency devices, should occur with high probability). The bad case is the opposite and, at worst, provides to E full information.

Let $\widetilde{\texttt{PKey}}_{1,0} = \langle e_0|e_0\rangle$ and $\widetilde{\texttt{PKey}}_{1,1} = \langle E|E\rangle$. For $q_N < P_{NC}$, these quantities cannot be directly observed; however they can be estimated. We will comment further on this shortly, however this notation allows us to write ρ_{ABE} as follows:

$$\rho_{ABE} = \frac{p_G}{M} \left(\underbrace{\frac{[\mathbf{00}]_{BA} \otimes [\mathbf{F}] + [\mathbf{01}]_{BA} \otimes [\mathbf{f_1}] + q_N[\mathbf{10}]_{BA} \otimes [\mathbf{e_0}] + q_N[\mathbf{11}]_{BA} \otimes [\mathbf{E}]}{\texttt{PKey}_{0,0} + \texttt{PKey}_{0,1} + q_N\widetilde{\texttt{PKey}}_{1,0} + q_N\widetilde{\texttt{PKey}}1,1}}_{\sigma_{good}} \right) + \left(1 - \frac{p_G}{M}\right) \sigma_{bad},$$

where:

$$p_G = \texttt{PKey}_{0,0} + \texttt{PKey}_{0,1} + q_N\widetilde{\texttt{PKey}}_{1,0} + q_N\widetilde{\texttt{PKey}}1, 1, \tag{6}$$

and σ_{bad} is a density operator (in particular it is of unit trace). By concavity of von Neumann entropy, we therefore have:

$$S(B|E)_\rho \geq \frac{p_G}{M} \cdot S(B|G)_{\sigma_{good}}. \tag{7}$$

Note that, if $q_N = 1$, then $p_G = M$ and so the "bad" case never occurs.

At this point, we use a Theorem proven in [18] to bound the entropy in σ_{good}. Invoking this theorem, and combining with Equation 7, leaves us with the following bound:

$$\begin{aligned} S(B|E)_\rho \geq \frac{p_G}{M}\left(\frac{\texttt{PKey}_{0,0} + q_N\widetilde{\texttt{PKey}}_{1,1}}{p_G}\left[h\left(\frac{\texttt{PKey}_{0,0}}{\texttt{PKey}_{0,0} + q_N\widetilde{\texttt{PKey}}_{1,1}}\right) - h(\lambda_1)\right]\right) \\ + \frac{p_G}{M}\left(\frac{\texttt{PKey}_{0,1} + q_N\widetilde{\texttt{PKey}}_{1,0}}{p_G}\left[h\left(\frac{\texttt{PKey}_{0,1}}{\texttt{PKey}_{0,1} + q_N\widetilde{\texttt{PKey}}_{1,0}}\right) - h(\lambda_2)\right]\right) \end{aligned} \tag{8}$$

where:

$$\lambda_1 = \frac{1}{2}\left(1 + \frac{\sqrt{(\texttt{PKey}_{0,0} - q_N\widetilde{\texttt{PKey}}_{1,1})^2 + 4q_N Re^2\left\langle E|F\right\rangle}}{\texttt{PKey}_{0,0} + q_N\widetilde{\texttt{PKey}}_{1,1}}\right) \tag{9}$$

$$\lambda_2 = \frac{1}{2}\left(1 + \frac{\sqrt{(\texttt{PKey}_{0,1} - q_N\widetilde{\texttt{PKey}}_{1,0})^2 + 4q_N Re^2\left\langle e_0|f_1\right\rangle}}{\texttt{PKey}_{0,1} + q_N\widetilde{\texttt{PKey}}_{1,0}}\right) \tag{10}$$

Computing $H(B|A)$ is trivial given $\texttt{PKey}_{i,j}$. We have therefore reduced the problem to estimating $\widetilde{\texttt{PKey}}_{i,j}$, $Re\left\langle E|F\right\rangle$, and $Re\left\langle e_0|f_1\right\rangle$. In general, this may be done by taking into account the unitarity of the attack operation U (in a manner similar to that done for proving B92 is secure [19]). We will compute bounds on these settings for certain practical attacks against this system leaving a complete analysis as future work.

3.1 Unambiguous State Discrimination Attack

One particularly devastating attack against standard B92 is the Unambiguous State Discrimination (USD) [20, 21] attack. Since our system is, essentially, the semi-quantum version of B92, it is no surprise that our protocol also suffers against it; in fact, things may be worse since the "source" from B is affected by E's first attack in the forward channel. Here we will compare the effectiveness of our protocol with that of B92 for this attack.

The USD attack may be modeled as a unitary operator U (applied in the reverse channel - see Equation 5):

$$\begin{aligned} U\left|E_N\right\rangle &= \sqrt{T}\left|+, 0\right\rangle + \sqrt{1-T}\left|v, f_v\right\rangle \\ U\left|1^N\right\rangle \otimes \left|\sigma_N^E\right\rangle &= \sqrt{T}\left|1, 1\right\rangle + \sqrt{1-T}\left|v, e_v\right\rangle \end{aligned}$$

(the action of U on $|1^n\rangle \otimes |\sigma_n^E\rangle$ is arbitrary - in fact our entropy bound assumes E gains full information on such states). Above, $1-T$ is the probability A observes a vacuum state.

Let $\alpha = \langle 1^N, \sigma_N | E_N \rangle$. Unitarity requires that:

$$\alpha = (1-T)\,\langle f_v | e_V \rangle$$

which, due to the fact that $|\,\langle f_v | e_v \rangle\,| \leq 1$, implies that, for U to be unitary, it must hold that $T \leq 1 - |\alpha|$. Thus, *to be secure*, it must hold that:

$$T > 1 - |\alpha| \tag{11}$$

For B92, where α is a function of an honest user's source (e.g., $|\alpha| = |\,\langle +|1\rangle\,| = \frac{1}{\sqrt{2}}$), this leads to the known bound that T must be greater than 29.3% [21]. In particular, if A observes T less than this (i.e., if the probability of a photon loss is greater than 70.7%), users must abort as, potentially, E is able to extract full information from the signal. For our protocol, the story is not as clear, since $|\alpha|$ can be chosen by the adversary and cannot be directly observed!

While $|\alpha|$ cannot be directly observed, it can be bounded as a function of the detector efficiency. Indeed, let P_{NC} be the probability that B's box outputs a 0 on its classical wire when it is given the command to `Measure and Resend`. Also define v to be the dark-count probability of the detector and $\eta > 0$ its efficiency (here we are assuming the implementation shown in Figure 2 is used - other implementations of the "box" will require different characterizations at this point). By definition of the action of B's "box" it holds that:

$$q_m = (1-v) \sum_{\substack{i \in \{0,1\}^N \\ w(i)=m}} |\alpha_i|^2 (1-\eta)^{N-m},$$

and so:

$$P_{NC} = (1-v) \sum_{m=0}^{N} (1-\eta)^{N-m} \sum_{\substack{i \in \{0,1\}^N \\ w(i)=m}} |\alpha_i|^2,$$

where $w(i)$ is the Hamming weight of the bit-string i (i.e., it is the number of 1's in the string i). From this, we see:

$$\begin{aligned}
\frac{P_{NC}}{1-v} &= |\alpha|^2 + (1-\eta) \left(\sum_{m=0}^{N-1} (1-\eta)^{N-m-1} \sum_{\substack{i \in \{0,1\}^N \\ w(i)=m \\ i \neq 11\cdots 1}} |\alpha_i|^2 \right) \\
&\leq |\alpha|^2 + (1-\eta) \left(\sum_{i \neq 11\cdots 1} |\alpha_i|^2 \right) \leq |\alpha|^2 + (1-\eta)\left(1 - |\alpha|^2\right)
\end{aligned}$$

$$\Rightarrow |\alpha|^2 \geq \frac{P_{NC}}{\eta(1-v)} - \frac{1-\eta}{\eta} \tag{12}$$

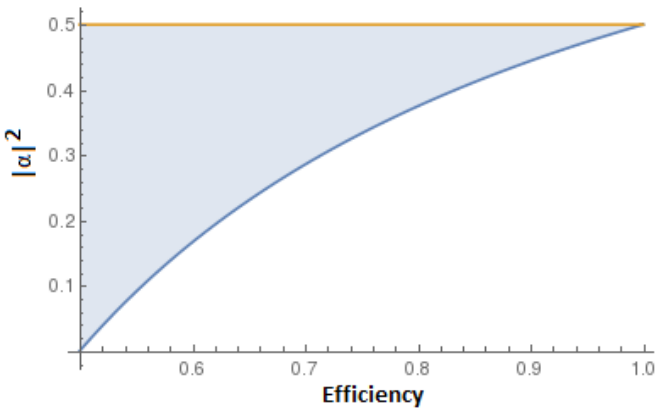

Figure 3: Possible values of $|\alpha|^2$ based on detector efficiency η.

Similarly, we may upper-bound the quantity:

$$\frac{P_{NC}}{(1-v)} = |\alpha|^2 + (1-\eta)\left(\sum_{m=0}^{N-1}(1-\eta)^{N-m}\sum_{\substack{i\in\{0,1\}^N \\ w(i)=m \\ i\neq 11\cdots 1}}|\alpha_i|^2\right) \geq |\alpha|^2$$

$$\Rightarrow |\alpha|^2 \leq \frac{P_{NC}}{1-v}. \tag{13}$$

Note that, when $v = 0$ and $\eta = 1$ (i.e., the detector is perfect), then $|\alpha|^2 = P_{NC}$ and is directly observable. As η decreases, the possible range of values for this quantity increase as shown in Figure 3. Naturally, we must assume the worst case and so our protocol is secure against the USD attack, so long as the probability of a photon loss $(1 - T)$ satisfies:

$$1 - T < \sqrt{\frac{P_{NC}}{\eta(1-v)} - \frac{1-\eta}{\eta}}. \tag{14}$$

If B's devices are perfect (in that $v = 0$ and $\eta = 1$), and if $P_{NC} = 1/2$ (which could even be enforced), then the maximal loss tolerated is 70.7% as with B92. This maximal loss drops as η increases. This is shown in Figure 4.

In particular, if $T = 10^{-a\cdot\ell/10}$, where ℓ is the distance of the quantum channel, then this protocol is secure against the USD attack, as long as:

$$\ell < -\frac{10}{a}\log_{10}\left(1 - \sqrt{\frac{P_{NC}}{\eta(1-v)} - \frac{1-\eta}{\eta}}\right)$$

The maximal distance supported is shown in Figure 5 for various η assuming $a = .25dB/km$.

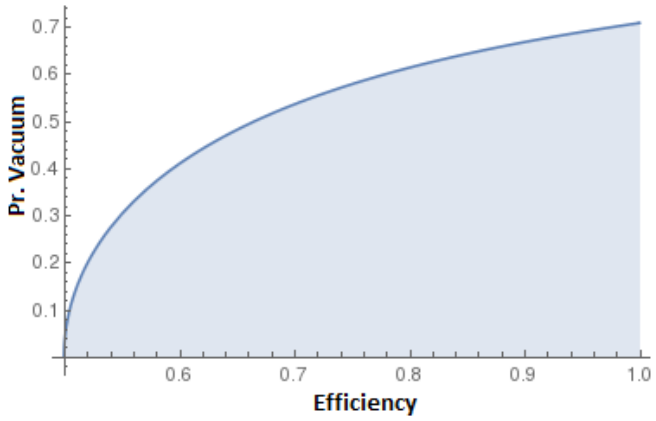

Figure 4: Protocol is secure against the USD attack so long as $1 - T$ (i.e., the probability of photon loss) is in the shaded region. Put differently, the protocol is definitely insecure outside the shaded region.

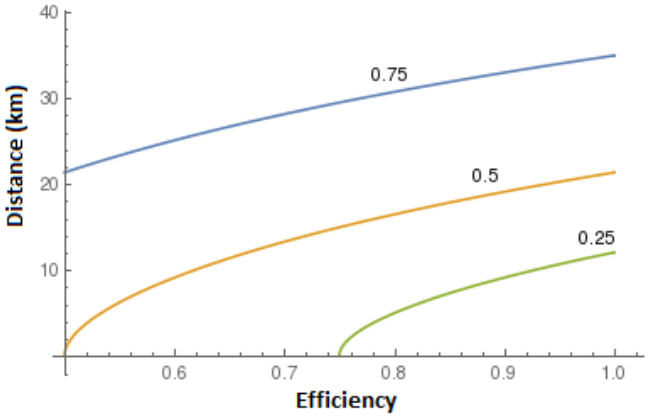

Figure 5: Maximum distance (from B to A) over which the protocol is secure against the USD attack. Shown are $P_{NC} = .75$, .5, and .25

.

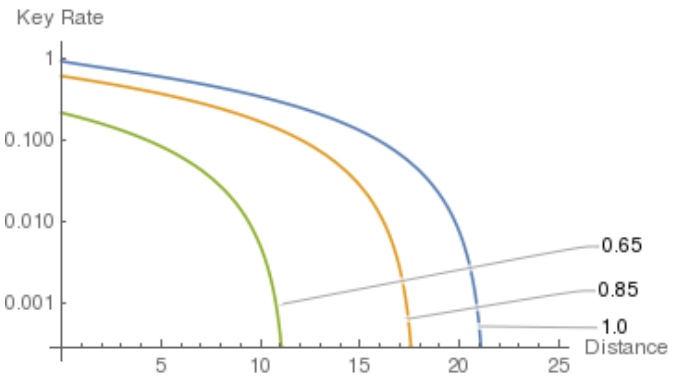

Figure 6: Showing the key-rate of our protocol against the attack described in Equation 15 as a function of Distance (in km) for $\eta = .65, .85$, and 1. Here $P_{NC} = .5$

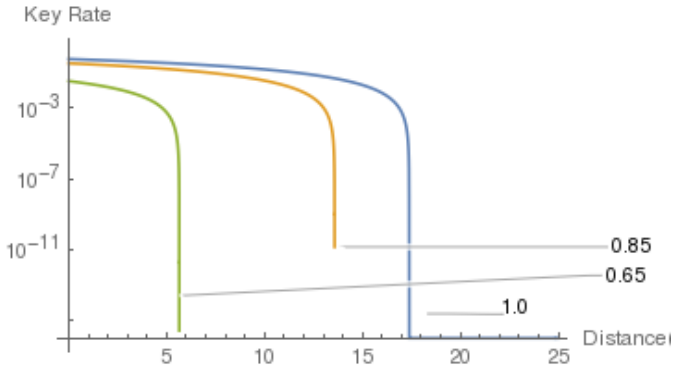

Figure 7: Showing the key-rate of our protocol against the attack described in Equation 15 as a function of Distance (in km) for $\eta = .65, .85$, and 1. Here $P_{NC} = .4$

Now, beyond the USD attack, we may more generally write this attack as follows:

$$\begin{aligned} U\left|E_n\right\rangle &= \left|+, f_0\right\rangle + \left|v, f_v\right\rangle \\ U\left|1^N\right\rangle &= \left|1, e_1\right\rangle + \left|v, e_v\right\rangle \end{aligned} \tag{15}$$

We assume a symmetry in that $\langle f_0|f_0\rangle = \langle e_1|e_1\rangle = T$. Unitarity requires that $Re\langle f_0|e_1\rangle = \sqrt{2}(Re(\alpha) - Re\langle f_v|e_v\rangle)$. Using our bound on $|\alpha|$, along with the Cauchy-Schwarz inequality to bound $|\langle f_v|e_v\rangle| \leq (1-T)$ and Equation 8, allows us to compute the key-rate as a function of distance assuming only photon loss in the channel (note $\langle E|F\rangle = \frac{1}{2}\langle e_1|f_0\rangle$ in this attack scenario). Also, we have $\widetilde{\texttt{PKey}}_{1,1} = \frac{1}{2}T$ and $\widetilde{\texttt{PKey}}_{1,0} = 0$. Finally, assuming the implementation shown in Figure 2, we have $q_N = (1-v)|\alpha|^2$. This is shown in Figures 6, 7, and 8. Note that, as P_{NC} decreases, the maximal distance also decreases. Keep in mind that P_{NC} is affected by N and E's forward channel attack - ideally, $P_{NC} = .5$, so lower values constitute "noise."

Note that, these attacks may be mitigated by extending our protocol in a manner similar to that done for extended-B92 as discussed in [14] and this remains an interesting open

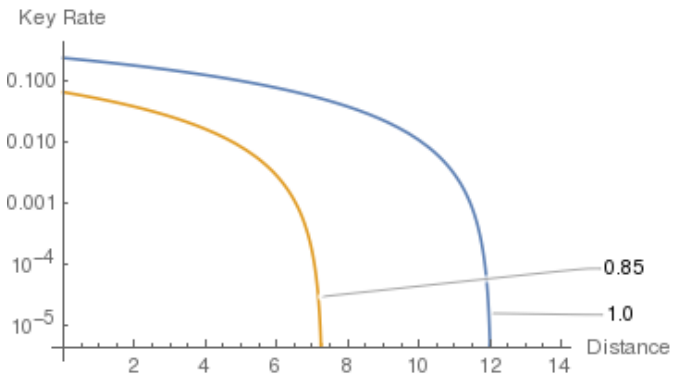

Figure 8: Showing the key-rate of our protocol against the attack described in Equation 15 as a function of Distance (in km) for $\eta = .85$, and 1 (for $\eta = .65$ the key-rate dropped to zero almost immediately). Here $P_{NC} = .25$

problem (it also greatly complicates the information theoretic security analysis due to the fact that, now, A will be preparing different states). Furthermore, bounding $Re\langle f_0|e_1\rangle$ when channel noise is also present (we only considered channel loss here), is also an important open question. Our equation for $S(B|E)$, however, derived in the previous section can be used towards this end.

3.2 Multi-Photon Attack

In this section, we consider how much information E can gain just by attacking the forward channel. As before, let $|e\rangle = |E_N\rangle = \sum_{x\in\{0,1\}^N} \alpha_x |x\rangle |e_x\rangle$. Assume E captures the entire state leaving B's lab. In the event his key-bit is 0, this state is simply $\rho_0 = [\mathbf{E_N}]$. If his key-bit is 1 (and conditioning on the event he will later accept - i.e., $c_2 = 0$), then $\rho_2 = \frac{1}{P_{NC}} \sum_n q_n [\mathbf{1^n}]$. Assume the worst case in that she attempts to extract information from the state at this point, instead of probing it further and forwarding a qubit to A. Though A, of course, requires a qubit to complete the protocol iteration, if we compute $S(B|E)$ at this point, it can only be lower than in the "real" case.

It is not difficult to show by definition that $S(B|E) = H(B) - I(B:E)$. Using a result from [22], we have $I(B:E) \leq \frac{1}{2}||\rho_0 - \rho_1||$. Thus:

$$S(B|E) \geq H(B) - \frac{1}{2}||\rho_0 - \rho_1|| \geq H(B) - \frac{1}{2}\left|\left|[\mathbf{E_N}] - \frac{q_N}{P_{NC}}[\mathbf{1_N}]\right|\right| - \frac{P_{NC} - q_N}{2P_{NC}},$$

where the last inequality follows from the triangle inequality and the fact that, for positive A, $||A|| = tr(A)$. Also, we used the fact that $P_{NC} = \sum_n q_n$.

Since $[\mathbf{E_N}] - \frac{q_N}{P_{NC}}[\mathbf{1_N}]$ is Hermitian and dimension no greater than two, the trace-norm is simply the sum of the absolute value of the (two) eigenvalues. These eigenvalues are easily

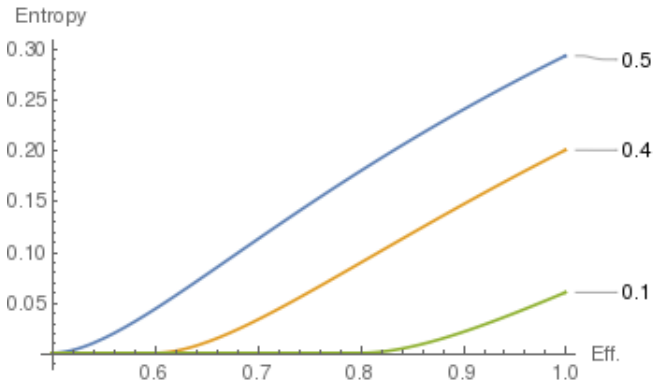

Figure 9: A lower-bound on $S(B|E)$ if E attacks the forward channel and tries to extract information from the returning state (before passing it to A) as a function of B's efficiency η ("Eff."). Shown are various levels of observed $P_{NC} = .5$, .4, and .1. Note that the larger N, the more likely P_{NC} will be low.

computed as:

$$\lambda_{\pm} = \frac{1}{2}\left(1 - \frac{q_N}{P_{NC}} \pm \sqrt{\left(1 + \frac{q_N}{P_{NC}}\right)^2 - \frac{4q_N|\alpha|^2}{P_{NC}}}\right),$$

where, as in the previous section, we define α to be $\langle 1^N|E_N\rangle = \alpha_{11\cdots1}$. In conclusion, therefore, if E only attacks the forward channel (i.e., launches a multi-photon attack against B's box), we have:

$$S(B|E) \geq H(B) - \frac{1}{2}(|\lambda_+| + |\lambda_-|) - \frac{P_{NC} - q_N}{2P_{NC}}.$$

$H(B)$ can be directly computed by B (in fact this may even be made to be 1 with some standard post-processing techniques). Also, P_{NC} is an observable statistic. The value of $|\alpha|^2$ may be bounded as described in the previous section. Finally, assuming the implementation shown in Figure 2, it holds that $q_N = (1 - v)|\alpha|^2$. To compute $S(B|E)$, therefore, one simply must numerically minimize the above expression over all valid $|\alpha|^2$. The result of this minimization for various η and v is shown in Figure 9.

4 CLOSING REMARKS

In this paper we introduced a new SQKD protocol and analyzed its security against certain practical attacks. Many interesting questions remain open including a full analysis against all attacks (though, in this paper, we have reduced the problem to only estimating $\langle E|F\rangle$, $\langle e_0|f_1\rangle$, and $\widetilde{\texttt{PKey}}_{i,j}$ thus aiding future researchers). Our security analysis assumed the existence of a "black box" and our evaluations involved one particular example of an implementation of this box. Can a better implementation be constructed improving the tolerance to low efficiency detectors?

References

[1] Valerio Scarani, Helle Bechmann-Pasquinucci, Nicolas J. Cerf, Miloslav Dušek, Norbert Lütkenhaus, and Momtchil Peev. The security of practical quantum key distribution. *Rev. Mod. Phys.*, 81:1301–1350, Sep 2009.

[2] Michel Boyer, D. Kenigsberg, and T. Mor. Quantum key distribution with classical bob. In *Quantum, Nano, and Micro Technologies, 2007. ICQNM '07. First International Conference on*, pages 10–10, 2007.

[3] Michel Boyer, Ran Gelles, Dan Kenigsberg, and Tal Mor. Semiquantum key distribution. *Phys. Rev. A*, 79:032341, Mar 2009.

[4] Wang Jian, Zhang Sheng, Zhang Quan, and Tang Chao-Jing. Semiquantum key distribution using entangled states. *Chinese Physics Letters*, 28(10):100301, 2011.

[5] Xiangfu Zou, Daowen Qiu, Lvzhou Li, Lihua Wu, and Lvjun Li. Semiquantum-key distribution using less than four quantum states. *Phys. Rev. A*, 79:052312, May 2009.

[6] Walter O Krawec. Mediated semiquantum key distribution. *Physical Review A*, 91(3):032323, 2015.

[7] Hua Lu and Qing-Yu Cai. Quantum key distribution with classical alice. *International Journal of Quantum Information*, 6(06):1195–1202, 2008.

[8] Walter O Krawec. Security proof of a semi-quantum key distribution protocol. In *Information Theory (ISIT), 2015 IEEE International Symposium on*, pages 686–690. IEEE, 2015.

[9] Wei Zhang, Daowen Qiu, Xiangfu Zou, and Paulo Mateus. A single-state semi-quantum key distribution protocol and its security proof. *arXiv preprint arXiv:1612.03087*, 2016.

[10] Yong-gang Tan, Hua Lu, and Qing-yu Cai. Comment on quantum key distribution with classical bob. *Phys. Rev. Lett.*, 102:098901, Mar 2009.

[11] Michel Boyer, Dan Kenigsberg, and Tal Mor. Boyer, kenigsberg, and mor reply:. *Phys. Rev. Lett.*, 102:098902, Mar 2009.

[12] Michel Boyer, Matty Katz, Rotem Liss, and Tal Mor. Experimentally feasible protocol for semiquantum key distribution. *Phys. Rev. A*, 96:062335, Dec 2017.

[13] Walter O Krawec. Restricted attacks on semi-quantum key distribution protocols. *Quantum Information Processing*, 13(11):2417–2436, 2014.

[14] Marco Lucamarini, Giovanni Di Giuseppe, and Kiyoshi Tamaki. Robust unconditionally secure quantum key distribution with two nonorthogonal and uninformative states. *Physical Review A*, 80(3):032327, 2009.

[15] Walter O Krawec. Security of a semi-quantum protocol where reflections contribute to the secret key. *Quantum Information Processing*, 15(5):2067–2090, 2016.

[16] Valerio Scarani and Renato Renner. Quantum cryptography with finite resources: Unconditional security bound for discrete-variable protocols with one-way postprocessing. *Physical review letters*, 100(20):200501, 2008.

[17] Charles H. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.*, 68:3121–3124, May 1992.

[18] Walter O. Krawec. Quantum key distribution with mismatched measurements over arbitrary channels. *Quantum Information and Computation*, 17(3 and 4):209–241, 2017.

[19] Matthias Christandl, Renato Renner, and Artur Ekert. A generic security proof for quantum key distribution. *arXiv preprint quant-ph/0402131*, 2004.

[20] Kiyoshi Tamaki, Masato Koashi, and Nobuyuki Imoto. Security of the bennett 1992 quantum-key distribution protocol against individual attack over a realistic channel. *Physical Review A*, 67(3):032310, 2003.

[21] Heasin Ko, Byung-Seok Choi, Joong-Seon Choe, and Chun Ju Youn. Advanced unambiguous state discrimination attack and countermeasure strategy in a practical b92 qkd system. *Quantum Information Processing*, 17(1):17, 2018.

[22] Jop Briët and Peter Harremoës. Properties of classical and quantum jensen-shannon divergence. *Physical review A*, 79(5):052311, 2009.