
Article:
Vinay, S.E. and Kok, P. (2018) Extended analysis of the Trojan-horse attack in quantum key distribution. Physical Review A, 97 (4). 042335. ISSN 2469-9926

https://doi.org/10.1103/PhysRevA.97.042335

Extended analysis of the Trojan-horse attack in Quantum Key Distribution

Scott E. Vinay* and Pieter Kok†
Department of Physics and Astronomy, University of Sheffield
(Dated: May 17, 2018)

The discrete-variable QKD protocols based on BB84 are known to be secure against an eavesdropper, Eve, intercepting the flying qubits and performing any quantum operation on them. However, these protocols may still be vulnerable to side-channel attacks. We investigate the Trojan-Horse side-channel attack where Eve sends her own state into Alice's apparatus and measures the reflected state to estimate the key. We prove that the separable coherent state is optimal for Eve amongst the class of multi-mode Gaussian attack states, even in the presence of thermal noise. We then provide a bound on the secret key rate in the case where Eve may use any separable state.

I. INTRODUCTION

Quantum Key Distribution (QKD) systems are generally created with the promise that the uncertainty inherent in quantum measurements allows for two or more parties to communicate with unconditional security. By this, it is meant that an eavesdropper, Eve, may be imbued with unbounded computational power and be able to do anything that is allowed by the laws of physics, yet still only achieve a level of mutual information with a bit string shared by valid parties Alice and Bob that is exponentially small with the key length [1]. This is in contrast with classical encryption, for which the above claim only holds when Eve has some bounded computational ability (which may be exceeded by a quantum computer [2]).

In any claim of security assumptions will necessarily be made on restrictions on the methods by which Eve may try to learn the key. For example, it is clear that if Eve has unrestricted access to Alice's lab then no level of sophistication in the protocol can prevent her from learning the key. Therefore we must always decide on a boundary demarcating the quantum or classical objects that Eve may access from the ones that she may not.

In the standard proofs of the security of many QKD protocols, it is assumed that Eve may interact with any of the "flying" photonic qubits that are sent between Alice and Bob and with the quantum channel carrying them. Her operations may include storing the qubits for arbitrarily long periods of time, performing multipartite rotations or measurements, entangling these with ancilla states, or replacing sections of the channel with loss-free channels. Renner used the de Finetti theorem to show that this very general case is equivalent to the case of Eve performing operations on one qubit at a time [3], and along with [4] and [5], this proves the security of BB84 [6] against such attacks.

However, we must assume that Eve is wily and cunning, and will seek alternative hidden avenues known as *side-channel attacks* (SCAs) [7–15]. One such SCA that has recently received theoretical [16–18] and experimental [19–21] attention is the so-called *Trojan Horse Attack* (THA). Here, Eve will tap into the optical channel that Alice and Bob use to communicate. She will then send her own optical state into Alice's system, whereupon it will reflect off the same apparatus used to encode the legitimate photonic qubits. Having picked up some information on the encoding of the latest quantum state that Alice sent, it will return out and be measured by Eve. Eve will then use the result of this measurement, possibly combined with some operation on the legitimate qubits, to make a best estimate of the state that Alice sent to Bob, thus giving her some non-negligible information mutual with the key.

This attack has previously been analysed by Lucamarini et al. [18, 22]. They assume that Eve uses a coherent state to probe the system, and describe using a one-way attenuating filter at the entry-point of Alice's apparatus as a defense. The effect of this is to absorb the majority of light that is sent into the system, such that Eve receives far less than one photon back per attempt, reducing her ability to estimate the key bit. They make use of the theoretical framework of Gottesman et al. [23] to get an expression for the rate at which Alice and Bob can generate a secret key in the presence of such an attack.

In this paper we make use of this same framework, but reduce the restrictions placed on the state that Eve may use. In section II we describe some of the fundamental notions necessary to understand the process of encoding in a phase-modulated BB84 protocol. In section III, we describe and analyse the effect of Eve performing a THA on the system, allowing her to use any Gaussian state including multimode entangled states. We prove that the (separable) coherent states are optimal amongst this class.

Motivated by the revelation that entanglement does not assist Eve when using Gaussian states to attack the system, in section IV we restrict Eve to separable states. We then derive a bound on the information that Eve may learn about the key when we allow her to use *any* separable state.

* svinay1@sheffield.ac.uk
† p.kok@sheffield.ac.uk

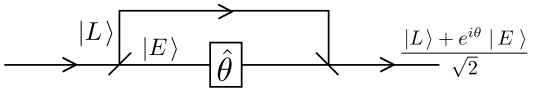

FIG. 1: A phase-shift between an early and a late mode, parameterised by θ, encodes the quantum bit.

II. PRELIMINARY NOTIONS

We consider here an implementation of the BB84 [6] protocol. Here, Alice chooses one of two mutually unbiased bases, X and Y. After choosing a basis she then sends a photon encoding either state $|0_{X,Y}\rangle$ or $|1_{X,Y}\rangle$. These states are encoded as

$$|0,1_{X,Y}\rangle = \frac{|L\rangle + e^{i\theta}\,|E\rangle}{2}, \tag{1}$$

where $|E\rangle$ and $|L\rangle$ are early and late modes as shown in Fig. 1.

The key parameter here is θ, which encodes the state as follows:

$$\begin{array}{ll} |0_X\rangle \to \theta = 0 & |0_Y\rangle \to \theta = \pi/2 \\ |1_X\rangle \to \theta = \pi & |1_Y\rangle \to \theta = 3\pi/2 \end{array} \tag{2}$$

It is this parameter that Eve wishes to estimate. In order to do this, she prepares her own state, ρ. This is assumed to exist in the photonic Fock space of a single mode. The single mode assumption is justified since we may say that Alice will filter out all frequencies that are not equal to the one sent to Bob. It may also be assumed without loss of power or generality to be pure. This state is sent into Alice's system. Here it passes through a filter which allows a fraction $\eta \ll 1$ of the light to be transmitted, resulting in a state ρ_η. It then reaches the polarizing filter, where it evolves according to the same Hamiltonian that encoded θ into the photon that was sent to Bob. That is to say, it is transformed as follows:

$$\rho_\eta \to \rho_\eta^\theta \equiv e^{i\theta\hat{a}^\dagger\hat{a}}\rho_\eta\, e^{-i\theta\hat{a}^\dagger\hat{a}}, \tag{3}$$

where $\hat{a}$ is the annihilation operator on the Fock space of Eve's photons. After Eve's state has picked up the phase information it returns to her. She then performs some operation to try to make an estimate of θ. Note that this framework analyses the effect of the phase modulator on the state. The effect of an intensity modulator, which is used in decoy-state QKD, is analysed in [22].

III. GAUSSIAN STATE ATTACK

Attenuation-based defense systems, which aim to muddy the phase information on Eve's state by blocking most of the incoming attack state, have been previously analysed by Lucamarini et al. [18]. In their analysis, it was assumed that Eve would send in a pure coherent state, which is by necessity separable. Here we wish to extend this analysis by relaxing this requirement, and consider the case where Eve may use any multi-mode entangled Gaussian state. Since only one mode enters Alice's apparatus, Eve needs only to use at most one idler mode, which she retains as a reference [24].

We will also consider that Alice's system may not be noise-free. In particular, with any use of apparatus that operates at a non-zero temperature, the mode that is sent to Bob will include a small amount of thermal noise. We will include the presence of this noise in the state that Eve receives, and examine its effect on the information that Eve can extract from her attack. This is included in order to show that Eve's optimal attack strategy does not change in the presence of thermal noise, and increase the generality of the analysis. The combination of ρ_η^θ with the thermal noise will be notated as ρ_{η,μ_T}^θ. Here, μ_T is the average thermal photon number (if any) in the mode occupied by the photon sent to Bob.

A. State description

Here we will describe specifically how we construct ρ_{η,μ_T}^θ from ρ, and how Eve should choose ρ to maximise her knowledge of the key.

Firstly, it is clear that the choice of initial state ρ will have a significant effect on Eve's ability to discern θ. There are certain properties of this state that we can identify that we expect to affect this in varying degrees.

The property that may be most apparent is that of the average photon number of the state. If Eve sends in a single photon, then given a high amount of attenuation, she is not likely to get much back and will not be able to reliably learn θ. On the other hand if she is allowed to send in an arbitrarily bright state with unbounded average photon number it is clear that she will always be able to distinguish the different settings of θ perfectly. Therefore, to be able to implement any QKD protocol, the first step in protecting against a THA is putting some upper bound on the average number of photons that may pass into the system. This may be done by way of some defense such as an optical fuse [25], which melts when sufficiently many photons pass through it, or by identifying some other component which will be irreversibly damaged when subject to a bright enough light [26]. A more detailed examination of the numbers and figures behind such defenses may be found in [18], but for our purposes we may simply assume that there does exist some bound N such that $\langle\hat{n}\rangle_\rho \equiv \mathrm{Tr}\,[\hat{n}\rho] < N$, where $\hat{n} = \hat{a}^\dagger\hat{a}$.

Another relevant property may be the purity of the state ρ_η *after* passing through the attenuator. Most states will become mixed after undergoing loss, however coherent states (as used in [18]) will not. They are instead mapped to coherent states with lower photon numbers. As a result of this, the loss does not introduce any

classical uncertainty into the estimation of the phase. It may also be the case that entanglement assists the estimation, as is the case with entanglement-assisted illumination [27]. It is important that we search for the most powerful possible attack that Eve may make, taking all of these factors into consideration. It is only then that we may have confidence in our security proofs against the THA or other SCAs.

Eve's multimode Gaussian state may be created by applying a two-mode squeezer to the vacuum followed by a displacement on the mode that enters Alice's system (applying a displacement to the idler mode turns out to have no effect on the the amount of information that Eve may learn about the key). Up to a change of variables in the squeezing and displacement parameters, this setup is equivalent to all other combinations of Gaussian operations [28], such as applying single-mode squeezers and displacing before squeezing. This, therefore, represents the most general Gaussian-state attack that Eve may make.

Eve's initial state is then

$$\rho = \hat{D}(\alpha)\ \hat{S}_2(\xi_E)\ |0\rangle\langle 0|\ \hat{S}_2^\dagger(\xi_E)\ \hat{D}^\dagger(\alpha), \tag{4}$$

where $\hat{D}(\alpha) = \exp\left(\alpha\hat{a}^\dagger - \alpha^*\hat{a}\right)$ is the displacement operator, $\hat{S}_2(\xi_E) = \exp\left(\frac{1}{2}\xi_E\hat{a}^\dagger\hat{b}^\dagger - \xi_E^*\hat{a}\hat{b}\right)$ is the two-mode squeeze operator (where $\hat{b}$ acts on Eve's idler mode) and $|0\rangle$ is the vacuum state. Without loss of generality we will let ξ_E be real.

As is typical, we will model the loss due to the attenuator as a beam splitter. A fraction η is allowed to pass through to reach Alice's apparatus, and $1-\eta$ is diverted into an auxiliary environment mode.

The final ingredient to be included is the thermal noise. Here we need some careful thought as to how exactly we will *mathematically* combine these two states. In other papers [29], thermal noise has been added to a signal by passing both the signal and the noise through a beam splitter. However, this does not seem to us to be appropriate in this situation for the following reason. Suppose the combined state is produced by passing these two states through a beam splitter with transmissivity η_T, so that $\eta_T = 1$ means that the resulting state is entirely a thermal state, and $\eta_T = 0$ means it is all signal. However, this introduces a new variable into the situation, which implies some degree of coupling between the thermal source and Eve's returned state. We want Eve to be oblivious as to the actual source of the thermal noise, and simply consider it as a simultaneously arriving light source. In particular, if we let $\eta_T = 1$ and $\mu_T = 0$, we arrive at the rather paradoxical conclusion that the signal has been completely overwhelmed by a thermal state containing no photons. For a similar reason we cannot combine ρ_η^θ with a thermal density matrix $\rho_{\rm Th}$ by way of a classical mixture such as $p\,\rho_\eta^\theta + (1-p)\,\rho_{\rm Th}$. As such, we expect that the strength of the thermal noise should depend only on the single parameter μ_T.

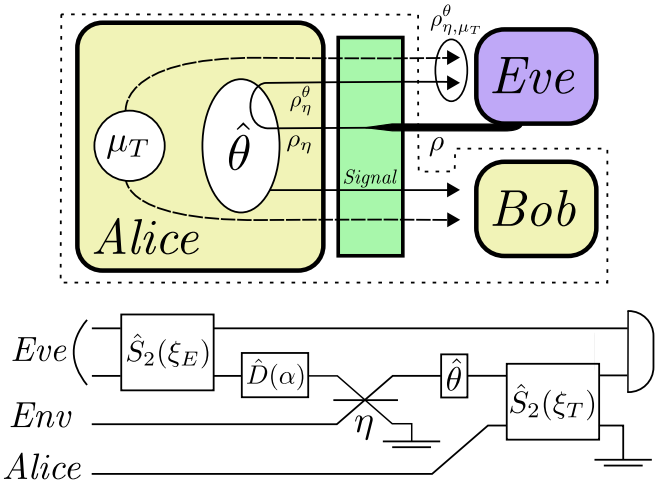

FIG. 2: Top: Schematic diagram illustrating the *physical* mechanisms that produce Bob and Eve's states. The dotted line outlines the "legitimate" part of the protocol, comprising the signal and the thermal noise (dashed line). Bottom: Circuit diagram showing the *mathematical* mechanisms that produce Eve's final state. Shown from left to right are the effects of Eve's squeezing, Eve's state displacement, Alice's attenuator, picking up the phase information, and adding the thermal noise. Double horizontal lines represent taking a partial trace over the relevant mode.

A method for the proper treatment of constructing a combined state from multiple simultaneously arriving photonic states was described by Glauber in his original treatment of the coherent states [30]. However, that method involved expressing the states in a diagonal coherent basis (the so-called P-representation). Whilst this is a powerful method, it results in an expression for the state that is no longer easily analytically tractable (although it *is* possible to use this to *numerically* analyse the effects of adding non-thermal noise). Since we are dealing here with Gaussian states we shall take advantage of a nice property of thermal states: that they may be produced be taking the partial trace over one mode of a two-mode squeezed vacuum with squeezing parameter $\xi_T = \text{arcsinh}(\sqrt{\mu_T})$. Therefore, we shall model the addition of the thermal noise as Eve's returning signal being passed through a two-mode squeezer with the vacuum, and discarding one of the resulting modes. Within this framework, Eve should choose α and ξ_E to maximise her mutual information with the secret key. The full set-up for the construction of Eve's state is illustrated in Fig. 2.

A great advantage of working with Gaussian states is that they may be completely characterised by their first and second moments. For an n-mode Gaussian state let $\underline{\hat{u}}$ be the vector of operators $[\hat{x}_1, \hat{p}_1, \ldots, \hat{x}_n, \hat{p}_n]^T$. Then to each Gaussian state, ρ, we may uniquely assign a pair $(\underline{u}, V)$ which we call the *mean vector* and *covariance matrix* respectively, with elements defined by

$$u_i = \mathrm{Tr}\left[\rho \hat{u}_i\right], \\ V_{i,j} = \frac{\mathrm{Tr}\left[\rho \hat{u}_i \hat{u}_j\right] + \mathrm{Tr}\left[\rho \hat{u}_j \hat{u}_i\right]}{2} - \mathrm{Tr}\left[\rho \hat{u}_i\right] \mathrm{Tr}\left[\rho \hat{u}_j\right]. \tag{5}$$

Let ϕ be the relative angle between the displacement and the squeezing parameters in the complex plane, $\mu_D = \eta |\alpha|^2$ be the average displacement after loss, and $\omega = \cosh(2\xi_E)$ be the normalised quadrature variance for a squeezed vacuum state. It then follows from Eq. 5 and Eq. 4 that the mean vector and covariance matrix for Eve's returned states corresponding to $\theta = 0$ and $\theta = \frac{\pi}{2}$ are as follows:

$$u_{\theta=0} = \begin{bmatrix} (\sin\phi + \cos\phi)\sqrt{2\mu_D} \\ (\sin\phi - \cos\phi)\sqrt{2\mu_D} \\ 0 \\ 0 \end{bmatrix}, \\ V_{\theta=0} = \frac{1}{2}\begin{bmatrix} \left[(1+\mu_T)\,\omega\eta + \mu_T\right]\mathbb{1}_2 & A\,\sigma_Z \\ A\,\sigma_Z & \omega\mathbb{1} \end{bmatrix}, \\ u_{\theta=\frac{\pi}{2}} = \begin{bmatrix} (\cos\phi - \sin\phi)\sqrt{2\mu_D} \\ (\cos\phi + \sin\phi)\sqrt{2\mu_D} \\ 0 \\ 0 \end{bmatrix}, \\ V_{\theta=\frac{\pi}{2}} = \frac{1}{2}\begin{bmatrix} \left[(1+\mu_T)\,\omega\eta + \mu_T\right]\mathbb{1}_2 & A\,\sigma_X \\ A\,\sigma_X & \omega\mathbb{1} \end{bmatrix}, \tag{6}$$

where $A = \sqrt{(1+\mu_T)\left(\omega^2 - 1\right)\eta}$ and σ_Z and σ_X are the Pauli Z and X matrices respectively.

B. Secret key rate

In all QKD systems the main quantity of merit is the secret key rate, K. This is the rate at which Alice and Bob can generate key bits with exponentially high security, which is in general lower than the rate at which Alice and Bob exchange raw key bits. This quantity is dependent on the specific choice of protocol that is being implemented. Here, we will analyse the performance of the BB84 protocol, which may be seen to be equivalent to entanglement-based protocols such as E91. This is because instead of deciding on a key bit $|0_B\rangle$ or $|1_B\rangle$ to send out for some basis B, Alice may instead prepare the entangled state $\left(|0_B\rangle|\uparrow\rangle + |1_B\rangle|\downarrow\rangle\right)/\sqrt{2}$, send the first mode to Bob and keep the second mode. She would then make a measurement of her retained mode to "decide" on the key bit. A similar process can be done to decide the basis. Since both frameworks are the same from the point of view of local density matrices as seen by Eve, the security of one reduces to the security of the other.

In vanilla BB84 with no threat of THA, the secret key rate has been found to be [23]

$$K = R\left[1 - 2H_2(\epsilon)\right], \tag{7}$$

where R is the raw key rate, ϵ is the bit-error rate and $H_2(\epsilon) = -\epsilon\log_2(\epsilon) - (1-\epsilon)\log_2(1-\epsilon)$ is the binary entropy function. We may say that one of these terms of $H_2(\epsilon)$ is due to Alice and Bob sacrificing key bits to perform error correction, and one factor is due to them applying classical privacy amplification algorithms.

Due to the nature of the THA being a SCA, Eve's attack will not affect the bit-error rate measured by Alice and Bob. However, it will still clearly compromise the security, so Eq. 7 cannot represent the achievable secret key rate. We expect in particular that the $H_2(\epsilon)$ term representing the error correction should remain unchanged — since a properly implemented SCA will not induce additional errors. However, Alice and Bob *will* have to do additional privacy-amplification, so this term will be modified.

The key rate for BB84 in the presence of an SCA was found by [23]. They show that the effect of the SCA may be summarised by a quantity known as the *distinguishability*, Δ. This is used to modify the error rate, ϵ, in the privacy-amplification term to become an *effective error rate*, $\tilde{\epsilon}$ given by

$$\tilde{\epsilon}(\epsilon, \Delta) = \epsilon + 4\Delta(1-\Delta)(1-2\epsilon) \\ + 4(1-2\Delta)\sqrt{\Delta(1-\Delta)\epsilon(1-\epsilon)}. \tag{8}$$

This means that we do not have to know exactly what Eve does with the states and the information available to her. For example, she may perform a THA to try to learn θ directly. Or, she might tailor her THA such that the measurement on the returned state only reveals information about the basis that Alice has chosen. After estimating this basis, she might then measure the *flying* qubit in that basis to learn θ without disturbing the state. She might do some combination of these approaches, or something else entirely. As such, it is of foundational importance to our analysis that we have some way of quantifying the strength of a THA, that only makes reference to the state she sends *out*, not to *what she does* to the state she gets back, including any measurement or series of measurements on any combination of the returned state and flying qubits. The distinguishability varies from 0 when all choices of θ are indistinguishable from the point-of-view of Eve, and $\frac{1}{2}$ when she can distinguish all settings with certainty. In practice, a value of Δ much greater than 0 will result in a secret key rate of 0, since it would require Alice and Bob to be sacrificing raw key bits for error correction and privacy amplification at a rate faster than they are being generated. This formulation of the strength of a THA in terms of Δ puts a lower bound on the secret key rate that Alice and Bob can hope to achieve. The distinguishability is given by [31]

$$\Delta \leq \frac{1 - F\left(\rho^0_{\eta,\mu_T}, \rho^{\pi/2}_{\eta,\mu_T}\right)}{2}, \tag{9}$$

where $F(\rho_1, \rho_2) = \mathrm{Tr}\left[\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\right]$ is the quantum fidelity function. Note that this is different from the form given in [18]. There, they reduce Eq. 9 to a form involving the optimal purifications of the two output states. Since they are using pure coherent states, such optimal purifications are easily found. However, there exists no general prescriptive formula to find these for a pair of general mixed states, so we must use the fidelity form of the distinguishability.

One may note that in this form, Δ has a nice physically intuitive interpretation. Suppose two states are prepared, one from the X basis and one from the Y basis, and one of them is given to Eve. She is aware of which two states are prepared but not which one she received. The quantity Δ then corresponds to the minimum probability that she makes an error in distinguishing them. If she succeeds in this task, she will know θ, without needing to perform any additional operations on the flying qubits. This leads to the non-trivial conclusion that Eve's optimal THA may be performed by *only* interacting with her own returned state, and she does not gain anything by interacting with Bob's qubits.

The rest of this section is dedicated to calculating an exact expression for Δ for the set of thermalised Gaussian states described above, and section IV is focused on calculating a bound on Δ for the set of general separable states. It should be noted that, unlike ϵ, $\tilde{\epsilon}$ (or equivalently Δ) cannot be directly measured in the process of running the QKD protocol. Therefore Alice should be able to perform some local action to be able to determine Δ to some high precision, and then use this value to determine how much privacy amplification they should perform.

Note that Bob may not detect all of the signal photons that Alice sends out. Let p_{succ} be the probability that Bob detects a single photon in any given attempt at sending a qubit. When $p_{\mathrm{succ}} < 1$, we must replace Δ with Δ/p_{succ}, since the lost signals may have been selectively eliminated by Eve to improve her mutual information with the key (Ref. [23], Eq. 32).

The problem of calculating the fidelity between two multimode Gaussian states was solved by [32]. There, they show that, for any Gaussian states ρ_1, ρ_2, we have:

$$\begin{aligned} F(\rho_1, \rho_2) &= \mathcal{F}\left(\tilde{V}_1, \tilde{V}_2\right) e^{-\frac{1}{4}\left(\underline{u}_1 - \underline{u}_2\right)^T \left(\tilde{V}_1 + \tilde{V}_2\right)^{-1} \left(\underline{u}_1 - \underline{u}_2\right)} \\ \mathcal{F}\left(\tilde{V}_1, \tilde{V}_2\right) &= \frac{\prod_{k=1}^{n} \sqrt{w_k + \sqrt{w_k^2 - 1}}}{\sqrt[4]{\det\left(\tilde{V}_1 + \tilde{V}_2\right)}}, \end{aligned} \tag{10}$$

where $\tilde{V}$ is equivalent to V, but expressed in the basis $[\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n, \hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n]^T$ and w_k are the eigenvalues of the auxiliary matrix W, defined as

$$\begin{aligned} W &= -2i\Omega^T \left(\tilde{V}_1 + \tilde{V}_2\right)^{-1} \left(\frac{\Omega}{4} + \tilde{V}_2 \Omega \tilde{V}_1\right) \Omega, \\ \Omega &= \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \otimes \mathbb{1}_n. \end{aligned} \tag{11}$$

When we combine the fidelity given in Eq. 10 with the mean vectors and covariance matrices given in Eq. 6, we find that the fidelity between two of Eve's returned states is given by

$$\begin{aligned} &F\left(\rho^0_{\eta,\mu_T}, \rho^{\pi/2}_{\eta,\mu_T}\right) = \\ &\frac{1}{4B} e^{-2\mu_D \omega / B} \left(\sqrt{C} + |4\mu_T\omega + 4\eta(1+\mu_T) - 1|\right), \end{aligned} \tag{12}$$

where

$$\begin{aligned} B &= 2\mu_T\omega + (1+\mu_T)(\omega^2+1)\eta, \\ C &= 16\eta^2(1+\mu_T)^2 + 8\eta(1+\mu_T)\omega(4\mu_T + \omega) \\ &\quad + (1 + 4\mu_T\omega)^2. \end{aligned} \tag{13}$$

Eve wants to choose her parameters ξ_E and μ_D in order to minimise the fidelity (and so maximise the distinguishability) between her returned states. Whilst increasing either of these parameters decreases F, she is not necessarily free to do both simultaneously. Both squeezing and displacement increase the average number of photons in each mode, and we have already established in subsection III A that this is limited by some number N.

Suppose, then, that Eve decides to use pN of her available photons to contribute towards squeezing and $(1-p)N$ towards displacement. Since a squeezing parameter of ξ_E gives an average photon number per mode of $\sinh^2(\xi_E)$, and a displacement parameter of α contributes $|\alpha|^2$ photons, we find that we can do no better that setting the parameters such that $\omega = \cosh\left[\mathrm{arcsinh}\left(2\sqrt{pN}\right)\right]$, $\mu_D = (1-p)N\eta$ for some p.

When we insert these values into Eq. 12, we can investigate the behaviour as a function of p and η for various values of N and μ_T. We find that F is minimised when $p = 0$. This means that Eve is best served by using *all* of her photons to contribute to the displacement of her state. As such, we can now simplify the fidelity, which may be written as

$$F\left(\rho^0_{\eta,\mu_T}, \rho^{\pi/2}_{\eta,\mu_T}\right) = \exp\left(-\frac{\mu_D}{1+2\mu_T}\right). \tag{14}$$

This means that Eve's optimal Gaussian-state attack is one involving coherent states only. This provides a rigorous footing for earlier works which analyse the results of coherent-state attacks with an attenuating defense [18].

One may note that an increase in the average thermal photon number emitted by Alice's apparatus will result in a decrease in Eve's information about the key. One may be tempted to suggest that Alice may try to exploit this and deliberately emit a small amount of noise along with the legitimate photon signal in order to cloud Eve's judgment of θ. This would likely be an unworkable solution, since in order to guarantee security by this method, Alice would have to guarantee that the thermal noise occupies exactly the same mode that is occupied by Bob's photon. When it comes to a continuous mode such as time-bin or frequency, this would be impractical if not impossible.

IV. GENERAL SEPARABLE ATTACKS

We have shown that, amongst multi-mode Gaussian attack states that Eve might use, the separable coherent state is optimal. Whilst it may seem initially surprising that entanglement does not assist her, note that entanglement between any two modes will drop off as more of the signal is attenuated. We are left with a distinguishability that depends only on the average output photon number, μ_D, so Δ does not *explicitly* depend on the transmissivity η.

It may be argued that coherent states are likely to be optimal amongst the separable states, since under loss, photon-counting statistics will tend to be Poissonian [33]. Therefore the best one can hope to do is with a Poissonian state that retains coherence, i.e. a coherent state. However, a state that is initially highly non-Poissonian in its statistics may require a very high attenuation before it approximates a Poissonian distribution, and there is no guarantee that the expression for Δ derived from coherent states will still hold.

In this section, we consider the set-up where Alice defends against a THA by use of an attenuator, and that Eve attacks the system using *any* separable state, but gets back a state with only a few photons. Whilst this seems to be a special case for Eve, note that it is more general that the situation considered in section III since this approach considers a set of states which includes, yet is larger than, the set of coherent states that are optimal within the Gaussian states.

Here, we will consider Eve's input state in its density matrix form instead of its covariance matrix form. We will consider the effect of the attenuator on ρ as a quantum channel, which we express in Kraus operator form:

$$\begin{aligned}\mathcal{E}(\rho) &= \sum_{k=0}^{\infty}\mathcal{E}_k(\rho) \equiv \sum_{k=0}^{\infty}\hat{A}_k\rho\hat{A}_k^\dagger \\ \hat{A}_k &= \sum_{j=k}^{\infty}\sqrt{\binom{j}{k}}\sqrt{\eta}^{\,j-k}\sqrt{1-\eta}^{\,k}\,|j-k\rangle\langle j|\,.\end{aligned} \tag{15}$$

Each term $\hat{A}_k\rho\hat{A}_k^\dagger$ represents k photons being lost from the state ρ, each with independent probability η.

We express each term in the map as follows

$$\begin{aligned}\mathcal{E}_k(\rho) &= \sum_{i,j=k}^{\infty} B_{i,j,k}(\eta)\,\langle i|\,\rho\,|j\rangle\cdot|i-k\rangle\langle j-k| \\ B_{i,j,k}(\eta) &= \sqrt{\binom{i}{k}\binom{j}{k}}\,\eta^{\frac{i+j}{2}-k}(1-\eta)^k.\end{aligned} \tag{16}$$

Since we require a very high level of attenuation to achieve any kind of useful secrecy, we may assume that η is very close to 0. Therefore we may expand the factor $B_{i,j,k}(\eta)$ as

$$B_{i,j,k}(\eta) \approx B_{i,j,k}(0) + \frac{\mathrm{d}B}{\mathrm{d}\eta}\Big|_0\eta + \frac{1}{2}\frac{\mathrm{d}^2B}{\mathrm{d}\eta^2}\Big|_0\eta^2, \tag{17}$$

leading to an expansion of each term in the map as

$$\mathcal{E}_k \approx \mathcal{E}_k^{(0)} + \mathcal{E}_k^{(1)} + \mathcal{E}_k^{(2)}. \tag{18}$$

Using the fact that $\lim_{x\to 0}x^p = \delta_{p,0}$ for $p \geq 0$, we can see that $B_{i,j,k}(0) = \sqrt{\binom{i}{k}\binom{j}{k}}\delta_{\frac{i+j}{2},k}$. Performing the sums over i,j,k we get $\mathcal{E}^{(0)} = |0\rangle\langle 0|\sum_{k=0}^{\infty}\langle k|\,\rho\,|k\rangle = |0\rangle\langle 0|$ (where we have used the fact that $\mathrm{Tr}[\rho]=1$)

In the same way, we find that

$$\mathcal{E}_k^{(1)} = \sum_{i,j=k}^{\infty}\sqrt{\binom{i}{k}\binom{j}{k}}\Bigg[\underbrace{\left(\frac{i+j}{2}-k\right)\delta_{\frac{i+j}{2},k+1}}_{\delta_{i,k}\delta_{j,k+2}+\delta_{i,k+2}\delta_{j,k}+\delta_{i,k+1}\delta_{j,k+1}} - k\delta_{\frac{i+j}{2},k}\Bigg]\eta$$

$$\begin{aligned}\therefore\ \mathcal{E}^{(1)} = &-\mu\,|0\rangle\langle 0| + \\ &\mu\,|1\rangle\langle 1| + \\ &\sum_{k=0}^{\infty}\sqrt{\binom{(k+2)}{k}}\,|0\rangle\langle 2|\,\langle k|\,\rho\,|k+2\rangle + \\ &\sum_{k=0}^{\infty}\sqrt{\binom{(k+2)}{k}}\,|2\rangle\langle 0|\,\langle k+2|\,\rho\,|k\rangle\,,\end{aligned} \tag{19}$$

where $\mu = \eta\sum_{k=0}^{\infty}k\,\langle k|\,\rho\,|k\rangle$ is the average number of photons that Eve receives back after attenuation.

Similarly,

$$\begin{aligned}\mathcal{E}^{(2)} = &\frac{\eta^2}{2}\left(v + \langle\hat{n}\rangle_\rho^2 + \langle\hat{n}\rangle_\rho\right)|0\rangle\langle 0| - \\ &\eta^2\left(v + \langle\hat{n}\rangle_\rho^2 + \langle\hat{n}\rangle_\rho\right)|1\rangle\langle 1| + \\ &\frac{\eta^2}{2}\left(v + \langle\hat{n}\rangle_\rho^2 + \langle\hat{n}\rangle_\rho\right)|2\rangle\langle 2| + \\ &\text{off-diagonals on } \ |0\rangle\langle 2| \ \text{ and } |2\rangle\langle 0| + \\ &\text{terms on } \ |i\rangle\langle j| \ \text{ where } i \text{ or } j \geq 3.\end{aligned} \tag{20}$$

where $v = \langle \hat{n}^2 \rangle_\rho - \langle \hat{n} \rangle_\rho$ is the variance in the initial state. Whilst the diagonal terms can be expressed in terms of the macroscopic observables of ρ, the off-diagonal terms have no such simple expression.

Firstly we should bound the effects of higher-order terms. We do this by supposing that Eve performs a measurement on her returned state to determine whether or not the state contains 2 or fewer photons. That is to say, her measurement of ρ has 2 outcomes corresponding to operators $\hat{E}_\checkmark = |0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 2|$ and $\hat{E}_\times = \sum_{k=3}^{\infty} |k\rangle\langle k|$.

In order to ensure that this measurement does not reduce the information that Eve learns about the state, we say that if she gets the result corresponding to $\hat{E}_\times$ then we assume that she learns the key bit θ perfectly. That is to say, instead of receiving $\mathcal{E}(\rho)$ she can be said to receive some state $|\theta\rangle\langle\theta|$, where $\langle\theta_1|\theta_2\rangle = \delta_{\theta_1,\theta_2}$.

If the measurement is successful, then Eve's state is projected onto the two-or-fewer-photon subspace, giving $\rho_{\text{sub}} = \hat{E}_\checkmark \mathcal{E}(\rho) \hat{E}_\checkmark / \operatorname{Tr}\left[\hat{E}_\checkmark \mathcal{E}(\rho)\right]$. In the case where $\theta = 0$, this may be expressed in the basis of $\{|0\rangle, |1\rangle, |2\rangle\}$ by

$$\rho_{\text{sub}}^{\theta=0} = \begin{bmatrix} 1 - \mu + \frac{\eta^2}{2}\alpha & 0 & \beta \\ 0 & \mu - \eta^2\alpha & 0 \\ \beta & 0 & \frac{\eta^2}{2}\alpha \end{bmatrix}, \tag{21}$$

where $\alpha = v + \langle \hat{n} \rangle_\rho^2 + \langle \hat{n} \rangle_\rho$ and β is some coefficient that cannot be easily expressed in terms of macroscopic properties of the state. In the case where $\theta = \pi/2$ we simply pick up a factor of -1 on the coefficient β. The overall state Eve gets back is then

$$\rho_{\text{returned}}^{\theta} = \operatorname{Tr}\left[\hat{E}_\checkmark \mathcal{E}(\rho)\right] \rho_{\text{sub}}^{\theta} + \operatorname{Tr}\left[\hat{E}_\times \mathcal{E}(\rho)\right] |\theta\rangle\langle\theta|. \tag{22}$$

In order to bound the contribution of the second term, we show in Appendix A that

$$\operatorname{Tr}\left[\hat{E}_\checkmark \mathcal{E}(\rho)\right] \geq e^{-\mu}, \tag{23}$$

and so $\operatorname{Tr}\left[\hat{E}_\times \mathcal{E}(\rho)\right] \leq 1 - e^{-\mu}$.

Since we want to find an upper limit to the information that Eve can learn, we say that she can receive any state that is consistent with both Eq. 21 and the laws of physics. We find that the fidelity between two such density matrices is minimised when the variance v is chosen such that the $|1\rangle\langle 1|$ component is 0 and the off-diagonal terms are maximised. Because of this, it turns out that we do not need to be able to express β in a way relating to macroscopic properties such as average photon number and variance. We simply choose β to be the largest value such that $\rho_{\text{sub}}^{\theta}$ remains positive semi-definite, which is $\beta = \frac{1}{2}\sqrt{\mu(2-\mu)}$.

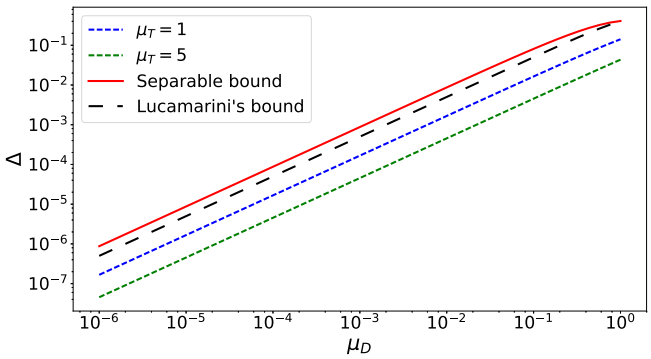

FIG. 3: Upper bounds on the distinguishability, with average returned photon number μ. Dotted lines show the distinguishability for a coherent-state attack with thermal noise, of thermal photon number μ_T. Upper dotted line (blue) shows $\mu_T = 1$, lower dotted line (green) shows $\mu_T = 5$. The black dashed line is the distinguishability for coherent-state attacks found by Lucamarini et. al. Solid line is the bound for separable states, which as expected is always greater that the other bounds.

This, along with Eq. 22 and Eq. 9 gives an ultimate distinguishability bound for the separable attack-state case of

$$\Delta \leq \frac{1 - e^{-\mu}\sqrt{1 - 3\mu(2-\mu)/4}}{2}. \tag{24}$$

Importantly, this is a function of a single variable, that is *measurable by Alice*; the average output photon number. By bounding quantities of Eve's state that cannot be measured, we have ensured that Alice can make an accurate assessment of how secure her QKD system is, whilst not knowing anything about the microscopic details of Eve's state.

Fig. 3 shows that the value of Δ is higher for our separable bound than for the case of a coherent state, whether one diluted by thermal noise or not. We also show it to be higher than the bound on Δ for a noiseless coherent-state attack found by Lucamarini et al. of $\Delta = [1 - e^{\mu}\cos(\mu)]/2$. Whilst our bound on Δ is not absolutely tight (since the 3-photon contributions surely will not convey perfect information of θ), we can see that is not too generous, since it tracks the known achievable bounds quite closely.

V. CONCLUSION

The discovery and implementation of the Trojan Horse Attack once threatened to eliminate the security so famously promised by quantum key distribution. Early seminal works have shown that the situation is not hopeless, and have indicated ways to quantify and abate this threat.

In this work we have fully characterised and quantified the effect of the THA on the key rate under two general attack vectors. We have shown that if Eve uses a multimode Gaussian attack state, her best bet is to use a coherent state. This is true even when there is thermal noise in the channel, although this noise causes a quantifiable decrease in the knowledge that Eve has about any given key bit. We have also quantified the maximum damage on the secrecy that could be caused by Eve using any separable state. We hope that this may be extended to the general entangled case in the future, but we have provided heuristic arguments for why we do not expect much of an improvement for Eve by doing this.

Side-channel attacks cannot be considered only as an afterthought in QKD systems. Even a relatively rudimentary SCA can, if not protected against, hugely reduce the security of a protocol. If we try to improve the security by privacy amplification alone we find that the secret key rate soon drops to zero. This highlights the importance of proper and specific defenses against SCAs that are easily quantified in terms of experimentally accessible quantities.

ACKNOWLEDGMENTS

We wish to acknowledge the contributions of valuable discussions with Mark Pearce and Stefano Pirandola. This research was made possible via the EPSRC Quantum Communications Hub, grant no. EP/M013472/1.

Appendix A: Proof of Eq. 23

We want to show that $\mathrm{Tr}\left[\hat{E}_{\checkmark}\mathcal{E}(\rho)\right] \geq e^{-\mu}$.

First note that since $\hat{E}_{\checkmark} = |0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 2|$, we can assert that

$$\mathrm{Tr}\left[\hat{E}_{\checkmark}\mathcal{E}(\rho)\right] \geq \mathrm{Tr}\left[|0\rangle\langle 0|\,\mathcal{E}(\rho)\right]. \tag{A1}$$

Since $\mathcal{E}$ does not map off-diagonal elements to diagonals and $|0\rangle\langle 0|$ is diagonal, we can consider only the effect of $\mathcal{E}$ on diagonal elements, and say that

$$\begin{aligned}\mathrm{Tr}\left[|0\rangle\langle 0|\,\mathcal{E}(\rho)\right] &= \mathrm{Tr}\left[|0\rangle\langle 0|\,\mathcal{E}\left(\sum_{k=0}^{\infty} p_{k,k}\,|k\rangle\langle k|\right)\right] \\ &= \sum_{k=0}^{\infty} p_{k,k}\,\mathrm{Tr}\left[|0\rangle\langle 0|\,\mathcal{E}(|k\rangle\langle k|)\right],\end{aligned} \tag{A2}$$

where $p_{k,k}$ is the k-th diagonal element of ρ.

We want our ultimate bounds to be in terms of the average photon number of the states. To relate the above quantity to this, we claim that

$$\sum_{k=0}^{\infty} p_{k,k}\,\mathrm{Tr}\left[|0\rangle\langle 0|\,\mathcal{E}(|k\rangle\langle k|)\right] \geq \mathrm{Tr}\left[|0\rangle\langle 0|\,\mathcal{E}\left(|\langle\hat{n}\rangle_{\rho}\rangle\langle\langle\hat{n}\rangle_{\rho}|\right)\right]. \tag{A3}$$

That is to say, the average of the probabilities of losing each of many different photon number states is greater than the probability of losing one state of the average photon number (which we assume without loss of generality to be an integer).

Since a state is mapped to $|0\rangle\langle 0|$ if and only if it loses all of its photons, we can say that

$$\mathrm{Tr}\left[|0\rangle\langle 0|\,\mathcal{E}(|k\rangle\langle k|)\right] = (1-\eta)^k, \tag{A4}$$

We will consider first the case of ρ being a mixture of only two Fock states, with weightings p and $1-p$ and respective photon numbers n and m. By using Eq. A4, Eq. A3 then becomes

$$p(1-\eta)^n + (1-p)(1-\eta)^m \geq (1-\eta)^{pn+(1-p)m}. \tag{A5}$$

If we let $n = m + \delta$, then this simplifies to

$$-y^p + py - p + 1 \geq 0, \tag{A6}$$

where we have used $y \equiv (1-\eta)^{\delta}$. The claim then reduces to proving that this polynomial is satisfied for all $y, p \in [0,1]$.

Let $f(p) = -y^p + py - p + 1$. We have $f(0) = f(1) = 0$. This function has a unique stationary point between 0 and 1, and the curvature $= -\left[\log(y)\right]^2 y^p$ is everywhere negative. Therefore $f(p) \geq 0$. This proves the claim for a bimodal initial state. The general claim follows by induction.

We now have that

$$\mathrm{Tr}\left[\hat{E}_{\checkmark}\mathcal{E}(\rho)\right] \geq (1-\eta)^{\langle\hat{n}\rangle_{\rho}} = \left(1 - \frac{\mu}{\langle\hat{n}\rangle_{\rho}}\right)^{\langle\hat{n}\rangle_{\rho}}. \tag{A7}$$

Since the average *input* photon number is generally of the scale of dozens of orders of magnitude above unity, we may confidently take the limit of $\langle\hat{n}\rangle_{\rho} \to \infty$, which reduces Eq. A7 to Eq. 23.

[1] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, Reviews of Modern Physics **81**, 1301 (2009).

[2] P. W. Shor, SIAM Journal on Computing **26**, 1484 (1997).

[3] R. Renner, International Journal of Quantum Informa-

tion **6**, 1 (2008).
[4] E. Biham and T. Mor, Physical Review Letters **79**, 4034 (1997).
[5] H. Inamori, N. Lütkenhaus, and D. Mayers, European Physical Journal D **41**, 599 (2007).
[6] C. H. Bennett and G. Brassard, International Conference on Computers, Systems & Signal Processing **1**, 175 (1984).
[7] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nature photonics **4**, 686 (2010).
[8] A. Lamas-Linares and C. Kurtsiefer, Optics express **15**, 9388 (2007).
[9] S. Nauerth, M. Fürst, T. Schmitt-Manderbach, H. Weier, and H. Weinfurter, New Journal of Physics **11**, 065001 (2009).
[10] L. C. Comandar, M. Lucamarini, B. Fröhlich, J. F. Dynes, A. W. Sharpe, S. W. Tam, Z. L. Yuan, R. V. Penty, and A. J. Shields, Nature Photonics **10**, 312 (2016).
[11] S. L. Braunstein and S. Pirandola, Physical Review Letters **108**, 1 (2012).
[12] I. Gerhardt, Q. Liu, A. A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Nature Communications **2** (2011), 10.1038/ncomms1348.
[13] F. Xu, B. Qi, and H. K. Lo, New Journal of Physics **12** (2010), 10.1088/1367-2630/12/11/113026.
[14] Y. Zhao, C. H. F. Fung, B. Qi, C. Chen, and H. K. Lo, Physical Review A - Atomic, Molecular, and Optical Physics **78**, 1 (2008).
[15] X. B. Wang, Physical Review Letters **94**, 1 (2005).
[16] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, Physical Review A **73**, 022320 (2006).
[17] F.-G. Deng, X.-H. Li, H.-Y. Zhou, and Z.-j. Zhang, Physical Review A **72**, 044302 (2005).
[18] M. Lucamarini, I. Choi, M. B. Ward, J. F. Dynes, Z. L. Yuan, and A. J. Shields, Physical Review X **5**, 1 (2015).
[19] N. Jain, B. Stiller, I. Khan, V. Makarov, C. Marquardt, and G. Leuchs, IEEE Journal of Selected Topics in Quantum Electronics **21**, 168 (2015).
[20] N. Jain, E. Anisimova, I. Khan, V. Makarov, C. Marquardt, and G. Leuchs, New Journal of Physics **16**, 123030 (2014).
[21] S. Sajeed, I. Radchenko, S. Kaiser, J. P. Bourgoin, A. Pappa, L. Monat, M. Legré, and V. Makarov, Physical Review A - Atomic, Molecular, and Optical Physics **91**, 1 (2015).
[22] K. Tamaki, M. Curty, and M. Lucamarini, New Journal of Physics **18**, 065008 (2016).
[23] D. Gottesman, H.-K. Lo, N. Lutkenhaus, and J. Preskill, in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on* (IEEE, 2004) p. 136.
[24] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information* (AAPT, 2002).
[25] T. J. Driscoll, J. M. Calo, and N. M. Lawandy, Optics letters **16**, 1046 (1991).
[26] C. W. Carr, H. B. Radousky, and S. G. Demos, Physical Review Letters **91**, 127402 (2003).
[27] S. Lloyd, Science **321**, 1463 (2008).
[28] S. L. Braunstein, Physical Review A - Atomic, Molecular, and Optical Physics **71**, 8 (2005).
[29] M. Lasota, R. Filip, and V. C. Usenko, **062312**, 1 (2017).
[30] R. J. Glauber, Physical Review **131**, 2766 (1963).
[31] This may be seen by considering Ref. [23], section VIII. The purified states corresponding to each basis are as defined in Ref. [18], Appendix B. From these it may be seen that $1-2\Delta$ is equal to the *average* fidelity between a state being emitted in the X basis and one in the Y basis. By the symmetry and unitary invariance of the fidelity function, this reduces to finding the fidelity between only the states corresponding to $\theta = 0$ and $\theta = \pi/2$.
[32] L. Banchi, S. L. Braunstein, and S. Pirandola, Physical Review Letters **115**, 1 (2015).
[33] Y. Hu, X. Peng, T. Li, and H. Guo, Physics Letters, Section A: General, Atomic and Solid State Physics **367**, 173 (2007).