Game Theoretic Security Framework for Quantum Key Distribution

Walter O. Krawec[1(✉)] and Fei Miao[1,2]

[1] Department of Computer Science and Engineering, University of Connecticut, Storrs, CT 06028, USA
walter.krawec@uconn.edu

[2] Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06028, USA
fei.miao@uconn.edu

Abstract. In this paper, we propose a game-theoretic model of security for quantum key distribution (QKD) protocols. QKD protocols allow two parties to agree on a shared secret key, secure against an adversary bounded only by the laws of physics (as opposed to classical key distribution protocols which, by necessity, require computational assumptions to be placed on the power of an adversary). We investigate a novel framework of security using game theory where all participants (including the adversary) are rational. We will show that, in this framework, certain impossibility results for QKD in the standard adversarial model of security still remain true here. However, we will also show that improved key-rate efficiency is possible in our game-theoretic security model.

Keywords: Quantum cryptography · Game theory · Security

1 Introduction

Quantum key distribution (QKD) protocols allow for the establishment of a shared secret key between two parties, referred to as Alice (A) and Bob (B), which is secure against an all-powerful adversary, customarily referred to as Eve (E). Such a task is impossible to achieve when using only classical communication; indeed, when parties have access only to classical resources, key-distribution is only secure if certain computational assumptions are made on the power of the adversary. With QKD protocols, however, the only required assumption is that the adversary is bounded by the laws of physics. Furthermore, QKD is a practical technology today with several experimental and commercial demonstrations. For a general survey of QKD protocols, the reader is referred to [1].

In general, most QKD protocols are designed, and their security proven, within a *standard adversarial model of security*. In this case, parties A and B run the protocol with the goal of establishing a shared secret key. An all-powerful adversary sits in the middle of the channel, intercepting, and probing, each quantum bit (or *qubit*) sent from A to B. As is standard in this usual model of

L. Bushnell et al. (Eds.): GameSec 2018, LNCS 11199, pp. 38–58, 2018.
https://doi.org/10.1007/978-3-030-01554-1_3

cryptography, it is assumed that E is simply malicious and has no motivation to attack, nor does E "care" about the cost of attacking.

In this paper, we investigate the use of *game theory* to study the security of QKD protocols. While we are not the first to propose a game theoretic analysis of cryptographic protocols (quantum or otherwise - see the "Related Work" section below for a summary), we propose a more general-purpose model which can be applied to arbitrary QKD protocols. Compared with prior work, our new approach is more general and, most importantly, allows for meaningful key-rate and noise tolerance computations to be performed which are vital when considering QKD security and comparing benefits of distinct protocols.

Beyond introducing our model, we also apply it to analyze certain important QKD protocols against both all-powerful, quantum, attacks and also more practical attacks based on current-day technology. For each, we compute the critical noise tolerance values and compare with the standard adversarial model. We also discuss the efficiency of the resulting protocols in our model. *Such computations were not possible in prior work, applying game theory to QKD thus showing the significance of our new methods.* We stress that this work's prime contribution is to develop a general framework for the modeling of QKD security, and various important computations involving these protocols (namely, key-rate computations and noise tolerances), through the use of game theory. We expect this work to be the foundation of future significant developments both in the fields of quantum key distribution, and also in game theory. Furthermore, our rational model of security may lead to more efficient secure communication systems as we discuss in the text.

1.1 Related Work

Game theory has seen great success when applied to *classical* cryptography (see [2] for a general survey). It has also raised a lot of interest recently in the study of Cyber-Physical System (CPS) security problems and network security [3–6].

Only recently have there been attempts and interest in applying game theory to *quantum* cryptography. Outside of key-distribution (our subject of interest in this paper), game theory has been used for secret sharing [7], rational state sharing [8], bit commitment [9], certain function computations [10], and secure direct communication [11].

The prior work discussed above all involve cryptographic primitives very different from QKD. However, some attempt has been made recently to apply game theory to QKD. In [12] a cooperative game was used to establish a quantum network consisting of point-wise QKD links which could relay information from one node to the other. However, QKD was only used as a tool in their work, the primary motivation for using game-theory was for the nodes to construct an optimal network topology in a vehicular network.

Closest to our work are [13,14]. In [13], game theory was used to analyze the BB84 QKD protocol. Their model, however, only considered strategies affecting certain choices within the protocol. In their work, a three-party game was constructed (consisting of A, B, and the adversary E). The strategy space of each

participant was to chose a *basis* (either Z or X) to send and receive quantum bits in (we will discuss quantum measurement in the next section). The goal of the parties A and B was to *detect* E while the goal of E was to avoid detection. There was no goal of establishing an actual secret key at the end of maximal length; furthermore, E did not have a goal of learning information on the key. Both of these goals will be incorporated in our more general model.

In the recently published work of [14], the model proposed in [13] was extended and applied to the so-called *Ping-Pong* protocol [15] and also the LM05 protocol [16]. Their work considered certain attacks E may perform against the system which were previously proposed in the literature against the ping-pong protocol. The strategy space for A and B (now considered one party in their work) consisted of choosing to run the protocol, or a variant of it (there was no choice to simply "abort" which is an important choice in QKD security [1]). The goal of E was to maximize her information on the final raw-key while avoiding detection; the goal of the party "AB" was to maximize their mutual information. Our model will also consider these two as goals; however we will not be concerned about probability of detection (which, in typical applications, is not a concern as there is always natural noise in the channel anyway). However, we will go beyond this by also setting a goal to maximize the efficiency of the protocol. Furthermore, the model we introduce in this work allows for critical key-rate and noise tolerance computations, not possible in prior work.

1.2 Notation and Definitions

We use $H(X)$ to denote the Shannon entropy of random variable X. In particular, if $P(X = x) = p_x$, then $H(X) = -\sum_x p_x \log p_x$, where all logarithms in this paper are base two unless otherwise stated. By $h(x)$ we mean the binary entropy function defined $h(x) = -x \log x - (1-x)\log(1-x)$. Given two random variables X and Y, then $H(XY)$ is the joint Shannon entropy of random variables X and Y defined in the usual way. $H(X|Y)$ denotes the conditional entropy defined $H(X|Y) = H(XY) - H(Y)$. By $I(X : Y)$ we mean the mutual information between X and Y, defined to be $I(X : Y) = H(X) + H(Y) - H(XY)$.

We assume a familiarity with game theory, and include the following definitions only for completeness. Given a tuple $q = (q_1, \cdots, q_n)$ we write q_{-i} to mean the $n-1$ tuple consisting of all q_j for $j \neq i$; i.e., $q_{-i} = (q_1, \cdots, q_{i-1}, q_{i+1}, \cdots, q_n)$.

Definition 1. *An n-player normal (strategic) form game G is an n-tuple $\{(S_1, u_1), \ldots, (S_n, u_n)\}$, where for each i,*

- *S_i is a nonempty set, called i's strategy space, and*
- *$u_i : S \to \mathbb{R}$ is called i's utility function, where $S = S_1 \times \cdots \times S_n$.*

Definition 2. *Dominant Strategy (DS). A strategy s'_i (weakly) dominates s''_i, if $\forall s_{-i} \in S_{-i}$, $u_i(s'_i, s_{-i}) \geq u_i(s''_i, s_{-i})$, and $\exists s'_{-i} \in S_{-i}$, $u_i(s'_i, s'_{-i}) > u_i(s''_i, s'_{-i})$.*

Definition 3. *Strict Nash Equilibrium (NE). $s^* \in S$ is a strict Nash equilibrium of $G = \{(S_1, u_1), \ldots, (S_n, u_n)\}$ if for all i and for all $s_i \in S_i$, $u_i(s^*_i, s^*_{-i}) > u_i(s_i, s^*_{-i})$.*

2 Quantum Communication and Cryptography

For completeness, we review some basic concepts in quantum key distribution and quantum communication. Due to length constraints, this section is necessarily short; however, the interested reader is referred to [17].

A *quantum bit* or *qubit* is modeled, mathematically, as a normalized vector in $\mathbb{C}^2$. More generally, an arbitrary n-dimensional quantum state may be modeled as a normalized vector in $\mathbb{C}^n$. Quantum states are typically denoted as "kets" of the form $|\psi\rangle$ where the ψ can be replaced with any arbitrary label. The inner product of two kets $|\psi\rangle$ and $|\phi\rangle$ is denoted $\langle\psi|\phi\rangle$.

The *measurement postulate* of quantum mechanics gives rules for how quantum states may be observed. We are interested only with *projective measurements* in this work. Let $\mathcal{B} = \{|v_1\rangle, \cdots, |v_n\rangle\}$ be an orthonormal basis of $\mathbb{C}^n$. Then, given a quantum state $|\psi\rangle \in \mathbb{C}^n$, after measurement in basis $\mathcal{B}$, one observes basis state $|v_i\rangle$ with probability $|\langle v_i|\psi\rangle|^2$. Note, therefore, that measurements are probabilistic processes and the outcome and distribution depends on the basis one performs a measurement in. For qubits, two common bases are the Z basis, denoted $\{|0\rangle, |1\rangle\}$ and the X basis, denoted $\{|+\rangle, |-\rangle\}$ where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Note that, once observed, the original quantum state is destroyed and "collapses" to the observed basis state. In theory, one may perform a measurement in any basis. Note, also, that the *No Cloning Theorem* prevents the exact duplication of an unknown quantum state. Thus, when used as communication resources, an adversary is forced to attack immediately (she cannot copy the qubits to attack later); furthermore, if she attempts to extract information from the qubits via a measurement, this may cause disturbances that may be detected by honest users later. (Measurements are not the only way E can attack a qubit - however, for understanding our work in this paper, measurements are sufficient.)

2.1 Quantum Key Distribution

Quantum key distribution takes advantage of certain properties unique to quantum mechanics to allow for the establishment of a shared secret key between A and B, secure against an all powerful adversary E, a task impossible to achieve with classical communication only. There are many different QKD protocols at this point, with the first being discovered in 1984 now known as the BB84 protocol [18]. Another important protocol, discovered in 1992, is the B92 protocol [19]. The basic operation of these protocols is shown in Protocols 1 and 2. It is important to note that, in addition to a quantum channel, allowing for the transmission of qubits from A to B, there is also an *authenticated classical channel* connecting the two users. This channel is *not secret*, however, so any message sent from A to B can be read by the attacker (though, the attacker cannot write on this channel). Authentication may be done in an information theoretic manner assuming the existence of an initial (small) secret key. Thus, QKD protocols are sometimes referred to as *quantum key expansion* protocols as, technically,

they require an initial shared secret key which they will then expand through the use of quantum communication.

In general, QKD protocols consist of a *quantum communication stage* followed by a classical reconciliation stage. The first stage utilizes, *through multiple iterations* the quantum and authenticated channels to produce a *raw key* - a string of classical bits that is partially correlated and partially secret. If the error rate is "low enough" (which depends on the protocol *and security model*), the second stage is employed which consists of an error-correcting protocol (done over the authenticated channel, thus leaking information to E "for free") and a privacy amplification protocol, yielding a *secret key*. The size of the secret key is directly correlated with the noise in the quantum channel and the amount of information an adversary potentially has on the raw-key. The more information the adversary has and the more noise, the smaller the secret key will be. In the standard adversarial model of security, the noise is assumed to be the product of the adversary's attack and the two are directly correlated; in fact, one important aspect of QKD research is to determine a protocols *maximally tolerated noise level*, that is the value of noise for which QKD is possible against a malicious adversary.

For more details on all these concepts, the reader is referred to [1].

Protocol 1. BB84 [18]

Public Knowledge: A key-bit "0" is encoded as a qubit $|0\rangle$ or $|+\rangle$ while a key-bit of "1" is encoded as $|1\rangle$ or $|-\rangle$.

Quantum Communication Stage (Repeat for N iterations):

1. A chooses a random key bit $k_A \in \{0, 1\}$ and a random basis $b_A \in \{Z, X\}$. She sends the encoding of k_A as a qubit using her randomly chosen basis.
2. B chooses a random basis $b_B \in \{Z, X\}$ and measures in that basis.
3. A and B share, over the authenticated channel, their choice b_A and b_B respectively. Parties only keep those iterations where $b_A = b_B$. All other results are discarded (approximately half should remain).

Protocol 2. B92 [19]

Public Knowledge: A key-bit "0" is encoded as a qubit $|0\rangle$ while a key-bit of "1" is encoded as a qubit $|+\rangle$.

Quantum Communication Stage (Repeat for N iterations):

1. A chooses a random key bit $k_A \in \{0, 1\}$ and prepares an appropriate qubit for B.
2. B chooses a random basis $b_B \in \{Z, X\}$ and measures in that basis. If $b_B = Z$ and he observes $|1\rangle$, the B sets $k_B = 1$. If $b_B = X$ and he observes $|-\rangle$, then B sets $k_B = 0$. All other results are considered "inconclusive."
3. B informs A, over the authenticated channel, which iterations he considered "inconclusive." All inconclusive iterations are discarded. It is expected that one-quarter of the iterations will remain.

3 Game Theoretic Model

We now introduce our game theoretic security model for QKD. While in practice, A and B are two separate entities, in our game theoretic model, we will consider them as one party which we denote by AB. We therefore consider a two-party game consisting of player AB and player E. The goal of party AB is to establish a long, secret key shared between each other. The goal of E is to limit the length of the final secret key. Since it is trivial for E to cause a denial-of-service attack in a point-to-point communication protocol (of which QKD is one), we limit E's strategy space to consisting of attacks which induce less noise than some maximal value Q which AB advertise as tolerating. This parameter Q can also represent certain "natural" noise in the quantum channel - AB will abort if the noise exceeds this value, thus if E attacks, she must "hide" in the natural noise. Noise for us is defined to be the average probability of a $|i\rangle$ flipping to a $|1-i\rangle$ and a $|\pm\rangle$ flipping to a $|\mp\rangle$. One key interest to us will be for what values of Q, QKD is possible in our game theoretic model and compare this with the standard adversarial model.

Beyond these goals, there are costs for using certain quantum (and potentially also classical) resources. For AB sending and receiving qubits can be a costly activity. Thus, though AB wish to establish a key, if doing so is "too expensive" they may wish to simply "abort" and do nothing. On the other hand, to gain information is E's goal (as this limits the size of the secret key), however attacking the quantum channel is a costly activity and extracting maximal information may require expensive quantum memory systems. Thus, it is the goal of our framework to construct a protocol (game strategy) where it is in AB's interest to run the protocol (and not abort), while it is in E's interest not to perform a complicated attack against it. Passive attacks (as opposed to more powerful quantum attacks) can greatly increase the efficiency of the protocol as we will see. *Thus, if users employ the rational model of security for QKD, more efficient quantum communication may be possible.*

One may consider applying our model to classical key-distribution (for instance, by using a hard problem that takes a large amount of classical resources to break); however this problem scenario, and the rewards for attacking, are very different from the quantum case. In a QKD protocol, the generated key is information theoretically secure, thus, for example, any message encrypted using the produced key is perfectly secret for all time. However, if a classical key-distribution system is used, an adversary may copy all communication sent by the protocol and attack offline; eventually if the system is broken, that adversary can learn all messages encrypted with that key. This is a very powerful motivating factor for an adversary. Contrast this with QKD: first, the adversary cannot attack offline and must attack actively. Furthermore, the adversary cannot learn the produced secret key nor any message encrypted with it.

We formulate our game-theoretic security model as follows. Let Σ_{AB} be a set of strategies (i.e., *protocols*) which party AB may choose to run and let Σ_E be the set of strategies (i.e., *attacks*) which party E may choose to employ against

AB. We always assume the "do nothing" strategy (denoted I_{AB} for AB and I_E for E) is an option for either party. (We use I for "identity operation.")

Now, in reality, player AB actually consists of two separate entities, thus it is important to ensure that our game-theoretic model can actually be employed in practice. In particular, we must ensure that A and B can agree on a strategy in a way that makes sense. There are many ways to achieve this; one in particular is they can sacrifice some of their initial shared secret key to send a constant-length message encrypted with one-time-pad (this message is the protocol to use). As mentioned earlier (see Sect. 2.1), for the authentication channel to work, A and B must begin with some shared random key already. They may use a constant amount $c = \log_2 |\Sigma_{AB}|$ to send, with perfect secrecy, the choice of protocol. So long as c is not a function of the number of iterations used in the quantum channel (which it is not), there is no contradiction to the key-expansion properties of QKD. Note that one cannot use this shared initial key to send securely a longer classical key - it can only be used for a small, constant, amount of initial communication such as picking from a small subset of strategies.

There are other ways for A and B to agree on a strategy, however, we may safely assume that party AB, though two distinct entities separated physically, may, at the start, agree on a single protocol to use from the set of allowed strategies Σ_{AB}. Note that a *mixed strategy* may also be agreed on by having A choose a random protocol and sending the choice, securely, to B.

Let $Q \in [0, .5]$ be the maximal noise level in the channel which is publicly known to both players before the game begins (alternatively, Q may be a value set by AB that is the "maximal tolerated" noise allowed in the channel, either naturally or artificially). Thus, even if E chooses not to attack (i.e., she chooses strategy I_E), she will still learn something about the raw key without incurring any costs (due to the information leaked by error correction). However, if she wishes to learn *more* (causing the secret key length to drop further) she must choose to attack the channel. We will assume that this attacker, if she chooses to attack, is able to replace the noisy quantum channel with an ideal one and then *hide* the noise her attack inevitably creates within this natural noise parameter Q. Such an operation (attacking, and setting up her equipment to hide within the natural noise) will be potentially expensive, though she will gain more information on the raw key thereby decreasing the efficiency of AB, her goal.

After running their respective protocol $\Pi \in \Sigma_{AB}$ (which includes running a quantum communication stage for N iterations, followed by error correction and privacy amplification), with E attacking using attack $\mathcal{A} \in \Sigma_E$, each party is given a utility for the outcome of the game. The outcome of the game for party AB is a function of the resulting *secret key length* (i.e., after error correction and privacy amplification), denoted M along with the cost of running the chosen strategy (denoted, $C_{AB}(\Pi)$). For our analysis, we will assume the utility is a simple linear function of the form:

$$u_{AB}(M, C_A(\Pi)) = w_g^{AB} M - w_c^{AB} C_{AB}(\Pi).$$

where w_g^{AB} and w_c^{AB} are non-negative weights for the "gain" and "cost" respectively of AB's utility. We will assume that these weights are simply 1.

For E, her utility is a function of the information she learned on the error-corrected raw key (before privacy amplification but after error correction) and the cost of running her chosen attack. Let K be the information E learns on the raw-key and $C_E(\mathcal{A})$ the cost of attack $\mathcal{A} \in \Sigma_E$. Then, her utility will be:

$$u_E(K, C_E(\mathcal{A})) = w_g^E K - w_c^E C_E(\mathcal{A}).$$

where w_g^E and w_c^E are non-negative weights for the "gain" and "cost" of E's utility. As with u_{AB}, we will assume that these weights are simply 1.

The reader may wonder what E's rational motivation would be for learning information about the raw-key (before privacy amplification) when it is the *secret key* (after privacy amplification) that is actually used by A and B later to, for example, encrypt messages. First, note that if we define the model to be E gains utility for learning information on the secret key, by the very definition of privacy amplification, her gain would be negligible (and, in the asymptotic scenario, we may even say it would be zero); thus this could never motivate her. On the other hand, we could not give her utility for causing A and B to simply abort due to high noise levels above Q as this is a form of *denial of service* attack which would cost E little to nothing to execute (E can simply cut the quantum channel!) and is a weakness for any point-to-point communication system, especially QKD. Thus, since gaining information on the secret key is not possible, and since a denial of service attack is outside the scope of the model, E's goal is to *minimize the key-rate of the protocol* (i.e., minimize its efficiency). Since the more information E has on the *raw-key* the smaller the secret key will be (after privacy amplification), it is E's goal to increase her information (thus shrinking the size of the final secret key) while minimizing her cost and staying below the natural noise level of Q.

We will use $U_{AB}(\Pi, \mathcal{A})$ to denote the expected utility given to player AB if that player chooses strategy $\Pi \in \Sigma_{AB}$ and if E chooses strategy $\mathcal{A} \in \Sigma_E$. $U_E(\Pi, \mathcal{A})$ is defined similarly for E.

The goal in our game-theoretic model of QKD security is to construct a protocol (strategy) "Π" such that the joint strategy (Π, I_E) is a strict Nash equilibrium (NE). In particular AB are motivated to actually run the protocol while E is motivated to not launch a complicated quantum attack against it. If such a protocol exists, then, under the assumption of a rational adversary, that adversary will choose not to implement a powerful quantum attack as it will be too expensive. This security model guarantees that if AB and E are rational, then, assuming the protocol is a strict NE, the resulting key is information theoretically secure. In the standard adversarial model the key is also information theoretically secure, however the effective key-rate will be lower after privacy amplification as one must "remove" E's additional information from her quantum attack. Thus, by assuming rational adversaries, one still maintains information theoretic security, but with greater communication efficiency.

In this work, we will consider standard QKD protocols (such as BB84 [18]) and add to these protocols additional "decoy" iterations. These decoy iterations will be, during the operation of the protocol, completely indistinguishable from standard iterations. At the end of the game (protocol run), AB will announce which iterations were "real" and which were decoys. Decoy iterations, which are useless to both parties, cost AB resources as they must still prepare and measure qubits (if they do not send qubits, this is distinguishable to E and she will know it is a decoy). However, since E cannot tell which are the decoy iterations, she is forced to attack them all the same, thus costing her resources also. If E's attack is very expensive (e.g., requires an expensive quantum memory to operate), then the more decoy iterations there are, the less incentive she will have to attack at all. Of course, the more decoy iterations there are, the less incentive AB will have to run the protocol as it will become too expensive for too little reward.

To incorporate this decoy method, we will introduce a parameter $\alpha \in [0, 1]$ which may be set by AB. On any iteration of a protocol, during the quantum communication stage, AB (in practice, just party A) will decide whether this iteration is a real iteration (with probability α) or a decoy iteration (with probability $1-\alpha$), however they run the iteration normally regardless so that E cannot distinguish the two cases. At the conclusion of the protocol, all decoy iterations are discarded (to achieve this in practice, A will transmit, at the conclusion of the protocol, through the authenticated classical channel, which iterations were decoys - thus E also learns this at the end of the game, *but at that point, she already used resources to attack*; furthermore, properties such as the No-Cloning Theorem, prevent her from making copies of qubits and later changing her attack based on this new knowledge). A protocol strategy, therefore, will be denoted $\Pi^{(\alpha)}$. Ultimately, the goal within this game-theoretic model is to find a value for α such that the joint strategy $(\Pi^{(\alpha)}, I_E)$ is a strict NE. Furthermore, we wish to determine what values of Q allow for an α to exist and to determine the efficiency of the resulting protocol.

3.1 All-Powerful Attacks Against BB84

In this section, we apply our framework to model security of the BB84 protocol allowing E the ability to launch all-powerful attacks (e.g., attacks requiring quantum memories). We will prove that the noise tolerance of the BB84 protocol in our game theoretic framework remains 11%, the same as in the standard adversarial model [20]. However, we will show that, for noise levels less than 11%, the efficiency of the protocol can be substantially higher in our game theoretic model than in the standard adversarial model.

We will consider the BB84 protocol parameterized by α, denoted here as $\Pi^{(\alpha)}_{BB84}$. We will consider what is required for $(\Pi^{(\alpha)}_{BB84}, I_E)$ to be a strict Nash equilibrium. First, consider AB's utility for this strategy; we assume N is the number of iterations they run the protocol for. In this case, since E is not attacking, after error correction and privacy amplification, the secret key will be of expected length $\frac{N\alpha}{2}(1 - h(Q))$ (recall, in BB84, only half the iterations are expected to be kept - see Protocol 1). Thus:

$$U_{AB}(\Pi^{(\alpha)}_{BB84}, I_E) = \frac{N}{2}\alpha(1 - h(Q)) - C_{AB}, \tag{1}$$

were we use C_{AB} to mean $C_{AB}(\Pi^{(\alpha)}_{BB84})$ (a value that AB must decide on, though its actual numerical value will not be important to us in this section). On the other hand, we have $U_{AB}(I_{AB}, I_E) = 0$. Thus, for a strict NE to exist, we require:

$$\alpha > \frac{2C_{AB}}{N(1 - h(Q))}.$$

Naturally, this requires $1 - h(Q) > \frac{2}{N}C_{AB}$. Thus, if this expression cannot be satisfied, then the natural noise in the quantum channel (denoted Q) is too great and AB cannot justify the cost of running the protocol. In the following analysis, we will assume this inequality is satisfied.

Let us now consider E's expected utility. If E does not attack (i.e., she chooses to play strategy I_E), then, since we are also considering "natural noise" in the channel at a rate of Q, party E will gain $\frac{N\alpha}{2}h(Q)$ bits of information on AB's raw key "for free" simply by listening in to the authenticated classical channel (we are assuming optimal error-correcting). Thus her expected utility is: $U_E(\Pi^{(\alpha)}_{BB84}, I_E) = \alpha\frac{N}{2}h(Q)$.

Now, assume that E chooses an optimal quantum attack strategy $\mathcal{A} \in \Sigma_E$. From this, she will gain more information on the raw key (thus shrinking the final secret key size, her ultimate goal), though it also will cost something to implement. Furthermore, she will waste resources on attacking decoy states. It is known that $I(A : E) = \frac{\alpha N}{2}h(Q)$ when E performs an optimal attack [1]. Thus, her utility (based on $I(A : E)$ and *also the information learned from error correction*) is:

$$U_E(\Pi^{(\alpha)}_{BB84}, \mathcal{A}) = I(A : E) + \alpha\frac{N}{2}h(Q) - C_E(\mathcal{A}) = \alpha N h(Q) - C_E(\mathcal{A})$$

Thus, to be a strict NE, we require $U_E(\Pi^{(\alpha)}_{BB84}, I_E) > U_E(\Pi^{(\alpha)}_{BB84}, \mathcal{A})$. For this inequality to hold it must be that: $\alpha < \frac{2C_E(\mathcal{A})}{Nh(Q)}$. Thus, for the strategy $(\Pi^{(\alpha)}_{BB84}, I)$ to be a strict NE, we require an α to exist that satisfies the following inequalities:

$$\frac{2C_A}{N(1 - h(Q))} < \alpha < \frac{2C_E(\mathcal{A})}{Nh(Q)}. \tag{2}$$

If such an α exists, and if AB choose that for their decoy state probability, they can be assured, in our rational model of security, that E will prefer to not attack the quantum channel but instead, simply eavesdrop on the authenticated channel. Furthermore, with such an α, rational AB are also motivated to run the protocol, as opposed to simply aborting.

To determine suitable values for α we require values for C_{AB} and $C_E(\mathcal{A})$. Let's assume a worst-case scenario in that $C_{AB} = C_E(\mathcal{A})$. Note that, to implement $\mathcal{A}$ in practice, E must somehow cut into the quantum channel, replace the natural noise with a more precise channel, setup attack equipment, and, in

this scenario where $I(A:E) = h(Q)$, construct and operate a perfect quantum memory. In reality, it seems reasonable to expect that $C_E(\mathcal{A}) > C_{AB}$. Thus, making these equal models a "worst-case" scenario of benefit to E.

Now, by assumption, we have $\frac{2}{N}C_{AB} < 1 - h(Q)$ (i.e., the cost per-bit for AB is less than $(1 - h(Q))/2$; if this assumption is not made, then AB have no motivation to run the protocol). Thus, the left-hand-side of Eq. 2 is strictly less than 1 and, so, a solution for α exists only if the following inequality is satisfied:

$$\frac{C_A}{1 - h(Q)} < \frac{C_E(\mathcal{A})}{h(Q)}.$$

Since we are assuming in this section that $C_{AB} = C_E(\mathcal{A})$, then $(\Pi^{(\alpha)}_{BB84}, I_E)$ is a strict NE only if the noise in the channel Q satisfies the following inequality:

$$1 - 2h(Q) > 0. \tag{3}$$

This is exactly the same noise tolerance bound as is derived in the standard adversarial model for BB84 as reported in [1,20,21]! In particular, a solution for α exists only if $Q \leq 11\%$.

However, despite the noise tolerance threshold being the same in our new game-theoretic model and the standard adversarial model, our game theoretic model may be used to gain a significantly improved key-rate as we now demonstrate. Assume that $Q \leq 11\%$ (and so $1 - 2h(Q) > 0$ and thus an α exists). Let α be the largest allowed by Eq. 2 (the higher α is, the better for AB as the more "real" iterations are being used on average). We may thus set:

$$\alpha = \min\left(\frac{2c_E(\mathcal{A})}{Nh(Q)} - \epsilon, 1 - \epsilon\right),$$

for some small $\epsilon > 0$. Since we are assuming $C_E(\mathcal{A}) = C_{AB}$ and we also require $\frac{2}{N}C_A < 1 - h(Q)$, we may write $C_E = \frac{\gamma}{2} \cdot N(1 - h(Q))$ for some constant $\gamma < 1$ and thus we have:

$$\alpha = \min\left(\gamma\frac{1 - h(Q)}{h(Q)} - \epsilon, 1 - \epsilon\right). \tag{4}$$

With α chosen as this, it is in E's interest to not attack, but to instead only gain the free information from the error-correction due to the natural noise level Q. In this case, the Csiszar-Korner bound [22] applies (as E no longer has a quantum system, but a classical one) which gives us a secret key size, after privacy amplification and error correction, of:

$$\ell_{GT}(N) = \alpha\frac{N}{2}(1 - h(Q)) = \frac{N}{2}\min\left(\gamma \cdot \frac{(1 - h(Q))^2}{h(Q)}, (1 - \epsilon)(1 - h(Q))\right). \tag{5}$$

On the other hand, in the standard adversarial model for a noise level of Q, the secret key size would be: $\ell_{SAM}(N) = \frac{N}{2}(1 - 2h(Q))$. Discounting the ϵ term (which may be made arbitrarily close to 0), we plot the conditional key-rate of the BB84 protocol in both our new game theoretic model and the standard adversarial model (i.e., we plot $2\ell_{GT}(N)/N$ and $2\ell_{SAM}(N)/N$ respectively) in Fig. 1.

Note that, even though the noise tolerance is the same in both security models, our game-theoretic security model may provide a much higher key-rate (i.e., efficiency) depending on the cost C_{AB} (i.e., γ). *Thus, by using a game-theoretic model of security, more efficient quantum secure communication systems may be employed!*

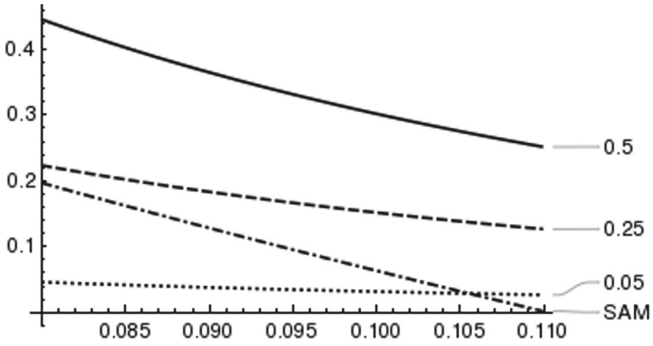

Fig. 1. Showing the key-rate of the BB84 protocol in the Standard Adversarial Model (SAM) compared with our game-theoretic model at high noise levels (x axis) for various values of γ. Higher means more efficient communication.

3.2 Intercept/Resend Attacks

In the previous section, we considered $\Sigma_{AB} = \{I_{AB}, \Pi^{(\alpha)}_{BB84}\}$ while E's strategy space was $\Sigma_E = \{I_E, \mathcal{A}\}$ where $\mathcal{A}$ was an optimal attack against the BB84 protocol utilizing a quantum memory system. We also assumed that the cost of performing attack $\mathcal{A}$ was similar to the cost of AB running the actual protocol (a very strong assumption in favor of the adversary). In practice, such an attack would be very difficult to launch against the protocol (and, with current technology, impossible as it would require a perfect quantum memory to perform successfully). In this section, we consider practical, so-called *Intercept-Resend (IR) attacks*. These attacks can be performed using today's technology; they also require hardware similar to that used by A and B, allowing us to more accurately compute the cost of an attack compared with the cost of running the actual protocol.

For this attack, on each iteration of the quantum communication stage, E will, with probability p, choose to attack and with probability $1 - p$ choose to ignore the incoming qubit. This value p will control how much noise E's IR attack actually creates (which, as before, must be kept below the natural noise level Q). This choice to attack or not is part of the strategy and is made independently for each iteration of the quantum communication stage. This is also different from the I_E strategy which chooses to not attack *every* iteration.

Should E decide to attack a particular iteration (with probability p), she will first measure the incoming qubit in a basis $\{|\nu_0\rangle, |\nu_1\rangle\}$ (this is fixed for each

iteration and part of the strategy) causing the qubit to collapse to one of the basis states $|\nu_0\rangle$ or $|\nu_1\rangle$. If E observes $|\nu_i\rangle$, she will "guess" that the key-bit for this iteration is $i \in \{0,1\}$. She will then send a fresh qubit in the state $|\nu_i\rangle$ to B.

There are two important parameters for an IR attack; first the value p and, second, the basis choice. We consider three common bases choices for IR attacks: $Z = \{|0\rangle, |1\rangle\}$, $X = \{|+\rangle, |-\rangle\}$ (see Sect. 2, and the Breidbart basis $B = \{|\phi_0\rangle, |\phi_1\rangle\}$, where: $|\phi_0\rangle = \cos\frac{\pi}{8}|0\rangle + \sin\frac{\pi}{8}|1\rangle$ and $|\phi_1\rangle = \sin\frac{\pi}{8}|0\rangle - \cos\frac{\pi}{8}|1\rangle$.

The value of p will be fixed to be the maximum value so that the induced noise is equal to Q. This makes sense, since the larger the value of p, the more information E may learn (since she is attacking more often), and since we cannot have p so large that the induced noise is higher than Q, the allowed maximum. Thus, once Q is given, the set Σ_E will consist of four distinct strategies: I_E (the "do nothing" attack); along with three strategies, one for each basis choice (we denote these attack strategies simply as Z, X, and B).

As for AB, we will consider three possible strategies: I_{AB} (i.e., "do nothing"); $\Pi^{(\alpha)}_{BB84}$ the BB84 protocol as analyzed previously (see Protocol 1); and $\Pi^{(\alpha)}_{B92}$, the B92 protocol [19] (see Protocol 2). Both BB84 and B92 are common protocols used in practical implementations of QKD [1]; B92 has the advantage that it requires less quantum resources to implement (and, so, is cheaper). However, at least in the standard adversarial model, B92 has a lower noise tolerance [20]. In this section, we will show that, so long as Q satisfies certain bounds, the joint strategy $(\Pi^{(\alpha)}_{BB84}, I_E)$ is a strict NE (for suitably chosen α); we will also show that $\Pi^{(\alpha)}_{BB84}$ is a dominate strategy for player AB and I_E is a DS for E for certain critical values of noise levels Q.

We begin by computing the utility of each possible action pair $(\Pi^{(\alpha)}, \mathcal{A})$. First, we must compute the cost associated to each strategy. To do so, we will define the following cost values for certain, basic, functionalities needed to implement the QKD protocol, and the IR attack:

$$
\begin{aligned}
C_S &: \text{The initial cost for } E \text{ to setup her attack equipment} \\
&\quad \text{(e.g., splicing into the quantum channel)} \\
C_M &: \text{The cost to perform a measurement in a single basis} \\
C_P &: \text{The cost to prepare a qubit basis state} \\
C_R(\delta) &: \text{The cost to produce a } \delta\text{-biased bit} \\
&\quad \text{We assume that } C_R(\delta) = h(\delta)C_R \text{ for some cost } C_R \\
C_{auth} &: \text{The cost for } AB \text{ to use the authenticated channel}
\end{aligned}
$$

We will assume that, if one requires an apparatus that is capable of producing a qubit in x different states, the cost is $\gamma_x C_P$ for some function γ_x. Similarly, for an apparatus capable of measuring a qubit in x different states, the cost is $\gamma_x C_M$. Our analysis below will be suitable for any non-decreasing γ_x; however when we evaluate our results, we will consider two cases: first $\gamma_x = 1$ for all x (i.e., there is no increase in cost) and, second, $\gamma_x = x$ (the cost increases

linearly in the number of required states). Note that we will assume $C_P \leq C_M$ which is a reasonable assumption since measurement devices are generally more complicated (and sensitive) than preparation devices [1]. These cost values may take into account such practical issues as device energy consumption over time for example (thus running the devices for longer, or having devices capable of performing additional measurements, will potentially cost users more).

From this, we can compute the following costs after N iterations of each protocol:

$$C_{AB}(\Pi^{(\alpha)}_{BB84}) = N[(3 + h(\alpha))C_R + \gamma_4 C_M + \gamma_4 C_P] + C_{auth} \tag{6}$$
$$C_{AB}(\Pi^{(\alpha)}_{B92}) = N[(2 + h(\alpha))C_R + \gamma_4 C_M + \gamma_2 C_P] + C_{auth}.$$

For BB84, AB must choose, each iteration, whether the iteration is a decoy or not (costing $h(\alpha)C_R$); what basis A should send in (with probability 1/2 each, thus costing C_R); what basis to measure in (costing C_R); and, finally, A must choose a random key bit (again, costing C_R). For B92, only one basis choice is required (from B). Finally, note that, BB84 is a *four-state* protocol in that A must prepare one of four possible qubit states each iteration. B92, however, is a *two-state* protocol - A must only be capable of preparing a state of the form $|0\rangle$ or $|+\rangle$. In both cases, however, B must be able to measure one of four states (from two bases). It is clear that the cost of running B92 is no greater than the cost of running BB84.

The cost for E to operate attack I_E is zero (i.e., $C_E(I_E) = 0$). The cost for the other strategies is the same: first, she must choose to attack or not, costing $h(p)C_R$; then she must measure and prepare a qubit in one basis. Those operations are performed for all N iterations of the quantum communication stage. Furthermore, she must also spend resources costing C_S to setup her attack initially (this is a one-time cost). The total cost for any attack $\mathcal{A} = Z, X, B$ is:

$$C_E(\mathcal{A}) = N[h(p)C_R + p(\gamma_2 C_M + \gamma_2 C_P)] + C_S, \text{ for any } \mathcal{A} \in \{Z, X, B\}. \tag{7}$$

To complete our utility computation, we must also compute the secret key length for each protocol under each attack. Since an IR attack results in three classical random variables (one for Alice, Bob, and Eve), we may use the Csiszar-Korner bound [22] to compute the number of secret bits that may be distilled from these sources. Let $\ell(N, \Pi^{(\alpha)}, \mathcal{A})$ be the amount of secret key bits that may be distilled after N iterations of protocol $\Pi^{(\alpha)}$ given that E used attack $\mathcal{A}$. Then from this bound, we have: $\ell(N, \Pi^{(\alpha)}, \mathcal{A}) = \eta N \alpha[I(A : B) - I(A : E)]$, where η is the proportion of non-discarded iterations; namely $\eta = 1/2$ for BB84 and $\eta = 1/4$ for B92 (see Protocols 1 and 2).

Note that the information computations above are dependent on only a single iteration of the protocol when faced with the specified attack since we are assuming iid attacks. Let $\mathcal{I}(\Pi^{(\alpha)}, \mathcal{A})$ be equal to $I(A : E)$ for the specified protocol and attack; then, the utility functions, for a fixed N, will be:

$$U_{AB}(\Pi^{(\alpha)}, \mathcal{A}) = \eta N \alpha[I(A : B) - \mathcal{I}(\Pi^{(\alpha)}, \mathcal{A})] - C_{AB}(\Pi^{(\alpha)}) \tag{8}$$
$$U_E(\Pi^{(\alpha)}, \mathcal{A}) = \eta N \alpha[\mathcal{I}(\Pi^{(\alpha)}, \mathcal{A}) + h(\widetilde{Q})] - C_E(\mathcal{A}), \tag{9}$$

where we use $\widetilde{Q}$ to denote the *raw-key error rate*; i.e., the error of the actual raw key which undergoes error correction (which, in the case of B92, is actually greater than the noise in the channel Q). The value $\eta N \alpha h(\widetilde{Q})$ denotes the information leaked to E "for free" during error correction.

To complete the utility computation, we require $I(A:B)$ and $I(A:E)$ for all possible protocols and strategy pairs. It is not difficult to show that $I(A:B) = 1 - h(\widetilde{Q})$. For BB84, a raw-key error occurs when a $|i\rangle$ flips to a $|1-i\rangle$ (for $i = 0, 1$) or when a $|\pm\rangle$ flips to a $|\mp\rangle$. By definition, this is exactly the channel noise level Q. Thus, for $\Pi^{(\alpha)}_{BB84}$, we have $I(A:B) = 1 - h(Q)$. For B92 it can be shown (see, for example, [23]) that the raw-key error is in fact: $\widetilde{Q} = 2Q/(1-2Q)$. Next, we must compute $\mathcal{I}(\Pi^{(\alpha)}, \mathcal{A})$. Clearly, $\mathcal{I}(\Pi^{(\alpha)}, I_E) = 0$ for any protocol. Consider, now, an IR attack where E measures and resends in a basis $\{|v_0\rangle, |v_1\rangle\}$ (in our case, either Z, X, or B, however the equations we derive here may be applied to other attack bases). By the measurement postulate, if A sends a qubit of the form $|i\rangle$ (for $i = 0, 1, +, -$), E will observe $|v_j\rangle$ with probability $v_{i,j} = |\langle i|v_j\rangle|^2$. To compute $\mathcal{I}(\Pi^{(\alpha)}, \mathcal{A})$ we will need the joint distribution held between A and E. This is straight-forward arithmetic: one must simply trace the execution of each protocol and use the measurement postulate. We summarize this distribution in Table 1.

Table 1. Showing the joint probability distribution for A's raw key bit and E's "guess" based on her attack (conditioning on the event she chooses to attack). For B92, we require a normalization term, denoted M which is: $M = v_{0,0}(v_{-,0} + v_{1,0}) + v_{0,1}(v_{-,1} + v_{1,1}) + v_{+,0}(v_{-,0} + v_{1,0}) + v_{+,1}(v_{-,1} + v_{1,1})$. The values here are found by tracing the protocol and using the measurement postulate.

AE	BB84	B92
00	$\frac{1}{4}(v_{0,0} + v_{+,0})$	$\frac{1}{M} v_{0,0}(v_{-,0} + v_{1,0})$
01	$\frac{1}{4}(v_{0,1} + v_{+,1})$	$\frac{1}{M} v_{0,1}(v_{-,1} + v_{1,1})$
10	$\frac{1}{4}(v_{1,0} + v_{-,0})$	$\frac{1}{M} v_{+,0}(v_{-,0} + v_{1,0})$
11	$\frac{1}{4}(v_{1,1} + v_{-,1})$	$\frac{1}{M} v_{+,1}(v_{-,1} + v_{1,1})$

By definition, we have $\mathcal{I}(\Pi^{(\alpha)}, \mathcal{A}) = p(H(A) + H(E) - H(AE))$ where the Shannon entropies may be computed easily from data in Table 1 and substituting in $|v_i\rangle$ for the appropriate basis state depending on the attack E uses (note that when E chooses to not attack, which occurs with probability $1-p$, she learns nothing, thus the need for the factor p in this expression). In summary, these are found to be:

$$\mathcal{I}(\Pi^{(\alpha)}_{BB84}, Z) \approx .189p \qquad \mathcal{I}(\Pi^{(\alpha)}_{BB84}, X) \approx .189p \qquad \mathcal{I}(\Pi^{(\alpha)}_{BB84}, B) \approx .399p$$
$$\mathcal{I}(\Pi^{(\alpha)}_{B92}, Z) \approx .459p \qquad \mathcal{I}(\Pi^{(\alpha)}_{B92}, X) \approx .459p \qquad \mathcal{I}(\Pi^{(\alpha)}_{B92}, B) = 0.$$

What remains is to find a value for p. As stated, we will assume that p is chosen to maximize E's information while keeping the induced noise from her

attack equal to Q. The natural noise in the channel is the average of the Z basis noise (which, in turn, is the average error of a $|i\rangle$ flipping to a $|1-i\rangle$ when it arrives at B's lab) and X basis noise (the average of a $|\pm\rangle$ flipping to a $|\mp\rangle$); that is: $Q = \frac{p}{4}(v_{0,0}v_{1,0} + v_{0,1}v_{1,1} + v_{1,0}v_{0,0} + v_{1,1}v_{0,1} + v_{+,0}v_{-,0} + v_{+,1}v_{-,1} + v_{-,0}v_{+,0} + v_{-,1}v_{+,1})$, from which it easily follows that $p = 2Q$ for $\mathcal{A} = Z, X$ and $p = 4Q$ for $\mathcal{A} = B$. Note that E may attack more often with the B basis as it induces less noise, on average, than the Z or X based IR attacks. From this analysis, we are now able to prove our two main results in this section involving sufficient conditions of the noise level for $(\Pi^{(\alpha)}_{BB84}, I_E)$ to be a strict NE and for each to be a DS.

Theorem 1. *Assume classical resources are free for both parties AB and E (that is, let $C_R = C_{auth} = C_S = 0$) and let $C_P \leq C_M$ (as discussed in the text). Define A_1 and A_2 as follows:*

$$A_1 = \frac{(\gamma_4 - \gamma_2)C_P}{\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q)} \qquad A_2 = \frac{2\gamma_4(C_M + C_P)}{1 - h(Q)}.$$

If $\max(A_1, A_2) < 1$ and Q, the noise in the channel is less than 0.232 and satisfies the following inequality:

$$\begin{cases} 10.025\left(\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q)\right) - \left(\frac{\gamma_4}{\gamma_2} - 1\right) > 0, & \text{If } A_1 \geq A_2 \\ 2.506(1 - h(Q)) - \frac{\gamma_4}{\gamma_2} > 0, & \text{Otherwise} \end{cases} \tag{10}$$

Then there exists an $\alpha \in [0, 1]$ such that $(\Pi^{(\alpha)}_{BB84}, I_E)$ is a strict NE.

Proof. Since $C_{auth} = C_S = 0$, the factor of N may be divided out of the utility functions (we are only interested in relations between them and the factor N appears in both U_{AB} and U_E. This allows us to construct the function table shown in Table 2. From this table, we see that, for $(\Pi^{(\alpha)}_{BB84}, I_E)$ to be a strict NE, the following inequalities must be satisfied:

$$\begin{aligned} \alpha &> \frac{2\gamma_4(C_M + C_P)}{1 - h(Q)} \\ \alpha &> \frac{(\gamma_4 - \gamma_2)C_P}{\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q)} \quad \left[\text{If } \frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q) > 0\right] \\ \alpha &< \frac{4\gamma_2(C_M + C_P)}{0.378} \approx 10.582\gamma_2(C_M + C_P) \\ \alpha &< \frac{8\gamma_2(C_M + C_P)}{1.596} \approx 5.013\gamma_2(C_M + C_P). \end{aligned}$$

Note that, if $Q < .232$ (as assumed in the hypothesis), then $\frac{1}{4} + \frac{1}{4}h(2Q/(1-2Q)) - \frac{1}{2}h(Q) > 0$. From this, it is clear that if we can find an α that satisfies:

$$\max(A_1, A_2) < \alpha < \frac{8\gamma_2(C_M + C_P)}{1.596},$$

Table 2. Function table for utility functions U_{AB} and U_E assuming $C_{auth} = C_S = 0$ and dividing out the factor of N on both functions.

Π_{AB}	$\mathcal{E} = I_E$
I_{AB}	$U_{AB} = 0$
	$U_E = 0$
$\Pi_{BB84}^{(\alpha)}$	$U_{AB} = \frac{\alpha}{2}(1 - h(Q)) - [(3 + h(\alpha))C_R + \gamma_4 C_M + \gamma_4 C_P]$
	$U_E = \frac{\alpha}{2} h(Q)$
$\Pi_{B92}^{(\alpha)}$	$U_{AB} = \frac{\alpha}{4}\left(1 - h\left(\frac{2Q}{1-2Q}\right)\right) - [(2 + h(\alpha))C_R + \gamma_4 C_M + \gamma_2 C_P]$
	$U_E = \frac{\alpha}{4} h\left(\frac{2Q}{1-2Q}\right)$
	$\mathcal{E} = Z = X$ (No difference between Z and X for these protocols)
I_{AB}	$U_{AB} = 0$
	$U_E = 0$
$\Pi_{BB84}^{(\alpha)}$	$U_{AB} = \frac{\alpha}{2}(1 - h(Q) - 0.378Q) - [(3 + h(\alpha))C_R + \gamma_4 C_M + \gamma_4 C_P]$
	$U_E = \frac{\alpha}{2}(h(Q) + 0.378Q) - [h(2Q)C_R + 2Q\gamma_2(C_M + C_P)]$
$\Pi_{B92}^{(\alpha)}$	$U_{AB} = \frac{\alpha}{4}\left(1 - h\left(\frac{2Q}{1-2Q}\right) - 0.918Q\right) - [(2 + h(\alpha))C_R + \gamma_4 C_M + \gamma_2 C_P]$
	$U_E = \frac{\alpha}{4}\left(h\left(\frac{2Q}{1-2Q}\right) + 0.918Q\right) - [h(2Q)C_R + 2Q\gamma_2(C_M + C_P)]$
	$\mathcal{E} = B$
I_{AB}	$U_{AB} = 0$
	$U_E = 0$
$\Pi_{BB84}^{(\alpha)}$	$U_{AB} = \frac{\alpha}{2}(1 - h(Q) - 1.596Q) - [(3 + h(\alpha))C_R + \gamma_4 C_M + \gamma_4 C_P]$
	$U_E = \frac{\alpha}{2}(h(Q) + 1.596Q) - [h(4Q)C_R + 4Q\gamma_2(C_M + C_P)]$
$\Pi_{B92}^{(\alpha)}$	$U_{AB} = \frac{\alpha}{4}\left(1 - h\left(\frac{2Q}{1-2Q}\right)\right) - [(2 + h(\alpha))C_R + \gamma_4 C_M + \gamma_2 C_P]$
	$U_E = \frac{\alpha}{4} h\left(\frac{2Q}{1-2Q}\right) - [h(4Q)C_R + 4Q\gamma_2(C_M + C_P)]$

the resulting joint strategy will be a strict NE (recall, by hypothesis, $\max(A_1, A_2) < 1$). For such a value to exist, it must be that $\max(A_1, A_2)$ is strictly less than the right-hand side of the above expression.

We show this in two cases. First, assume $A_2 > A_1$. Then, by our assumptions on the channel noise Q, we have:

$$\frac{\gamma_4}{\gamma_2} < 2.506(1 - h(Q))$$
$$\implies \frac{\gamma_4(C_M + C_P)}{1 - h(Q)} < \frac{4\gamma_2(C_M + C_P)}{1.596} \implies \frac{2\gamma_4(C_M + C_P)}{1 - h(Q)} < \frac{8\gamma_2(C_M + C_P)}{1.596},$$

as desired.

For the second case, assume $A_1 \geq A_2$. Then, by assumption on the channel noise Q, we have:

$$\frac{\gamma_4}{\gamma_2} - 1 < 10.025\left(\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q)\right)$$

$$\Longrightarrow \frac{2(\gamma_4 - \gamma_2)C_P}{\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q)} < \frac{8\gamma_2(2C_P)}{1.596}.$$

Noting that $C_P \leq C_M$ completes the proof.

Table 3. Showing the allowed noise tolerance for which $(\Pi^{(\alpha)}_{BB84}, I_E)$ is a strict NE. When $\gamma_4 = \gamma_2$ then it is always true that $A_2 \geq A_1$ (since $A_1 = 0$ and A_2 is always non-negative) and so we do not need to evaluate the case for $A_1 > A_2$. When $\gamma_4 = 2\gamma_2$, we must evaluate both cases. See text for explanation.

	$A_2 \geq A_1$	$A_1 > A_2$
$\gamma_4 = \gamma_2$	$Q \leq .146$	n/a
$\gamma_4 = 2\gamma_2$	$Q \leq .031$	$Q \leq .207$

Theorem 1 gives conditions on the noise parameter Q for which $(\Pi^{(\alpha)}_{BB84}, I_E)$ becomes a strict NE. The restrictions on $\max(A_i) < 1$ may be satisfied if the cost C_P and C_M are low enough. The restrictions on Q depend only on the value γ_4 and γ_2. So long as Q satisfies Eq. 10, then AB are motivated to run the BB84 protocol and E is motivated to not perform an intercept/resend attack (but, instead, to simply "listen" on the authenticated channel). We evaluate the noise tolerance in Table 3. Surprisingly, if $\gamma_2 = \gamma_4$, the noise tolerance is 14.6% also the maximal noise tolerance of BB84 in the standard adversarial model against optimal individual attacks (which are more general/powerful than IR attacks). Note, however, while the noise tolerance may be lower in our game theoretic model, *as before, the efficiency in our game theoretic model may improve as E is not motivated to attack.*

Theorem 2. *Assume classical resources are free for both parties (i.e., let $C_R = C_{auth} = C_S = 0$) and let $C_P \leq C_M$ (as discussed in the text). Define A_1 and A_2 as follows:*

$$A_1 = \frac{(\gamma_4 - \gamma_2)C_P}{\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q) - 0.798Q} \qquad A_2 = \frac{2\gamma_4(C_M + C_P)}{1 - h(Q) - 1.596Q}. \tag{11}$$

If $\max(A_1, A_2) < 1$ *and if Q, the noise in the channel, is strictly less than* 0.185 *and if it satisfies the following inequality:*

$$\begin{cases} 10.025\left(\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q) - 0.798Q\right) - \left(\frac{\gamma_4}{\gamma_2} - 1\right) > 0, & \text{If } A_1 \geq A_2 \\ 2.506(1 - h(Q) - 1.596Q) - \frac{\gamma_4}{\gamma_2} > 0, & \text{Otherwise} \end{cases} \tag{12}$$

then there exists a value for α such that $\Pi^{(\alpha)}_{BB84}$ is a dominate strategy (DS) for AB and I_E is a DS for E.

Proof. Fix α. For $\Pi^{(\alpha)}_{BB84}$ to be a DS for AB, we must show that, for every strategy $\mathcal{E} \in \Sigma_E$, it holds that $U_{AB}(\Pi^{(\alpha)}_{BB84}, \mathcal{E}) \geq U_{AB}(\Pi^{(\alpha)}, \mathcal{E})$ for $\Pi^{(\alpha)} = \Pi^{(\alpha)}_{B92}$ and $\Pi^{(\alpha)} = I_{AB}$. We see from Table 2, for this to be true, the following inequalities must be satisfied:

$$\alpha > \frac{2\gamma_4(C_M + C_P)}{1 - h(Q)} \qquad \alpha > \frac{(\gamma_4 - \gamma_2)C_P}{\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q)}$$

$$\alpha > \frac{2\gamma_4(C_M + C_P)}{1 - h(Q) - 0.378Q} \qquad \alpha > \frac{(\gamma_4 - \gamma_2)C_P}{\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q) + 0.0405Q}$$

$$\alpha > \frac{2\gamma_4(C_M + C_P)}{1 - h(Q) - 1.596Q} \qquad \alpha > \frac{(\gamma_4 - \gamma_2)C_P}{\frac{1}{4} + \frac{1}{4}h\left(\frac{2Q}{1-2Q}\right) - \frac{1}{2}h(Q) - 0.798Q}$$

Note that, the denominators of the above six inequalities are all positive by assumption that $Q < 0.185$. Note also, that there are only six inequalities, and not eight, since two are repetitions.

It is not difficult to see that if we take $\alpha \geq \max(A_1, A_2)$, where A_1 and A_2 are defined in Eq. 11, then all the above inequalities are automatically satisfied and, so, $\Pi^{(\alpha)}_{BB84}$ will be a DS for party AB.

Now, we consider E's strategy I_E. For I_E to be a DS for party E, the following inequalities must be satisfied (again, consulting Table 2):

$$\alpha < \frac{4Q\gamma_2(C_M + C_P)}{.378Q} \approx 10.582\gamma_2(C_M + C_P)$$

$$\alpha < \frac{8Q\gamma_2(C_M + C_P)}{1.596Q} \approx 5.013\gamma_2(C_M + C_P)$$

$$\alpha < \frac{8Q\gamma_2(C_M + C_P)}{0.918Q} \approx 8.715\gamma_2(C_M + C_P)$$

Clearly if $\alpha < \frac{8Q\gamma_2(C_M+C_P)}{1.596Q}$, the other two are also satisfied. All that remains to be shown is that an α exists allowing both $\Pi^{(\alpha)}_{BB84}$ to be a DS for AB and I_E to be a DS for E. In particular, we must show that: $\max(A_1, A_2) < \frac{8\gamma_2(C_M+C_P)}{1.596}$. However, this can be proven in a similar manner as in the proof of Theorem 1, using the new bounds on Q from Eq. 12. This completes the proof.

The allowed noise tolerances for $\Pi^{(\alpha)}_{BB84}$ to be a DS for *AB and I_E* to be a DS for E, are reported in Table 4.

4 Closing Remarks

In this paper, we introduced a new game-theoretic model of QKD security. Many interesting problems remain open. It would be interesting to analyze best-reply

Table 4. Showing the allowed noise values Q from Theorem 2.

	$A_2 \geq A_1$	$A_1 > A_2$
$\gamma_4 = \gamma_2$	$Q \leq .094$	n/a
$\gamma_4 = 2\gamma_2$	$Q \leq .024$	$Q \leq .13$

strategies under different noise values and decoy probabilities. We may also consider adding additional strategies for AB, different, non-linear, utility functions, and support for multi-user protocols [24]. One may also analyze the NE strategies based on Stackelberg game model, when the attacker E observes the strategy of party AB and chooses her strategy accordingly. One can envision a system whereby parties re-evaluate their choices after large sequences of N iterations, taking into account noise conditions, to chose new optimal strategies.

References

1. Scarani, V., Bechmann-Pasquinucci, H., Cerf, N.J., Dušek, M., Lütkenhaus, N., Peev, M.: The security of practical quantum key distribution. Rev. Mod. Phys. **81**, 1301–1350 (2009)
2. Katz, J.: Bridging game theory and cryptography: recent results and future directions. In: Canetti, R. (ed.) TCC 2008. LNCS, vol. 4948, pp. 251–272. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78524-8_15
3. Miao, F., Zhu, Q., Pajic, M., Pappas, G.J.: A hybrid stochastic game for secure control of cyber-physical systems. Automatica **93**, 55–63 (2018)
4. Zhu, Q., Basar, T.: Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: games-in-games principle for optimal cross-layer resilient control systems. IEEE Control Syst. **35**(1), 46–65 (2015)
5. Manshaei, M., Zhu, Q., Alpcan, T., Basar, T., Hubaux, J.: Game theory meets network security and privacy. ACM Comput. Surv. **45**(3), 25:1–25:39 (2013)
6. Zhu, M., Martinez, S.: Stackelberg-game analysis of correlated attacks in cyber-physical systems. In: American Control Conference, ACC, pp. 4063–4068, June 2011
7. Maitra, A., De, S.J., Paul, G., Pal, A.K.: Proposal for quantum rational secret sharing. Phys. Rev. A **92**(2), 022305 (2015)
8. Dou, Z., Xu, G., Chen, X.B., Liu, X., Yang, Y.X.: A secure rational quantum state sharing protocol. Sci. China Inf. Sci. **61**(2), 022501 (2018)
9. Zhou, L., Sun, X., Su, C., Liu, Z., Choo, K.K.R.: Game theoretic security of quantum bit commitment. Inf. Sci. (2018)
10. Maitra, A., Paul, G., Pal, A.K.: Millionaires problem with rational players: a unified approach in classical and quantum paradigms. arXiv preprint (2015)
11. Qin, H., Tang, W.K., Tso, R.: Establishing rational networking using the DL04 quantum secure direct communication protocol. Quantum Inf. Process. **17**(6), 152 (2018)
12. Das, B., Roy, U., et al.: Cooperative quantum key distribution for cooperative service-message passing in vehicular ad hoc networks. Int. J. Comput. Appl. **102**, 37–42 (2014). ISSN 0975 8887

13. Houshmand, M., Houshmand, M., Mashhadi, H.R.: Game theory based view to the quantum key distribution BB84 protocol. In: 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, IITSI, pp. 332–336. IEEE (2010)
14. Kaur, H., Kumar, A.: Game-theoretic perspective of Ping-Pong protocol. Phys. A: Stat. Mech. Appl. **490**, 1415–1422 (2018)
15. Boström, K., Felbinger, T.: Deterministic secure direct communication using entanglement. Phys. Rev. Lett. **89**(18), 187902 (2002)
16. Lucamarini, M., Mancini, S.: Secure deterministic communication without entanglement. Phys. Rev. Lett. **94**(14), 140501 (2005)
17. Nielsen, M., Chuang, I.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)
18. Bennett, C.H., Brassard, G.: Quantum cryptography: public key distribution and coin tossing. In: Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, New York, vol. 175 (1984)
19. Bennett, C.H.: Quantum cryptography using any two nonorthogonal states. Phys. Rev. Lett. **68**, 3121–3124 (1992)
20. Renner, R., Gisin, N., Kraus, B.: Information-theoretic security proof for quantum-key-distribution protocols. Phys. Rev. A **72**, 012332 (2005)
21. Shor, P.W., Preskill, J.: Simple proof of security of the BB84 quantum key distribution protocol. Phys. Rev. Lett. **85**, 441–444 (2000)
22. Csiszár, I., Korner, J.: Broadcast channels with confidential messages. IEEE Trans. Inf. Theory **24**(3), 339–348 (1978)
23. Krawec, W.O.: Quantum key distribution with mismatched measurements over arbitrary channels. Quantum Inf. Comput. **17**(3), 209–241 (2017)
24. Phoenix, S.J., Barnett, S.M., Townsend, P.D., Blow, K.: Multi-user quantum cryptography on optical networks. J. Mod. Opt. **42**(6), 1155–1163 (1995)