BELL NONLOCALITY IS NOT SUFFICIENT FOR THE SECURITY OF STANDARD DEVICE-INDEPENDENT QUANTUM KEY DISTRIBUTION PROTOCOLS

MÁTÉ FARKAS, MARIA BALANZÓ-JUANDÓ, KAROL ŁUKANOWSKI, JAN KOŁODYŃSKI, AND ANTONIO ACÍN

Abstract. Device-independent quantum key distribution is a secure quantum cryptographic paradigm that allows two honest users to establish a secret key, while putting minimal trust in their devices. Most of the existing protocols have the following structure: first, a bipartite nonlocal quantum state is distributed between the honest users, who perform local projective measurements to establish nonlocal correlations. Then, they announce the implemented measurements and extract a secure key by post-processing their measurement outcomes. We show that no protocol of this form allows for establishing a secret key when implemented on any correlation obtained by measuring local projective measurements on certain entangled nonlocal states, namely on a range of entangled two-qubit Werner states. To prove this result, we introduce a technique for upper-bounding the asymptotic key rate of device-independent quantum key distribution protocols, based on a simple eavesdropping attack. Our results imply that either different reconciliation techniques are needed for device-independent quantum key distribution in the large-noise regime, or Bell nonlocality is not sufficient for this task.

1. Background and summary

Device-independent quantum key distribution (DIQKD) is the most secure quantum cryptographic paradigm. The security of DIQKD protocols is based only on the assumption that quantum theory is correct, and on observed measurement outcome statistics (often called correlations). In particular, the honest users do not need to trust the inner workings of their devices, and therefore these protocols are not vulnerable to implementation flaws that may be exploited in eavesdropping attacks.

All the known secure DIQKD protocols are based on the observed correlations violating a certain Bell inequality, usually a biased CHSH inequality [? ? ? ? ? ? ?]. It is in fact relatively straightforward to see that Bell nonlocality (the violation of a Bell inequality) is necessary for the security of any DIQKD protocol. However, it remains an open question to characterise the necessary and sufficient conditions a correlation needs to satisfy in order to be useful for secure DIQKD.

One approach to rule out that certain correlations are useful for DIQKD is to derive *upper bounds* on the extractable key rate from a given correlation. This approach has gained some interest recently, and a few generic techniques have been proposed [? ? ?]. However, all

The second author has been partially supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodovska-Curie grant agreement No 847517.

The talk at the 8IMM 2022 has been given by the second author.

the upper bounds constructed thus far remain positive for nonlocal correlations, suggesting that Bell nonlocality might be sufficient for the security of DIQKD protocols.

In this work, we prove the opposite for a generic class of DIQKD protocols that we call *standard* protocols, as most of the known protocols belong to this class. In particular, we show that for a large class of nonlocal correlations, the upper bound on the key rate is zero, and hence, Bell nonlocality is in general not sufficient for the security of standard DIQKD protocols. Our proof is based on a simple eavesdropping attack, which allows for efficiently computing upper bounds on DIQKD key rates for arbitrary correlations.

2. Tools and main ideas

Every DIQKD protocol can be described by two trusted parties—Alice and Bob—measuring a fresh copy of a bipartite quantum state ρ_{AB} in each of a total of n measurement rounds. In each round, they select one of many possible measurement settings (labelled by x for Alice and y for Bob), and record their measurement outcomes (labelled by a for Alice and b for Bob). After sufficiently many rounds, they publicly reveal the measurement settings and outcomes for a subset of the rounds in order to estimate the joint distribution

$$p_{AB}(a,b|x,y) = \mathrm{tr}[\rho_{AB}(A_a^x \otimes B_b^y)], \tag{1}$$

specifying the probability of observing the outcome pair a, b given the setting pair x, y. In Eq. (1), A_a^x (B_b^y) denotes the measurement operator of Alice (Bob) corresponding to the setting x (y) and the outcome a (b). Note that while quantum theory prescribes that the correlation can be written in the form of Eq. (1), Alice and Bob do not need to know the form of the quantum state and the measurements. From their perspective, they are simply selecting the settings x and y, and recording the outcomes a and b.

Once the estimation is done, they attempt to extract a secure key from the non-disclosed outcomes a and b, using two-way public communication schemes, usually referred to as privacy amplification (reducing the knowledge of any potential eavesdropper about their outcomes) and error correction (increasing the correlation between their outcomes) [?]. At the end of the protocol, Alice and Bob aim to hold a pair of identical bit strings that appear completely random to any potential eavesdropper. In this work, we are interested in the *asymptotic key rate*, that is, the length of the final key string divided by the number of measurement rounds n, in the limit of infinitely many measurement rounds, $n \to \infty$.

We provide a generic tool for upper-bounding the asymptotic key rate for *standard* DIQKD protocols. In these protocols, after the estimation Alice and Bob publicly announce the measurement settings for each round. Note that most DIQKD protocols proposed thus far are of this kind [? ? ? ? ? ? ?], with only a few exceptions (see e.g. [?]).

In order to upper-bound the key rate, it is sufficient to describe a specific eavesdropping attack conceivable within the limitations of quantum theory. We propose the *convex combination* (CC) attack, which was originally considered for eavesdroppers limited only by the no-signalling principle [? ?]. It is a simple *individual* attack, that does not require any quantum memory from the eavesdropper [?]. Since in the device-independent paradigm Alice and Bob do not trust the precise form of the shared state and their measurements, the eavesdropper (Eve) can make use of this lack of information. In particular, one way of thinking of the CC attack is assuming that Eve knows the precise form of the measurements A_a^x and B_b^y, and she is the one distributing the quantum state ρ_{AB} in Eq. (1). This attack

can effectively be described as Eve "distributing" correlations via the quantum states. In the CC attack, Eve distributes a local correlation $p^{\mathcal{L}}_{AB}(a,b|x,y)$ with probability $q_{\mathcal{L}}$, and a nonlocal correlation $p^{\mathcal{NL}}_{AB}(a,b|x,y)$ with probability $1-q_{\mathcal{L}}$. The correlation observed by Alice and Bob is hence given by

$$p_{AB}(a,b|x,y) = q_{\mathcal{L}} \cdot p^{\mathcal{L}}_{AB}(a,b|x,y) + (1-q_{\mathcal{L}}) \cdot p^{\mathcal{NL}}_{AB}(a,b|x,y). \tag{2}$$

Every local correlation can be realised as a convex combination of deterministic correlations [**?**]. Hence, in every round in which the correlation distributed by Eve is local, we assume that Eve knows the outcomes of Alice and Bob after they have announced their measurement settings. For this reason, in the CC attack we maximise $q_{\mathcal{L}}$ by finding a suitable decomposition of the observed correlation $p_{AB}(a,b|x,y)$. Note that this maximisation is a linear program, since the set of local correlations forms a convex polytope [**?**]. Therefore, computing the CC attack [the right-hand side of Eq. (2)] for a given observed correlation [the left-hand side of Eq. (2)] can be performed efficiently.

According to the above, we assume that in the CC attack Eve records the outcomes of Alice and Bob in every local round, while she records "?" (representing her limited knowledge) in every nonlocal round. This eavesdropping strategy can be described by the tripartite correlation

$$p_{ABE}(a,b,e|x,y) = q_{\mathcal{L}} \cdot p^{\mathcal{L}}_{AB}(a,b|x,y) \cdot \delta_{e,(a,b)} + (1-q_{\mathcal{L}}) \cdot p^{\mathcal{NL}}_{AB}(a,b|x,y) \cdot \delta_{e,?}. \tag{3}$$

Once such a tripartite correlation is known, there exist well-established results from classical cryptography bounding the asymptotic key rate extractable via two-way public communication. Here, we use the *intrinsic information* [**?**], which is an upper bound on the key rate, given by

$$I(A:B\downarrow E) = \min_{T:E\to\bar{E}} I(A:B|T(E)), \tag{4}$$

where the minimisation is taken over all stochastic maps $T: E \to \bar{E}$, and $I(A:B|T(E))$ is the conditional mutual information of the resulting distribution $p_{AB\bar{E}}(a,b,\bar{e})$. The key rate r in the DIQKD protocol is then bounded by

$$r \leq \sum_{x,y} p_{xy} \cdot I_{xy}(A:B\downarrow E), \tag{5}$$

where $I_{xy}(A:B\downarrow E)$ is the intrinsic information of the tripartite distribution in Eq. (3), and p_{xy} is the probability of Alice and Bob choosing the setting pair x,y in the protocol. In summary, the CC attack provides an efficient means for computing upper bounds on the key rate of standard DIQKD protocols. Note that while the minimisation in Eq. (4) is non-linear, every stochastic map $T: E \to \bar{E}$ provides a valid upper bound on the key rate r.

3. Results

We apply the CC attack to DIQKD protocols that use the two-qubit Werner state

$$\rho^{v}_{AB} = v|\psi_-\rangle\langle\psi_-| + \frac{1-v}{4}\mathbb{I}, \tag{6}$$

where $|\psi_-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, $\mathbb{I}$ is the identity operator, and $v \in [0,1]$ is the *visibility*. It is known that for visibilities no greater than $v^{w}_{\mathcal{L}} \approx 0.6829$, the Werner state is local, i.e., any

arrangement of arbitrarily (even infinitely) many local projective measurements leads to a local correlation [**?**].

Any observed correlation from the Werner state with visibility v can be written as a unique convex combination of a local correlation from the Werner state with visibility $v^w_{\mathcal{L}}$, and a correlation from the state $|\psi_-\rangle\langle\psi_-|$. It is precisely this convex combination that we will assume that Eve uses in the CC attack for an arbitrary standard DIQKD protocol based on a Werner state and arbitrarily many projective measurements. Furthermore, we assume that Eve applies the map $T : E \to \bar{E}$ on her variable e that maps every pair $e = (a, b)$ such that $a \neq b$ to $\bar{e} =?$, and leaves every other variable invariant. The CC attack is therefore fully characterised, and we can compute the upper bound on the key rate in Eq. (5) for any setting pair x, y. We find that for any possible pair of projective measurements, the intrinsic information $I_{xy}(A : B \downarrow E)$ is zero for visibilities no greater than

$$v^w_{\text{crit}} = \frac{v^w_{\mathcal{L}} + 1}{3 - v^w_{\mathcal{L}}} \approx 0.7263 > v^w_{\mathcal{NL}} \approx 0.7012, \tag{7}$$

where $v^w_{\mathcal{NL}}$ is the nonlocal visibility, for which it is known that there exist arrangements of projective measurements that give rise to a nonlocal correlation [**?**]. In summary, Alice and Bob cannot use any Werner state with visibility $v^w_{\mathcal{NL}} \leq v \leq v^w_{\text{crit}}$ for standard DIQKD, even though the state is nonlocal, i.e., there exist measurements that give rise to nonlocal correlations. This means that in general Bell nonlocality is not sufficient for the security of standard DIQKD protocols.

The bound on the critical visibility can be improved if the number of measurement settings is fixed. This is the case in the most commonly used biased CHSH-based DIQKD protocols, that use the Werner state with two measurements on one side, and three or four measurements on the other side [**? ? ? ? ? ? ? ? ?**]. In this case, the polytope of local correlations is fully characterised [**?**], and the visibility at which the correlation obtained from the Werner state becomes local is known exactly. Denoting the bias parameter in the biased CHSH inequality by $\theta \in (0, \pi/2)$, the local visibility is given by $v^\theta_{\mathcal{L}} = 1/(\cos\theta + \sin\theta)$. This together with our results on general Werner state protocols gives us the critical visibility for the biased CHSH-based protocols, $v^\theta_{\text{crit}} = (v^\theta_{\mathcal{L}} + 1)/(3 - v^\theta_{\mathcal{L}}) > v^\theta_{\mathcal{L}}$. That is, in the range $v^\theta_{\mathcal{L}} < v \leq v^\theta_{\text{crit}}$, the observed correlation in the biased CHSH-based protocol is nonlocal, but the key rate is zero. Moreover, in these protocols, Alice and Bob usually extract the key from a fixed setting pair, and the upper bound on the key rate in Eq. (5) can be calculated for every θ as a function of the visibility v. In particular, we present the bound on the standard CHSH-based protocol ($\theta = \pi/4$) in Fig. 1, compared to two recently derived upper bounds. Note that the bound becomes zero at the critical visibility, and near this visibility it outperforms all the previously known upper bounds. In the same figure, we present a region in the set of quantum correlations (red region in the inset) that is nonlocal but cannot be used for standard DIQKD.

4. Discussion and impact

We presented a generic tool for upper-bounding the asymptotic key rate in standard DIQKD protocols, based on a simple eavesdropping attack. Our bound is efficiently computable, as the attack can be computed via a linear program. Using our tools, we showed that a large class of nonlocal correlations is useless for standard DIQKD, and therefore Bell

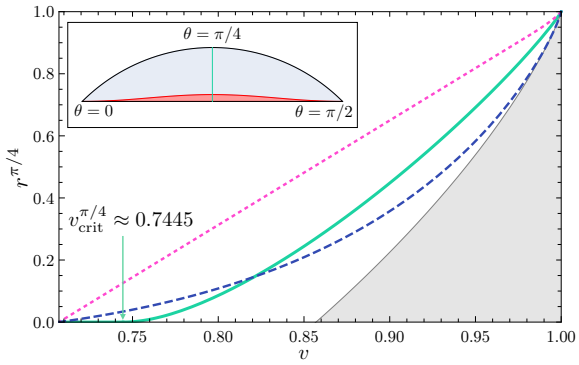

FIG. 1. Upper bounds on the asymptotic key rate for the standard CHSH protocol in terms of the visibility. The dotted line is the upper bound from [**?**], the dashed line is the upper bound from [**?**], and the solid line is our bound. The shaded area represents the lower bound from [**?**]. In the inset, we depict a region of nonlocal quantum correlations (in red) that cannot be used for key extraction if the honest parties announce their settings. Here, the vertical line corresponds to the correlations obtained from the standard CHSH protocol, the boundary points on the top represent the correlations for $0 < \theta < \frac{\pi}{2}$ and visibility $v = 1$, and the horizontal line on the bottom represents a facet of the local polytope.

nonlocality is not sufficient for the security of these protocols. In particular, the commonly used biased CHSH-based protocols become insecure in the noisy case, while still exhibiting nonlocality.

Our results do not immediately rule out the studied correlations as being useful in DIQKD. However, if one aims to show that they are useful, one needs to propose protocols beyond standard DIQKD. One possibility is to consider protocols similar to that of Ref. [**?**], in which only one party announces their settings. Indeed, for this protocol we were unable to derive a vanishing upper bound in the nonlocal regime using our techniques. Another possibility is to consider non-projective measurements, however, it is worth noting that there is no known quantum state that is local for all arrangements of projective measurements, while exhibiting nonlocality for some arrangement of non-projective measurements.

Maria Balanzó-Juandó; ICFO – Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Spain
Email address: maria.balanzo@icfo.eu