Stronger Attacks on Causality-Based Key Agreement

Benno Salwey and Stefan Wolf
Faculty of Informatics, Università della Svizzera Italiana, Via G. Buffi 13, 6900 Lugano, Switzerland

***Abstract*—Remarkably, it has been shown that in principle, security proofs for quantum key-distribution (QKD) protocols can be independent of assumptions on the devices used and even of the fact that the adversary is limited by quantum theory. All that is required instead is the absence of any hidden information flow between the laboratories, a condition that can be enforced either by shielding or by space-time causality. All known schemes for such Causal Key Distribution (CKD) that offer noise-tolerance (and, hence, must use privacy amplification as a crucial step) require multiple devices carrying out measurements *in parallel* on each end of the protocol, where the number of devices grows with the desired level of security. We investigate the power of the adversary for more practical schemes, where both parties each use a single device carrying out measurements *consecutively*. We provide a novel construction of attacks that is strictly more powerful than the best known attacks and has the potential to decide the question whether such practical CKD schemes are possible in the negative.**

I. Introduction

The use of quantum theory in cryptography allows for realising a task classically impossible unless assumptions are made on the computational power of the adversary: starting from a small shared secret key, two parties Alice and Bob can generate much longer secret keys. Such *quantum cryptography* goes back to the celebrated seminal work by Charles Bennett and Gilles Brassard in 1984 [6]. They devised a protocol based on the exchange of single quantum bits, *e.g.*, coded into the polarisation of single photons. The security of the protocol depends on the assumptions sketched in Figure 1.

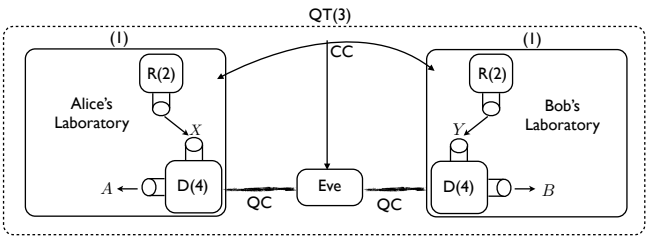

Fig. 1. *Schematic setup of QKD protocols with assumptions (1)-(4). The boxes around the legitimate parties' laboratories indicate* **protection against unwanted information leakage** *(1). The R's are the* **sources of free randomness**[2] *(2) used as the inputs (x, y) to the* **devices D which generate, and operate on, the specified quantum systems** *(4). CC refers to a classical insecure (but authenticated) channel to which the adversary Eve also has access. QC is a completely insecure quantum channel which Eve may interfere with to an unspecified extent. The dotted box indicates that* **the protocol takes place within the rules of quantum theory** *(3).*

It lies in the spirit of cryptography to reduce the assumptions under which security can be proven. In the physics community, quantum key distribution became prominent and popular through the work of Artur Ekert [13], who presented a protocol based on *entangled* pairs of quantum bits, and on the phenomenon of *non-local correlations* [5]: If the joint behaviour, under measurements, of two parts of a system is stronger than what can be explained by shared (classical) information, one speaks of *non-local* correlations since no *local* hidden-variable model alone can lead to the behaviour (alone). A joint two-partite input-output behaviour, also called *system* in the following, is recognised to be non-local if it violates some Bell inequality, the latter being respected by all local systems. The rationale of Ekert's method is as follows (see also Figure 2): If, after exchange and measurement on the two parts of the entangled pair, respectively, a (virtually) maximal violation of a specific Bell inequality, due to Clauser, Horne, Shimony, and Holt [11], occurs, then the shared state must be (close to) a maximally entangled pair of quantum bits. Furthermore, (the completeness of) quantum theory implies that the outcomes when such a *singlet* state is measured are (a) perfectly correlated with each other yet at the same time (b) completely *un*correlated with any (classical or quantum) information *outside* the two laboratories (and, hence, potentially under an adversary's control); the latter follows from a state violating maximally the CHSH inequality necessarily being *pure*. Ekert's result (and [18] when dealing with noise) has been a big step towards *device-independent* security [1] and the possibility of dropping assumption (4) (see Figure 1). Vazirani and Vidick [22] devised a scheme similar to Ekert's, where the two parties could each reuse a single device to achieve full device-independent security even tolerating (a certain level of) noise. They proved that the partial security of the raw key consisting of the (measurement) outputs of the devices can be amplified using standard *privacy-amplification* techniques [8], [7], [16]. However, even their security proof, like Ekert's, rests on the validity of the entire Hilbert-space formalism of quantum theory. It is natural to ask whether it is possible to derive security of the final key *directly and only* from the (extent of) non-locality of the generated values (see Figure 2), together with the assumption that no hidden communication has taken place between the laboratories. Barrett, Hardy, and Kent [4] have

[2]We refer to the notion of free randomness used by Colbeck and Renner in [12]: A random variable, generated at some point in space-time, displays free randomness if it is independent of any variable which lies outside its future light-cone.

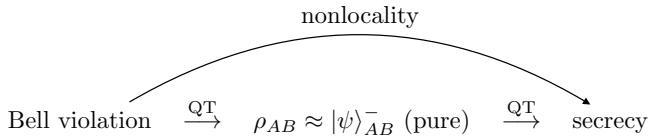

Fig. 2. *Ekert's reasoning: If a system violates the CHSH inequality virtually maximally (i.e., close to Tsirelson's bound [10]), then the framework of quantum theory implies that the state of the system must be close to a maximally entangled and, hence,* **pure** *state, a Bell state. The purity of the entangled state implies the secrecy of the local measurement outcomes. This reasoning is strongly based on the formalism of quantum theory.*
Barrett, Hardy, and Kent's reasoning: A Bell-inequality violation indicates a non-local correlation that **directly** *implies a constraint on the predictive power of any external piece of information (such as, e.g, Eve's entire knowledge) about Alice and Bob's measurement outcomes. This reasoning is independent of quantum theory.*

shown that in principle, the answer is *yes*: They presented a protocol generating a secret key under the sole assumption that *no illegitimate communication* takes place between the laboratories. Note that such "*causal key agreement*" requires neither Assumption (3) nor (4) above, see Figure 1.

Motivated by this proof of principle, several authors have worked on developing protocols that are based on the CHSH inequality instead of the chained Bell inequality [9], and that are not only more efficient but also tolerant to noise [14], [17]. However, besides the no-signalling assumption *between* the parties, the protocols' security proofs must be based on the same condition *within* their laboratories in order to perform privacy amplification.[3] Actually, in [15], the impossibility of privacy amplification was shown if there are no additional no-signalling conditions assumed. Yet, if Alice and Bob reuse their devices, then previously obtained *outputs cannot depend on future inputs* as a consequence of (2); the corresponding additional conditions are termed *time-ordered no-signalling* (TONS) conditions. In [3], it was shown that under the TONS conditions, super-linear privacy amplification is impossible: Using class of attacks which we refer to as "*prefix-code attacks*" (see Definition III.6), they showed that if n is the length of the input to the amplification function, then the adversary's knowledge on the output is at least of order $o(1/n)$. Furthermore, prefix-code attacks rule out the use of linear privacy-amplification functions (which are used for 2-universal hashing) as here the adversary's knowledge on the output remains constant (*i.e.*, independent of n). However, the knowledge prefix-code attacks yield about non-linear functions is limited, *e.g.*, $\Theta(1/\sqrt{n})$ for majority functions. We present a novel construction of TONS attacks which comprise prefix-code attacks and, furthermore, can also provide a constant knowledge on the output for highly non-linear functions, *i.e.*, an improvement of $\Theta(\sqrt{n})$ over prefix-code attacks in the case of majority. That our attack proves TONS privacy amplification with linear functions as well as a highly non-linear function like majority impossible is an indicator that the attack is sufficiently strong to rule out TONS privacy amplification at all. From a practical point of view impossibility of TONS privacy amplification means that Alice and Bob necessarily need additional devices which are shielded against information loss to carry out CKD.

Due to spatial limitations we are forced to omit the detailed proof of Theorem III.3, Lemma III.7, and Theorem III.8 and refer the reader to Chapters 3.4.1 and 3.5.4 in [21].

[3]The number of required no-signalling conditions is proportional to the negative logarithm of the tolerable noise level.

II. Preliminaries

A. No-signalling systems

We refer to a system A as a black box with an interface consisting of an input $x \in \mathcal{X}$ and an output $a \in \mathcal{A}$, where its complete input-output behaviour is specified by the conditional probability distribution $\mathsf{P}(a \,|\, x)$. If a system A is shared between m parties, each holding n marginal systems, then we denote the interface of the i-th marginal system held by party j by A_i^j. No-signalling conditions between different systems simply mean that the input one party inserts into her system does not affect the output the other party obtains from her system.

Definition II.1 (m-Party no-signalling)**.** An m-system box

$$\mathsf{P}(a^1 \dots a^m \,|\, x^1 \dots x^m)$$

is *m-party no-signalling* if no subset of parties, $I^1 \subseteq [m]$, can signal to any other (disjoint) subset of parties. Defining I^2 to be the complementary set to I^1 we have formally

$$\sum_{a^{I^1}} \mathsf{P}(a^{I^1} a^{I^2} \,|\, x^{I^1} x^{I^2}) = \sum_{a_{I^1}} \mathsf{P}(a^{I^1} a^{I^2} \,|\, (x')^{I^1} x^{I^2}) \quad \forall I^1, a^{I^2}, x^{I^1}, (x')^{I^1}, x^{I^2} \ . \qquad (1)$$

We introduce the short-hand notation $A^{I^1} \xrightarrow{ns} A^{I^2}$ if (1) is satisfied, *i.e.*, the systems A^{I^1} do not signal to the systems A^{I_2}.

Definition II.2 (Marginal)**.** $A^{I^1} \xrightarrow{ns} A^{I^2}$ induces a valid *marginal* distribution $\mathsf{P}(a^{I^2} \,|\, x^{I^2})$ on the systems A^{I^2} that is independent of the inputs chosen by the parties in I^1.

Definition II.3 (No-signalling extension)**.** A *no-signalling extension* of a given system A (possibly consisting of arbitrarily many subsystems), identified with $\mathsf{P}(a \,|\, x)$, is any joint system AE, identified with $\mathsf{P}'(ae \,|\, xu)$, such that $A \xleftrightarrow{ns} E$ and the marginals on A coïncide, *i.e.*, $\mathsf{P}'(a \,|\, x) = \mathsf{P}(a \,|\, x)$.

We consider the case of three parties that we identify with Alice, Bob, and Eve ($A^1 = A, A^2 = B, A^3 = E$), where Alice and Bob each hold n subsystems. We use the shorthand notation $A_{\leq n} := A_1 A_2 ... A_n$ to define the no-signalling conditions that are relevant if Alice and Bob each *reuse* their devices to create the systems $A_i B_i$ *consecutively*.

Definition II.4 (TONS)**.** A $(2n+1)$-system

$$\mathsf{P}(a_{\leq n} b_{\leq n} e \,|\, x_{\leq n} y_{\leq n} u)$$

is *time-ordered no-signalling (TONS)* if no subset of marginal systems can signal to systems outside its causal future. Any

union of systems $A_{\leq i} \cup B_{\leq j} \cup E_{\leq k}$, with $k \in \{0,1\}$ and $0 \leq i,j \leq n$, must have a valid marginal distribution $\mathsf{P}(a_{\leq i}b_{\leq j}e_{\leq k} \,|\, x_{\leq i}y_{\leq j}u_{\leq k})$ induced by the equations

$$\sum_{a_{>i}b_{>j}e_{>k}} \mathsf{P}(a_{\leq i}a_{>i}b_{\leq j}b_{>j}e_{\leq k}e_{>k} \,|\, x_{\leq i}x_{>i}y_{\leq j}y_{>j}u_{\leq k}u_{>k}) =$$

$$\sum_{a_{>i}b_{>j}e_{>k}} \mathsf{P}(a_{\leq i}a_{>i}b_{\leq j}b_{>j}e_{\leq k}e_{>k} \,|\, x_{\leq i}x'_{>i}y_{\leq j}y'_{>j}u_{\leq k}u'_{>k})$$

$$\forall (a_{\leq i}, b_{\leq j}, e_{\leq k}, x_{\leq i}, y_{\leq j}, u_{\leq k}), (x_{>i}, y_{>j}, u_{>k}), (x'_{>i}, y'_{>j}, u'_{>k}), 0 \leq i,j \leq n, k \in \{0,1\} \; . \tag{2}$$

B. Some explicit no-signalling distributions

- We denote by $\mathsf{U}(a \,|\, x)$ a box that outputs a uniformly random element of the output alphabet $\mathcal{A}$

$$\mathsf{U}(a \,|\, x) := \frac{1}{|\mathcal{A}|} \qquad \forall a, x \; . \tag{3}$$

- We denote by $\mathsf{PR}(ab \,|\, xy)$, with $\mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Y} = \{0,1\}$, as a box with probabilities

$$\mathsf{PR}(ab \,|\, xy) := \begin{cases} \frac{1}{2} & \text{if } a \oplus b = x \cdot y \\ 0 & \text{otherwise} \; . \end{cases} \tag{4}$$

- We denote by $\mathsf{V}(ab \,|\, xy)$, with $\mathcal{A} = \{0,1\}$ and unspecified alphabets $\mathcal{B}$, $\mathcal{X}$, and $\mathcal{Y}$, as an arbitrary box that satisfies the no-signalling conditions (1) and has a uniform marginal on A,

$$\sum_b \mathsf{V}(ab \,|\, xy) = \frac{1}{2} \qquad \forall a, x, y \; . \tag{5}$$

  An example for this type of boxes is the PR box or the boxes corresponding to the chained Bell inequalities [9] considered in [3] and also multi-partite boxes corresponding to the multipartite *Guess Your Neighbours Input*-game [2], since the system B is not specified and can be composed of an arbitrary number of subsystems.
- We denote by $\mathsf{P}_\epsilon(ab \,|\, xy)$ the noisy version of an arbitrary box $\mathsf{P}(ab \,|\, xy)$ as the box with probabilities[4]

$$\mathsf{P}_\epsilon(ab \,|\, xy) := (1 - 2\epsilon)\, \mathsf{P}(ab \,|\, xy) + 2\epsilon\, \mathsf{U}(ab \,|\, xy) \; . \tag{6}$$

C. No-signalling privacy amplification

The task of privacy amplification is as follows. Suppose an adversary holding some system E can guess a single bit a_i with probability $1/2 + 2\epsilon$, but a complete bit-string $a_1 \ldots a_n$ only with exponentially small probability, let us say with probability at most $(1/2 + 2\epsilon)^n$. Usually, in a privacy-amplification protocol, one applies a randomly chosen function f^r, where r denotes the random choice, to obtain a shorter bit-string $s = f^r(a_1 \ldots a_n)$, think of a single bit, that cannot be guessed except with probability (exponentially in n) close to $1/2$. If the adversary E is governed by classical or quantum theory, it is possible to generate a single bit s that is (exponentially in n) close to uniform if the function f^r is chosen uniformly amongst all linear functions [8], [7], [16], [20]. In no-signalling privacy amplification, Alice and Bob hold a box $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$, and Alice outputs a Boolean function $f(a_{\leq n})$. To analyse the privacy of such a bit $f(a_{\leq n})$ against a no-signalling adversary, one considers, in analogy to the quantum case, an adversary Eve that holds a "no-signalling purifying marginal system" E with input U.

Definition II.5 (TONS attack)**.** The box

$$\mathsf{P}'(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}u)$$

is a *time-ordered no-signalling (TONS) attack* on the box $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ if it is a no-signalling extension of $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ and satisfies the TONS conditions (2).

We study privacy amplification in the context of secret-key distribution. Hence, Alice must communicate her choice r of the privacy-amplification function $f^r(a_{\leq n})$ to Bob eventually, such that they can arrive at a shared secret key in the end of the protocol. Since we assume that Eve can wiretap the classical communication between Alice and Bob and learn the value r, she can wait to use her system E until that happens and choose her input as a function of r, $u(r)$, accordingly. Her actions are completely specified by the box $\mathsf{P}'(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}u(r))$ and the figure of merit is Eve's maximal guessing probability $\mathsf{P}'(f^r(a_{\leq n}) = e \,|\, x_{\leq n}u(r))$ on the output of the privacy-amplification protocol. Since the marginal distribution $\mathsf{P}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ must be, in particular, independent of $u(r)$, each choice of r can be investigated independently and we can confine our analysis on attacks $\mathsf{P}'(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on *fixed* functions $f(a_{\leq n})$, where E has no input. Security against a TONS adversary stems from systems being non-local, *i.e.*, from systems violating a Bell inequality. If a no-signalling adversary Eve attacks, *e.g.*, a single $\mathsf{PR}_\epsilon(ab \,|\, xy)$ box, the probability $\mathsf{P}'(a = e \,|\, x)$ to guess the output a of Alice is at best $1/2 + 2\epsilon$ [14], *i.e.*, which is nontrivial exactly if the box is nonlocal. For simplicity of the representation, we assume that Alice and Bob hold n V_ϵ boxes, *i.e.*, the Bell inequality used has binary outcomes on Alice side and we confine ourselves to the hardest case, where Alice outcome is completely random in the noiseless case. The best known previous result on TONS privacy amplification is as follows.

Lemma II.6. *[3] Assume that Eve attacks $\mathsf{V}_\epsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ held by Alice and Bob. Then, for any function $f(a_{\leq n})$, there exists a TONS-attack $\mathsf{P}'(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$*

$$\mathsf{P}'(f(a_{\leq n}) = e \,|\, x_{\leq n}) \geq \frac{1}{2} + \frac{\epsilon}{2n} \qquad \forall x_{\leq n} \; . \tag{7}$$

III. The novel attack

A. Novel construction of TONS attacks

We present a novel construction of no-signalling attacks on $\mathsf{V}_\epsilon^{\otimes n}$. The idea is to decompose each of the n V_ϵ boxes in a pure and a noise part via (6) and then attack each of

[4] We chose this decomposition to be conform with the usual definition of the *"noisy PR-box"* PR_ϵ when P corresponds to the PR box introduced originally by Popescu and Rohrlich in [19].

the 2^n terms separately. We identify restrictions (8) and (9) on marginal (classical) distributions $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ on systems $A_{\leq n}E$ that permit extension to a TONS attack for each of the 2^n terms in the decomposition of $\mathsf{V}_\epsilon^{\otimes n}$.

Definition III.1 (Ordered $\mathcal{S}$-influenceable distributions)**.** For a set $\mathcal{S} \in \mathcal{P}([n])$ we define an *ordered $\mathcal{S}$-influenceable distribution* $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ as a probability distribution that satisfies uniformity on $a_{\leq n}$

$$\sum_e \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e) = 2^{-n} \quad \forall a_{\leq n} \quad \text{and} \tag{8}$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) = \frac{1}{2} \quad \forall a_{\leq i}, e, \quad \text{and} \quad i \in \overline{\mathcal{S}}\ . \tag{9}$$

We call the distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ ordered $\mathcal{S}$-influenceable since condition (9) implies that Eve can only bias the bits a_i with $i \in \mathcal{S}$, and, furthermore, for $j \notin \mathcal{S}$ the bits a_i can only be biased with respect to bits a_j if $j < i$.

Definition III.2 (Ordered $(\epsilon, \mathcal{S})$-divisible distribution)**.** Fix a full set of ordered $\mathcal{S}$-influenceable distributions $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$. We define an *ordered $(\epsilon, \mathcal{S})$-divisible distribution* $\mathsf{Q}_{o-\epsilon}(a_{\leq n}e)$, as

$$\mathsf{Q}_{o-\epsilon}(a_{\leq n}e) := \sum_{\mathcal{S}\in\mathcal{P}([n])} \omega(\mathcal{S}, n, \epsilon)\, \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\ , \tag{10}$$

with weights

$$\omega(\mathcal{S}, n, \epsilon) := (1-2\epsilon)^{n-|\mathcal{S}|}\,(2\epsilon)^{|\mathcal{S}|}\ . \tag{11}$$

Theorem III.3. *Any ordered $\mathcal{S}$-influenceable distribution $\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)$ can be extended to a TONS-attack $\mathsf{P}_{o-\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on the systems $A_{\leq n}B_{\leq n}$ with marginal distribution*

$$\mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}) := \prod_{i\in\mathcal{S}} \mathsf{U}(a_ib_i \,|\, x_iy_i) \prod_{i\in\overline{\mathcal{S}}} \mathsf{V}(a_ib_i \,|\, x_iy_i) \tag{12}$$

The proof of Theorem III.3 consists of an explicit construction of $\mathsf{P}'_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$:

$$\mathsf{P}'_{\mathcal{S}}(e) = \mathsf{Q}_{o-\mathcal{S}}(e) \tag{13}$$

$$\mathsf{P}'_{\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}e) = \prod_{i=1}^{n} \mathsf{P}'_{\mathcal{S}}(a_ib_i \,|\, a_{<i}b_{<i}x_{\leq n}y_{\leq n}e) \tag{14}$$

$$\mathsf{P}'_{\mathcal{S}}(a_ib_i \,|\, a_{<i}b_{<i}x_{\leq n}y_{\leq n}e) = \begin{cases} \mathsf{V}(a_ib_i \,|\, x_iy_i) & i \in \overline{\mathcal{S}} \\ \mathsf{U}(b_i \,|\, y_i)\, \mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) & \\ \qquad\qquad \text{otherwise}\ . & \end{cases} \tag{15}$$

It is a bit tedious but straightforward to show that (13)-(15) implies that $\mathsf{P}'_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$

1) satisfies the TONS-conditions (2),
2) has the correct marginal on systems $A_{\leq n}B_{\leq n}$:

$$\sum_e \mathsf{P}'_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \mathsf{P}_{\mathcal{S}}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n}), \tag{16}$$

3) and has the correct marginal on systems $A_{\leq n}E$:

$$\sum_{b_{\leq n}} \mathsf{P}'_{\mathcal{S}}(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\ . \tag{17}$$

Corollary III.4. *For any ordered $(\epsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\epsilon}(a_{\leq n}e)$, there exists a TONS-attack $\mathsf{P}'(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n})$ on $\mathsf{V}_\epsilon^{\otimes n}(a_{\leq n}b_{\leq n} \,|\, x_{\leq n}y_{\leq n})$ such that*

$$\sum_{b_{\leq n}} \mathsf{P}'(a_{\leq n}b_{\leq n}e \,|\, x_{\leq n}y_{\leq n}) = \mathsf{Q}_{o-\epsilon}(a_{\leq n}e) \qquad \forall x_{\leq n}, y_{\leq n} \tag{18}$$

Accordingly, we also denote $\mathsf{Q}_{o-\epsilon}(a_{\leq n}e)$ as a TONS attack.

B. Prefix-code attacks and their limits

Definition III.5 (Influence)**.** We define the *influence* $\Delta^f(a_{<i})$ *of a_i given the prefix $a_{<i}$ on the function $f(a_{\leq n})$* as

$$\Delta^f(a_{<i}) := \frac{1}{2}\Big(\mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}, a_i = 0) - \mathsf{Q}(f(a_{\leq n}) = 0 \,|\, a_{<i}, a_i = 1)\Big)\ , \tag{19}$$

where $\mathsf{Q}(a_{\leq n}) = 2^{-n}$.

Definition III.6 (Prefix-code attack)**.** *Given a prefix-code $C = \{c_1, c_2, ..., c_k\}$ and the function $f(a_{\leq n})$, we define the corresponding prefix-code attack as the ordered $(\epsilon, \mathcal{S})$-divisible distribution $\mathsf{Q}_{o-\epsilon}(a_{\leq n}e)$ induced by the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ defined as*

$$\mathsf{Q}_{o-\mathcal{S}}(e) = \frac{1}{2}\ , \tag{20}$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_i \,|\, a_{<i}e) = \begin{cases} \frac{1}{2}\left(1 + \mathit{sign}(\Delta^f(c_m))(-1)^{e\oplus a_i}\right) & \\ \quad \textit{if } \exists m:\ a_{<i} = c_m \cap i \in \mathcal{S}\ , & \\ \frac{1}{2} \qquad \textit{otherwise}\ . & \end{cases} \tag{21}$$

Lemma III.7. *Let the distribution $\mathsf{Q}_{o-\epsilon}(a_{\leq n}e)$ be a prefix-code attack on the majority function $\mathrm{Maj}_n(a_{\leq n})$. Then, for any choice of a prefix-code $C = \{c_1, ..., c_k\}$ the performance of this attack is*

$$\begin{aligned} \mathsf{Q}_{o-\epsilon}(\mathrm{Maj}_n(a_{\leq n}) = e) &= \frac{1}{2} + \epsilon \cdot 2^{-n+1}\binom{n-1}{\frac{n-1}{2}} \\ n \longrightarrow \infty \qquad &= \frac{1}{2} + \Theta\left(\frac{\epsilon}{\sqrt{n}}\right)\ . \end{aligned} \tag{22}$$

The insight behind the proof of Lemma III.7 is that in a prefix-code attack $\mathsf{Q}_{o-\epsilon}(a_{\leq n}e)$ on $\mathrm{Maj}_n(a_{\leq n})$, a *single* bit a_i is ϵ-biased towards the value e, while all other bits $a_{\neq i}$ are uniform when conditioned on e; the influence of a single bit a_i on the value of $\mathrm{Maj}_n(a_{\leq n})$ is of the order $\Theta\left(\frac{\epsilon}{\sqrt{n}}\right)$.

C. A stronger attack on Majority

We construct another attack $\mathsf{Q}_{o-\epsilon}(a_{\leq n}e)$ via the set $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$

$$\mathsf{Q}_{o-\mathcal{S}}(e) = \frac{1}{2} \tag{23}$$

$$\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}\,|\,e) = 2^{-n+1}\cdot\delta(\mathrm{Maj}_{\mathcal{S}}(a_{\mathcal{S}}),e)\ , \tag{24}$$

for $|\mathcal{S}|$ being odd (for even $|\mathcal{S}|$ we define $\mathrm{Maj}_{\mathcal{S}}(a_{\mathcal{S}})$ as the majority of all but the last bit). Intuitively, Eve makes a maximum-likelihood estimate of $\mathrm{Maj}_n(a_{\leq n})$ on the string $a_{\mathcal{S}}$, which is to compute $\mathrm{Maj}_{\mathcal{S}}(a_{\mathcal{S}})$. Due to the symmetry of the majority function with respect to exchange of indices, the guessing probability $\mathsf{Q}_{o-\mathcal{S}}(\mathrm{Maj}_n(a_{\leq n})=e)$ of the adversary depends only on $s := |\mathcal{S}|$.

Theorem III.8. *Let $|\mathcal{S}| = s = c\,n$ for some constant $0 < c < 1$ such that s is odd. Then there exists a series of ordered $\mathcal{S}$-influenceable distributions $\{\mathsf{Q}_{o-\mathcal{S}}(a_{\leq n}e)\}$ such that*

$$\mathsf{Q}_{o-\mathcal{S}}(\mathit{Maj}_n(a_{\leq n})=e) \overset{n\to\infty}{=} 1-\frac{\arctan\left(\sqrt{\frac{1-c}{c}}\right)}{\pi} \tag{25}$$

Through the concentration of measure around $s = 2\epsilon\,n$, induced by the central limit theorem, a direct consequence of Theorem III.8 is Corollary III.9.

Corollary III.9. *For any $\delta > 0$, there exists a series of $\mathsf{Q}_{o-\epsilon}(a_{\leq n}e)$ such that*

$$\mathsf{Q}_{o-\epsilon}(\mathit{Maj}_n(a_{\leq n})=e) \overset{n\to\infty}{\geq} 1-\frac{\arctan\left(\sqrt{\frac{1-(2\epsilon-\delta)}{(2\epsilon-\delta)}}\right)}{\pi}\ . \tag{26}$$

Lemma III.7 and Corollary III.9 imply an $\Theta(\sqrt{n})$ advantage of our attack on the best previously known attack, the prefix-code attack.

IV. Conclusion

Causal key distribution (CKD) requires only a *minimal* set of assumptions, *i.e.*, (1) a shielded laboratory and (2) free randomness, see Figure 1, which both can be considered also *necessary*: If the parties' laboratories leak information about the key the adversary eventually learns it. Without free randomness everything becomes deterministic from the view of the adversary, and she can compute the key herself. All CKD protocols that offer noise tolerance [14], [17] have the impractical requirement for Alice and Bob to use many devices in parallel, where each device needs to be shielded against unwanted information leakage individually. We address the (still) open problem whether CKD is also possible if Alice and Bob each *reuse a single device* and construct a novel attack on the necessary time-ordered no-signalling (TONS) privacy-amplification step in the CKD protocol. Our construction is a generalisation of the best known attack [3], and we prove it to be superior if majority functions are used for TONS privacy amplification; the amount of knowledge that our attack provides is optimal (up to a constant factor). That our attack performs well against TONS privacy amplification with linear functions as well as with a highly non-linear function like majority may suggest that it also powerful enough to prove impossibility of TONS privacy amplification in general, *if* this is indeed the case.

Acknowledgments

The authors thank Rotem Arnon-Friedman, Ämin Baumeler, Gilles Brassard, Omar Fawzi, Arne Hansen, Karol Horodecki, Jibran Rashid, Renato Renner, and Dave Touchette for stimulating discussions and helpful comments. BS and SW are supported by the Swiss National Science Foundation (SNF), the NCCR *QSIT*, by the COST action on "Fundamental Problems in Quantum Theory," and the CHIST-ERA project *DIQIP*.

References

[1] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98:230501, Jun 2007.

[2] Mafalda L. Almeida, Jean-Daniel Bancal, Nicolas Brunner, Antonio Acín, Nicolas Gisin, and Stefano Pironio. Guess your neighbor's input: A multipartite nonlocal game with no quantum advantage. *Phys. Rev. Lett.*, 104:230404, Jun 2010.

[3] Rotem Arnon-Friedman and Amnon Ta-Shma. Limits of privacy amplification against nonsignaling memory attacks. *Phys. Rev. A*, 86:062333, Dec 2012.

[4] Jonathan Barrett, Lucien Hardy, and Adrian Kent. No signaling and quantum key distribution. *Phys. Rev. Lett.*, 95:010503, Jun 2005.

[5] John S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[6] Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the International Conference on Computers, Systems and Signal Processing*, pages 175–179, 1984.

[7] Charles H. Bennett, Gilles Brassard, Claude Crepeau, and Ueli M. Maurer. Generalized privacy amplification. *IEEE Trans. Inf. Theor.*, 41(6):1915–1923, Nov 1995.

[8] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy amplification by public discussion. *SIAM J. Comput.*, 17(2):210–229, Apr 1988.

[9] Samuel L. Braunstein and Carlton M. Caves. Wringing out better Bell inequalities. *Nuclear Physics B - Proceedings Supplements*, 6(0):211 – 221, 1989.

[10] Boris S. Cirel'son. Quantum generalizations of Bell's inequality. *Letter in Mathematical Physics*, 4:93–100, 1980.

[11] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[12] Roger Colbeck and Renato Renner. No extension of quantum theory can have improved predictive power. *Nat. Commun.*, 2:411, Aug 2011.

[13] Artur K. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661–663, Aug 1991.

[14] Esther Hänggi, Renato Renner, and Stefan Wolf. Efficient device-independent quantum key distribution. In *Proceedings of the 29th Annual International Conference on Theory and Applications of Cryptographic Techniques*, EUROCRYPT'10, pages 216–234, 2010.

[15] Esther Hänggi, Renato Renner, and Stefan Wolf. The impossibility of non-signaling privacy amplification. *Theoretical Computer Science*, 486(0):27–42, 2013.

[16] Johan Hastad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, Mar 1999.

[17] Lluis Masanes. Universally composable privacy amplification from causality constraints. *Phys. Rev. Lett.*, 102:140501, Apr 2009.

[18] Dominic Mayers and Andrew Yao. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS '98, page 503, 1998.

[19] Sandu Popescu and Daniel Rohrlich. Nonlocality as an axiom. *Foundations of Physics*, 24(379), (1994).

[20] Renato Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6(01):1–127, 2008.
[21] Benno Salwey. *No-Signalling Attacks and Implications for (Quantum) Nonlocality Distillation*. PhD thesis, USI Lugano, 2015.
[22] Umesh Vazirani and Thomas Vidick. Fully device-independent quantum key distribution. *Phys. Rev. Lett.*, 113:140501, Sep 2014.