Quantum Cryptography and Security 2021/22
Report 3 - Quantum Key Distribution (QKD)
Prof. Giuseppe Vallone
Prof. Nicola Laurenti

Samuele Piccinelli

(Dated: 28 January, 2022)

Quantum key distribution (QKD) provides the only physically secure and proven method for the transmission of a secret key between two distant parties, Alice and Bob, who are connected by an authenticated classical channel and an (insecure) quantum channel. In particular, due to its ability to tolerate high channel losses, decoy-state QKD has been object of a lot of recent studies and it has been demonstrated to be secure and feasible in real-world conditions.
In this work we review a case of 3-state 1-decoy efficient BB84 protocol, extensively analysed in [1]. We will compute the relevant parameters for classical post-processing and graph both the quantum bit error rate (QBER) and the secret key rate (SKR) as a function of time.

1 Theoretical background

We follow slavishly the analysis that has been carried out in [1] and we refer to the article for the full derivation of the security parameters.
In this section we will quickly review all the relevant quantities needed to estimate the QBER and SKR, as we will see later on.

The finite-key analysis of the 2-decoy method bounds the secret key length of the protocol to the quantity

$$\ell \leq s^l_{Z,0} + s^l_{Z,1}(1 - h(\phi^u_Z)) - \lambda_{\text{EC}} - \log_2(b/\varepsilon_{\text{sec}}) - \log_2(2/\varepsilon_{\text{cor}}), \tag{1}$$

where

- $s^l_{Z,0}$ is the lower bound on the vacuum events $s_{Z,0}$, defined by the events where Bob had a detection and the pulse sent by Alice contained no photons;
- $s^l_{Z,1}$ is the lower bound on the single-photon events $s_{Z,1}$, defined by the number of detections at Bob side when the pulse sent by Alice contained only one photon;
- $h(\cdot)$ is the binary entropy;
- ϕ^u_Z is the upper bound on the phase error rate ϕ_Z;
- λ_{EC} is the number of disclosed bits in the error correction stage;
- ε_{sec}, ε_{cor}, are the secrecy and correctness parameters, set to the values 10^{-9} e 10^{-15} respectively, similarly to what is commonly used in literature;
- a and b are parameters whose value depend on the specific security analysis taken into account: $a = 6$, $b = 19$ for the 1-decoy protocol.

1.1 Counts and errors

In the 1-decoy protocol, we consider only a set of two intensity levels $\kappa = \{\mu_1, \mu_2\}$, where $\mu_1 > \mu_2$. We will consider, without loss of generality, the case where the states are encoded in the Z basis. Let $s_{Z,n}$ be the detection observed by Bob given that Alice sent an n photon state. The total number of detections in the Z basis are given by

$$n_Z = \sum_{n=0}^{\infty} s_{Z,n}. \tag{2}$$

In the asymptotic limit the number of detection with a specific intensity k should be $n^{\star}_{Z,k}$, where

$$n^{\star}_{Z,k} = \sum_{n=0}^{\infty} p_{k|n} s_{Z,n} \qquad \forall k \in \kappa, \tag{3}$$

with $p_{k|n}$ being the conditional probability of sending an n photon state with a given intensity k.
By considering now a finite statistics scenario, we can use Hoeffding's inequality for independent variables to bound the difference between our observed data $n_{Z,k}$ and the corresponding asymptotic case $n^{\star}_{Z,k}$ in the following way:

$$\left| n^{\star}_{Z,k} - n_{Z,k} \right| \leq \delta(n_Z, \varepsilon_1); \tag{4}$$

this latter relation holds with a probability $1 - 2\varepsilon_1$ and δ is a function given by the following expression

$$\delta(a, b) \coloneqq \sqrt{a \log(1/b)/2}. \tag{5}$$

Analogous consideration can be made for the error rate estimation: if we define the value $v_{Z,n}$ as the number of errors detected at Bob's side when Alice generated an n photon state and $m_Z = \sum_{n=0}^{\infty} v_{Z,n}$ as the total number of errors in the Z basis, the number of errors, $m^{\star}_{Z,k}$ for a pulse of intensity k can be expressed in the asymptotic case as:

$$m^{\star}_{Z,k} = \sum_{n=0}^{\infty} p_{k|n} v_{Z,n} \qquad \forall k \in \kappa. \tag{6}$$

The correction due to the finite size statistics, similarly to Eq. (4), is given by

$$\left| m^{\star}_{Z,k} - m_{Z,k} \right| \leq \delta(m_Z, \varepsilon_2), \tag{7}$$

where the expression holds with probability $1 - 2\varepsilon_2$.
With the total number of detection and the total number of errors we can define the quantum-bit-error-rate (QBER) as the ratio

$$\mathrm{QBER}_Z \equiv \mathcal{Q}_Z = m_Z / n_Z. \tag{8}$$

In a QKD protocol, the QBER refers to the fraction of positions where Alice's and Bob's bit strings differ and is generally a direct measure for the secrecy of Alice and Bob's strings since any eavesdropping strategy would perturb the correlations between them.
Lastly, we denote with

$$\tau_n = \sum_{k \in \kappa} p_k e^{-k} k^n / n! \tag{9}$$

the total probability to send an n photon state, with p_k the probability for Alice to choose one or the other intensity.

1.2 Bounds on vacuum and single-photon events

The upper bound the vacuum contribution $s_{Z,0}$ is given by

$$s_{Z,0} \leq s'^{u}_{Z,0} \coloneqq (2m_Z + \delta(n_Z, \varepsilon_1)) \tag{10}$$

and can be analogously obtained by considering only the errors relative to one intensity, taking into account the finite-key statistic: this approach yields to

$$s_{Z,0} \leq s''^{u}_{Z,0} \coloneqq 2\left(\tau_0 \frac{e^k}{p_k}\left(m_{Z,k} + \delta(m_Z, \varepsilon_1)\right) + \delta(n_Z, \varepsilon_1)\right). \tag{11}$$

In the analysis, both expressions have been implemented: the final $s^u_{Z,0}$ is then expressed by the tighter resulting bound between Eq. (10) and the values resulting for each intensity of Eq. (11): $s^u_{Z,0} \coloneqq \min(s'^{u}_{Z,0}, s''^{u}_{Z,0}(k = \mu_1), s''^{u}_{Z,0}(k = \mu_2))$.

On the other hand, the analytical bound on the single-photon events is given by

$$s_{Z,1} \geq s^l_{Z,1} \coloneqq \frac{\tau_1 \mu_1}{\mu_2(\mu_1 - \mu_2)}\left(n^-_{Z,\mu_2} - \frac{\mu_2^2}{\mu_1^2} n^+_{Z,\mu_1} - \frac{(\mu_1^2 - \mu_2^2)}{\mu_1^2} \frac{s^u_{Z,0}}{\tau_0}\right), \tag{12}$$

where we defined $n^{\pm}_{Z,k} = (n_{Z,k} \pm \delta(n_Z, \varepsilon_1))$, $\forall k \in \kappa$; an analogous relation holds also for the number of errors for a pulse of intensity k, $m_{Z,k}$, substituting $\varepsilon_1 \leftrightarrow \varepsilon_2$.

The lower bound on the vacuum events in the finite-key scenario, is given by the formula

$$s_{Z,0} \geq s^l_{Z,0} \coloneqq \frac{\tau_0}{\mu_1 - \mu_2}\left(\mu_1 n^-_{Z,\mu_2} - \mu_2 n^+_{Z,\mu_1}\right). \tag{13}$$

1.3 Phase error estimation and error correction bits

In order to upper bound the phase error in the Z basis, we can upper bound the number of bit errors in the X basis due to single-photon events through

$$v_{X,1} \leq v^u_{X,1} = \frac{\tau_1}{\mu_1 - \mu_2}\left(m^+_{X,\mu_1} - m^-_{X,\mu_2}\right), \tag{14}$$

obtaining for the phase error rate in the Z basis the formula

$$\phi_Z \leq \phi^u_x \coloneqq \frac{v^u_{X,1}}{s^l_{X,1}} + \gamma\left(\varepsilon_{\text{sec}}, \frac{v^u_{X,1}}{s^l_{X,1}}, s^l_{Z,1}, s^l_{X,1}\right), \tag{15}$$

where

$$\gamma(a, b, c, d) = \sqrt{\frac{(c+d)(1-b)b}{cd\log(2)} \log_2\left(\frac{c+d}{cd(1-b)b}\right)\frac{21^2}{a^2}}. \tag{16}$$

Lastly, we give an estimate of the parameters λ_{EC}, i.e. the size of the information exchanged (number of disclosed bits) during the error-correction step. Since we are not going to implement an actual error correction algorithm in the analysis, we follow the example showed in [2] and set it to a simple function $f_{\text{EC}} h(\text{QBER}_Z)$, where f_{EC} is the error correction efficiency and $h(\text{QBER}_Z)$ is the binary entropy computed on the QBER in the key basis (Z in our case). We fix $f_{\text{EC}} = 1.5$.

We have now all the terms needed to estimate the secret key length.

2 Dataset

The experiment is run for 3519 s. We dispose of three binary files of raw keys, each containing:

– Alice's choices of basis and state;

– Alice's choices of decoy state (decoy or signal);

– Bob's detected states.

The raw keys are obtained from a QKD run after synchronisation and after having discarded the qubits that were not received by Bob.
Each file contains key blocks of different lengths. A file begins with 8 bytes that code for a `uint64` big-endian, which is the length N of the block (in bytes). After these first 8 bytes, N bytes of raw keys follow. Each block represents 1 s of acquisitions for the QKD protocol.

The basis selection probabilities at the transmitter and receiver are 90% for H/V, 10% for D and 50% for H/V, 50% for D/A respectively. The decoy probabilities are 70%, 30% for signal and decoy respectively and the decoy intensities (i.e. the average number of photons per pulse) are $\mu_1 = 0.4699$ and $\mu_2 = 0.1093$.

3 Analysis and discussion

We load the data and convert them, depending on the encoding, in a tabular format. Since the amount of data makes it impractical to load the entire data-frame into memory, we first read the different block sizes and subsequently process the raw data in blocks of $T = 200\,\mathrm{s}$ ($\mathcal{O}(10^6)$ events). The choice of this particular value will be motivated in the later analysis. Each separate data block is saved to disk for further processing.
The analysis follows the same procedure for all the different blocks: we load the i-th 200 s-chunk, $i = [1, \dots, n]$, $n \coloneqq \lfloor N/T \rfloor$ with N the number of 1 s-chunk in the whole data acquisition. Successively, we sift the data: using the information on the basis used by Alice and Bob at each event we remove all the bits corresponding to a mismatch in the basis choice between the two. We compute n_b, m_b, $n_{b,k}$, $m_{b,k}$ for $b = X,\, Z,\, k = \mu_1,\, \mu_2$. From here, we compute the QBER for both basis. This way of processing data optimises the operations and allows for a better memory usage.
We fix $\varepsilon_1 = \varepsilon_2 = 1/19 \cdot \varepsilon_{\text{sec}}$: this choice is motivated by setting all error terms α_1, α_2, α_3 and ν (carried out in the security analysis in [2]) by a common value ε in the definition of the security parameter ε_{sec} given in [1],

$$\varepsilon_{\text{sec}} = 2\left[\alpha_1 + 2\alpha_2 + \alpha_3\right] + \nu + 6\varepsilon_1 + 4\varepsilon_2. \tag{17}$$

3.1 Finite key effects and SKR

The specific choice of the temporal resolution T is motivated by the fact that for smaller acquisition time intervals, finite key effects become preponderant and return nonsensical (negative) bounds within Eq. (12). For instance, a single 1 s-second block does not contain a sufficient number of events to obtain positive key rates because the confidence interval of the parameters is too large. Obviously, this leads to a trade-off: the larger the key block, the closer one gets to the asymptotic rate, but at the same time the computational cost and latency will be higher - so it will take longer

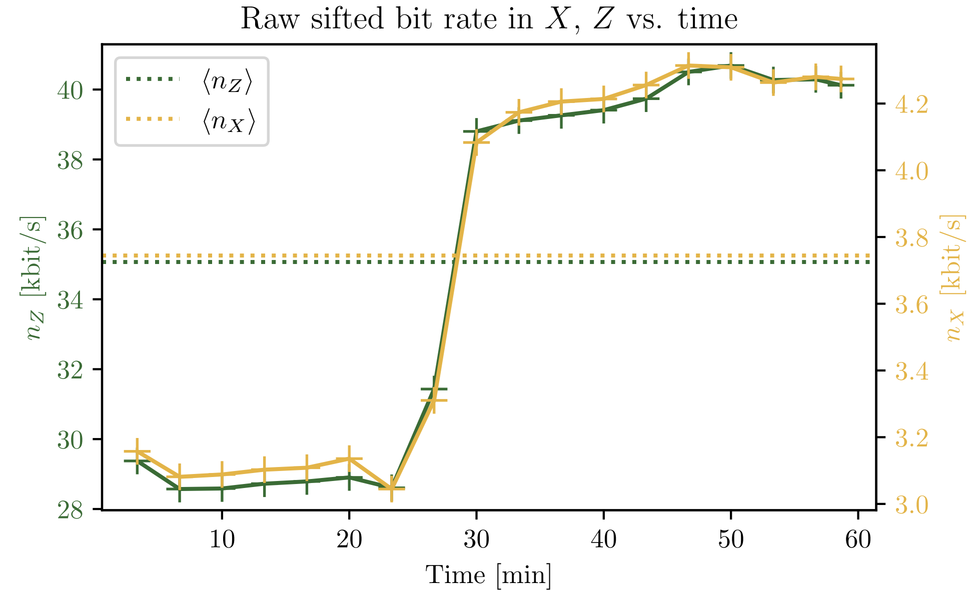

Figure 1. Raw sifted bit rate for the X and Z basis as a function of time. The two curves share a similar trend, but the control basis string is systematically smaller by one order of magnitude with respect to the key basis one. In fact, the ratio $\langle n_Z \rangle / \langle n_X \rangle \sim 9 : 1$ reflects the basis selection probabilities at the transmitter.

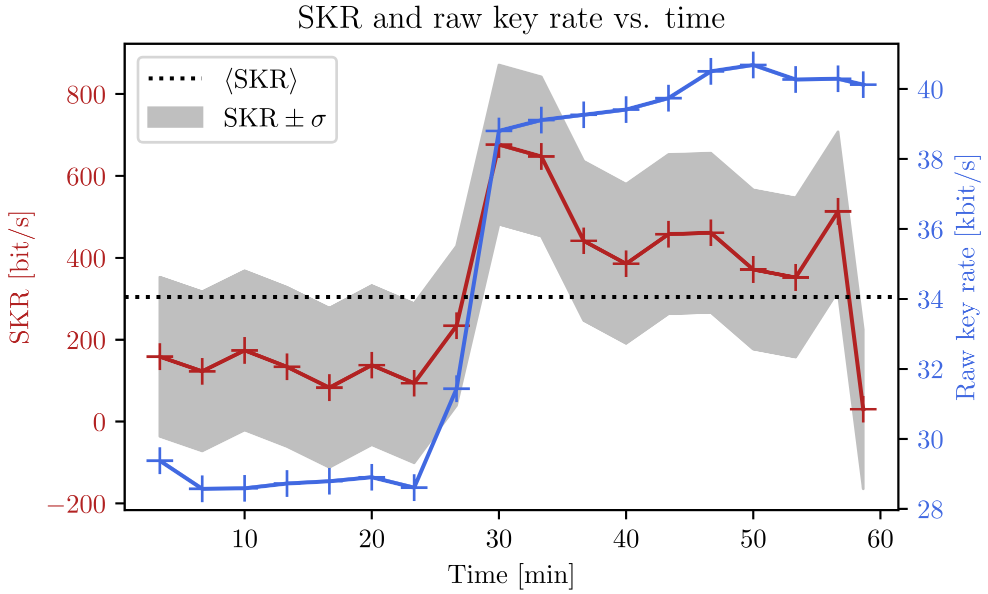

Figure 2. SKR (red curve) and raw key rate (blue curve) as a function of time. The dotted line represents the average value and the shaded area the region of ±1 standard deviation around the SKR. During the whole acquisition time the average of secure bit produced is around 304 bit/s.

to generate a sequence of secure bits.
On the other hand, a similar effect can be observed in Eq. (13) for larger sized blocks: evidently, for the case under study the boundaries estimated in [1] are not valid and return useless values, since the number of detections at Bob's side should in any case be lower-bounded by 0. To overcome this we set the bound to $\max(s^l_{Z,0}, 0)$.
In Fig. 1 we show the raw key stream for both the X and Z basis. The trend presents a steady

behaviour before and after a time $t \sim \tau = 30$ min due to deteriorating channel condition that were not further investigated.

For each of the n blocks, we estimate the bound on the secret key length ℓ: we then divide by T to obtain a proper rate. In Fig. 2 we show the obtained SKR as a function of time, i.e. for each consecutive T-long block. We can observe how the value suddenly increases at $t \sim \tau$, corresponding to a strong increase in the raw key rate. The very last point has been normalised to 119 instead of 200, according to the different duration of the last (smaller) block: due to the reduced number of counts the points has a significant drop compared to the prevailing trend.

3.2 Cumulative analysis and asymptotic behaviour

Additionally, we study the behaviour of the cumulative SKR as a function of the sifted block length: practically, we consider iteratively the cumulative counts of the first j blocks for $j = [1, \ldots, n]$, calculating for each the SKR as the SKL over the equivalent time in seconds. The results are showed in Fig. 3. The plot answer to the question: what would be the rate of secure bits generated by considering larger and larger chunks of the input sifted block length?
Due to the security requirements we impose and eventual imperfections in the communication channel, there is a limited SKR we can achieve: we expect that for a long enough input key we saturate the bounds in Eq. (1), to approach the ∞-key case. Indeed, as one can observe, the curve tends to a constant value as the block length in input increases (the difference between the last two consecutive points is ~ 6.2‰).
We can use this property to quantify how much the finite key effects are impacting the rate: we estimate the asymptotic limit by considering the average value of n_b, m_b, $n_{b,k}$, $m_{b,k}$ for $b = X$, Z, $k = \mu_1$, μ_2 over each chunk in order to obtain an estimate representative of the fluctuations of this values. We then scale it by a large, constant factor: this is to simulate the same experiment, with the same parameters, but with much less losses. In the infinite count limit we can find the asymptotic rate: the gap between this latter value and the rate with the correct counts is an index you how much impact the finite key effects have. Results are shown in Fig. 3. The trend of the curve is compatible with the asymptotic limit: the percentage difference with the last point of the curve is of 13.5%.
As mentioned, provided that one is capable of generating long enough strings of raw bits, the computational time to concretely perform the needed post-processing operations increases with the input key length: one should take this into consideration, given the resources at disposal.

3.3 QBER estimation

Lastly, we study the behaviour of the QBER as a function of time with a time resolution of 1 s. This is made possible by the fact that finite key effects do not arise in this case. In Fig. 4 we show the resulting curves. In this case the kink at $t \sim \tau$ is well evident for the QBER in the key basis; moreover, the spread of the distribution is clearly smaller for the key basis, as one can immediately asses by computing the two variances for the first 1000 s: $\mathrm{var}(\mathcal{Q}_X) \approx 5.7 \cdot \mathrm{var}(\mathcal{Q}_Z)$. The overall averages are $\mathcal{Q}_Z = 2.0\%$ and $\mathcal{Q}_X = 1.1\%$.

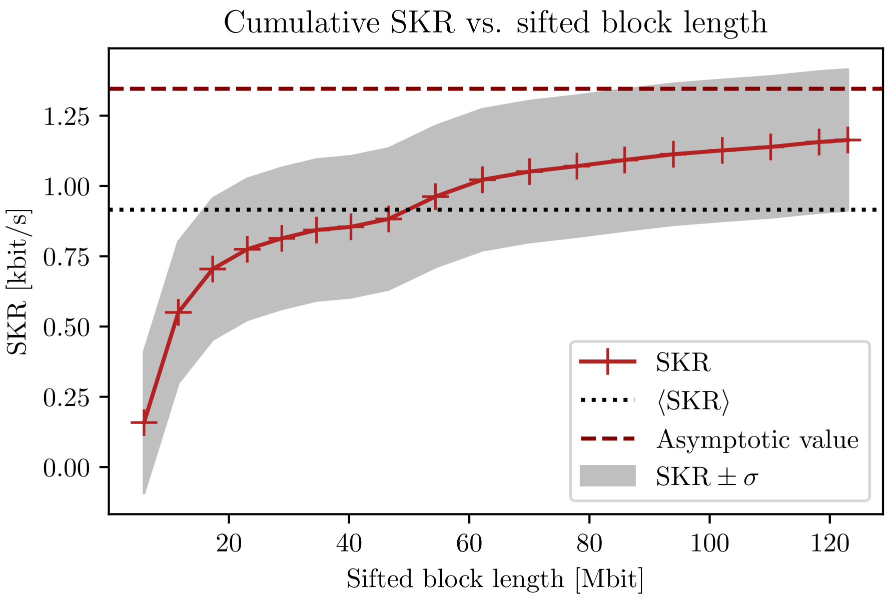

Figure 3. Cumulative SKR as a function of the sifted block length. The SKR has a sudden growth for short raw key length: the slope becomes smaller and smaller for larger input sizes. The asymptotic limit is shown with a dashed line. The grey bands represent ±1 standard deviations.

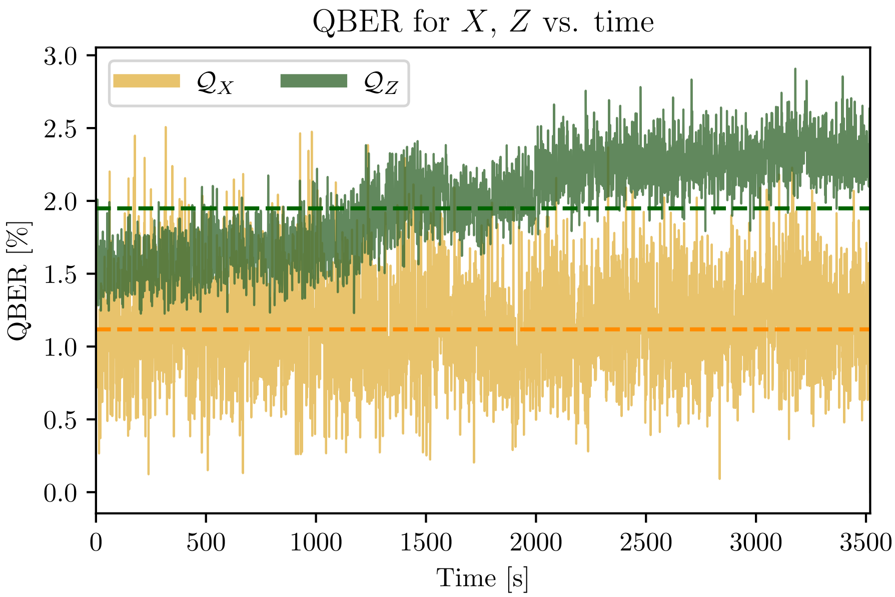

Figure 4. The QBERs $\mathcal{Q}_Z$ and $\mathcal{Q}_X$ in the key and control bases respectively, measured and plotted for every second. The dashed lines represent the overall averages.

4 Conclusion

In this work we have shown an example of a polarization-based QKD system implementing a simplified 3-state and 1-decoy BB84 QKD protocol. We have shown the raw sifted bit rate as well as the (cumulative) SKR as a function of time. Moreover, we plotted the distributions of the QBER for both the key- and control basis.
While we did not find analogous analysis to compare the goodness of our estimates with, we can

affirm that the implementation produced reasonable results and was carried out taking into account the optimisation of the computational resources available.

[1] D. Rusca, A. Boaron, F. Grünenfelder, A. Martin, and H. Zbinden, Applied Physics Letters **112**, 171104 (2018).

[2] C. C. W. Lim, M. Curty, N. Walenta, F. Xu, and H. Zbinden, Physical Review A **89** (2014), 10.1103/physreva.89.022307.