Presentation for the Quantum Seminar

José Manuel Rodríguez Caballero

University of Tartu

Spring 2020

Subject

My presentation is about the paper[1]:

> *Bouman, Niek J., and Serge Fehr. "Sampling in a quantum population, and applications." Annual Cryptology Conference. Springer, Berlin, Heidelberg, 2010.*

My presentation may differ in some points with respect of the presentation given by the authors.

[1] I will not repeat the notation from the paper in this presentation. If someone needs a clarification, either read the paper or ask me during the presentation.

Contributions of the paper

Contribution 1. The authors introduce a framework in for *sampling quantum population*.

Contribution 2. This framework is used in a new proof of the security of the *quantum key distribution protocol BB84* (entanglement-based version[2]).

Contribution 3. This framework is used in a new proof of the security of the *quantum oblivious-transfer from bit-commitment*[3].

[2] This implies the security of the original BB84.

[3] Because of constrains in time, we will focus on contributions 1 and 2. Contribution 3 will be let for those who are interested in reading the paper, because it is rather similar to contribution 2.

Main idea of the paper

The **error estimation** in some quantum key distribution protocols can be understood as applying a **sampling and estimation strategy**.

Presentation of Contribution 1.

The authors introduce a framework in for *sampling quantum population*.

Overview of Contribution 1

The authors define a *sampling and estimation strategy* (*strategy* for short). This strategy can be used in order to study both classical and quantum populations. The "quantum error" is bounded by the square root of the "classical error", thus reducing the quantum information-theoretical problem to a well-studied probabilistic problem.

Strategy

Let n be a positive integer, $\mathcal{A}$ be a finite alphabet and $\mathcal{S}$ be a finite set of seeds. Assume that $|\mathcal{A}| \geq 2$ and $|\mathcal{S}| \geq 1$. A *sampling strategy* is a triplet (P_T, P_S, f), where P_T is a probability distribution over $\mathcal{T} := 2^{[n]}$, P_S is a probability distribution over $\mathcal{S}$ and f is a real-valued function defined on the finite set[4]

$$\mathrm{Dom}_f := \{(t, q, s) \in \mathcal{T} \times \mathcal{A}^* \times \mathcal{S} : \quad |t| = |q|\}.$$

[4] The notation $\mathcal{A}^*$ is for the free monoid over $\mathcal{A}$.

Classical populations

What is a classical population?

In the context of the paper, a *classical population* is a finite word q over the alphabet $\mathcal{A}$. The alphabet $\mathcal{A}$ contains a distinguished element, that we will denote by 0. We are interested in estimating how many letters in q are different from 0, i.e., the Hamming weight of q. In order to simplify the presentation, we will use the relative hamming weight of q, denoted $\omega(q)$, which is the Hamming weight of q divided by its length.

Example of classical population

The classical population could be the population of the People's Republic of Wakanda[5] (each person corresponds to a position in the word q) and the alphabet could be $\mathcal{A} = \{0, 1\}$, where 0 and 1 stand for uninfected and infected of COVID-19, respectively. In order to estimate how many people are infested with COVID-19, we will study a random sample of the population consisting of k people, which will be small in comparison to the size of the population, i.e. k is "small" with respect to n. Finally, we will extrapolate the data from the sample to the whole population.

[5] I am considering an idealized situation, in order to avoid all the complications of the a real-life scenario, e.g., fake data and heterogeneous properties of the population. In Wakanda everything will be as homogeneous as possible.

How to estimate in a classical population?

Given a sampling strategy (P_T, P_S, f) and a classical population $q = q_1 q_2 q_3 ... q_n \in \mathcal{A}^n$. We take a random pair $(t, s) \in \mathcal{T} \times \mathcal{S}$ (using the joint distribution of P_T and P_S). Let

$$\tau_1 < \tau_2 < ... < \tau_k$$

be the elements of t written in increasing order. Let

$$\overline{\tau}_1 < \overline{\tau}_2 < .... < \overline{\tau}_{n-k}$$

be the elements of $[n] \backslash t$ written in increasing order. We define

$$q_t = q_{\tau_1} q_{\tau_2} ... q_{\tau_k}, \quad q_{\overline{t}} = q_{\overline{\tau}_1} q_{\overline{\tau}_2} ... q_{\overline{\tau}_{n-k}}.$$

By definition, we *estimate* that the relative Hamming weight of $q_{\overline{t}}$ is given by $f(t, q_t, s)$.

Example of estimation in a classical population

Let's return to the example of the population of Wakanda, identified with the word q. We define $\mathcal{S} = \{\perp\}$, i.e., we do not use a random seed here. So, P_S is trivially the uniform distribution on $\mathcal{S}$. Now, let assume that P_T is the uniform distribution on subsets of size k of $[n]$, i.e.,

$$P_T(t) = \frac{1}{\binom{n}{k}}$$

if $|t| = k$ and $P_T(t) = 0$ otherwise. Finally, define

$$f(t, q, s) := \omega(q).$$

This framework is known as *random sampling without replacement*.

Example of estimation in a classical population (continuation)

This sampling strategy is just a formalization of the heuristic approach assuming that the frequency of cases of COVID-19 in the whole population is the same as in our uniformly random sample. This is not necessarily the case in reality, e.g., people working in hospitals have a higher probability of getting infested than the remain of the population.

Set of classical δ-close states

Let δ be a positive real number. We define the *set of classical δ-close states* as

$$B^{\delta}_{t,s} := \{q \in \mathcal{A}^n : \quad |\omega(q_{\bar{t}}) - f(t, q_t, s)| < \delta\},$$

This is the set of states for which the estimation is fine up to the precision δ.

Example of set of classical δ-close states

In Wakanda example, $B^{\delta}_{t,s}$ is the set of all possibilities (after choosing t and s) in which the frequency of infection of COVID-19 in the sample is δ-close to the frequency of infection in the remaining part of the population of Wakanda.

Random set of classical δ-close states

Let T and S be random variables associated to the probability distributions P_T and P_S respectively. Notice that the pair (T, S) is a random variable. Furthermore, the evaluation of $(t, s) \mapsto B^{\delta}_{t,s}$ at (T, S) determines a random variable, denoted $B^{\delta}_{T,S}$ and associated to the probability distribution

$$\Pr\left[B^{\delta}_{T,S} \in \Gamma\right] := \Pr\left[(T, S) \in \left\{(t, s) \in \mathcal{T} \times \mathcal{S} : \quad B^{\delta}_{t,s} \in \Gamma\right\}\right]$$

where $\Gamma \subset 2^{\mathcal{A}^n}$. The call $B^{\delta}_{T,S}$ the *random set of classical δ-close states*.

Example of random set of classical δ-close states

In Wakanda example, $B^{\delta}_{T,S}$ is the random set of all possibilities (before choosing t and s) in which the frequency of infection of COVID-19 in the sample is δ-close to the frequency of infection in the remaining part of the population of Wakanda.

Classical δ-error

The *classical* δ*-error* of a strategy is defined as

$$\varepsilon_c^\delta := \max_{q \in \mathcal{A}^n} \Pr\left[B_{T,S}^\delta \in \left\{ X \in 2^{\mathcal{A}^n} : \quad q \notin X \right\} \right].$$

ε_c^δ measures the probability that the sampling strategy fails in the worst-case scenario[6].

[6] Notice that the expression inside the bracket is equivalent to $q \notin B_{T,S}^\delta$.

Example of classical δ-error

In Wakanda example, ε_c^δ is the largest probability, with respect to all possible populations, that the frequency of infection of the sample is not δ-close to the frequency of infection of the remaining part of the population of Wakanda.

For $k \leq \frac{n}{2}$ (as it should be in practice), we have

$$\varepsilon_c^\delta \leq 2 \exp\left(-\frac{\delta^2 k}{2}\right).$$

Quantum populations

What is a quantum population?

In the context of the paper, a *quantum population* is a quantum state from $\mathcal{H}_A \otimes \mathcal{H}_E$, where $\{|q\rangle\}_{q \in \mathcal{A}^n}$ is an orthonormal basis of the Hilbert space $\mathcal{H}_A$ (populations as such) and $\mathcal{H}_E$ (environment) is also a finite dimensional Hilbert space.

Example of quantum population

Imagine that People's Republic of Wakanda isolated itself from the rest of the universe in such a way that information exchange with the exterior is no longer possible[7]. Inside a lab of virology, there was an accident and some particles of the virus SARS-CoV-2 (associated to the illness COVID-19) escaped to the lab of physics. More precisely, the viral particles fall into a radioactive sample. If the radioactive sample decay, the radiation will kill the virus, but until then, all the physicists using the sample will get infected of COVID-19.

[7] This mental experiment is a variation of the famous Schrödinger's cat.

Example of quantum population (continuation)

From the point of view of an outsider, the population of Wakanda will be a quantum population: the health state (infested or not) of the people (population as such) will be entangled with the quantum state of the above-mentioned radioactive sample (environment).

How to estimate in a quantum population?

Given a sampling strategy (P_T, P_S, f) and a quantum population $|\varphi_{AE}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_E$. We take a random pair $(t, s) \in \mathcal{T} \times \mathcal{S}$ (using the joint distribution of P_T and P_S). We apply the measurement of the system A in the positions indexed by the elements of t. The result will be classical population q_t. By definition, we *estimate* that the post-measurement state of $|\varphi_{AE}\rangle$, denoted $|\varphi_{A_{\bar{t}}E}\rangle$ is a superposition of states δ-close to $f(t, q_t, s)$.

Example of estimation in a classical population

The equivalent of a measurement of the part t of a quantum population is to ask for people corresponding to the numbers in t to get out Wakanda (without destroying the quantum entanglement of the remaining population) and to be tested for COVID-19 outside. From this information, we should be able to estimate the frequency of people in the remaining quantum population of Wakanda (post-measurement state).

δ-fidelity

For any $|\varphi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_E$, the δ-*fidelity* of the strategy at (t, s), with respect to the environment $\mathcal{H}_E$, is

$$f_{t,s}^{\delta}\left(|\varphi\rangle\right) = \sup_{|\psi\rangle \in \mathsf{span}\left(B_{t,s}^{\delta}\right) \otimes \mathcal{H}_E} |\langle\psi|\varphi\rangle|^2.$$

Quantum δ-error

The *quantum* δ*-error* of a strategy is defined as

$$\varepsilon_q^\delta := \sup_{\mathcal{H}_E} \sup_{|\varphi_{AE}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_E} \sum_{(t,s)\in \mathcal{T}\times\mathcal{S}} P_{T,S}(t,s)\sqrt{1 - f_{t,s}^\delta\left(|\varphi_{AE}\rangle\right)}.$$

Classical-quantum inequality for δ-error

For any strategy and for any precision $\delta > 0$, we have

$$\varepsilon_q^\delta \leq \sqrt{\varepsilon_c^\delta}.$$

Example of application of the classical-quantum inequality for δ- error

The bound for the classical δ-error for the strategy applied to the population of Wakanda

$$\varepsilon_c^{\delta} \leq 2 \exp\left(-\frac{\delta^2 k}{2}\right), \quad \text{provided } k \leq \frac{n}{2},$$

implies the following bound for the quantum δ-error for the same strategy

$$\varepsilon_q^{\delta} \leq \sqrt{2} \exp\left(-\frac{\delta^2 k}{4}\right), \quad \text{provided } k \leq \frac{n}{2}.$$

Presentation of Contribution 2.

This framework is used in a new proof of the security of the *quantum key distribution protocol BB84* (entanglement-based version).

Outline of the presentation of contribution 2

- We will present/recall the protocol BB84 (entanglement based).
- We will present the formalization of the statement that this protocol is secure.
- We will explain how the framework of sampling and estimating in a quantum population can be used in order to prove the security of this protocol.

Entanglement based BB84 (outline)

Step 1. Qubit distribution.

Step 2. Error estimation.

Step 3. Error correction.

Step 4. Key distillation.

Entanglement based BB84 (Step 1)

Qubit distribution. Alice prepares n EPR pairs of the form $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and sends one qubit of each pair to Bob, who confirm the receipt of the qubits. Then, Alice picks random $\theta \in \{0,1\}^n$ and sends it to Bob, and Alice and Bob measure their respective qubits in basis θ to obtain x on Alice's side respectively y on Bob's side.

In the proof of security, the random variables corresponding to θ, x and y will be denoted by their respective capital letters Θ, X and Y.

Entanglement based BB84 (Step 2)

Error estimation. Alice chooses a random subset $s \subset [n]$ of size k and sends it to Bob. Then Alice and Bob exchange x_s and y_s and compute $\beta := \omega(x_s \oplus y_s)$.

Entanglement based BB84 (Step 3)

Error correction. Alice sends the syndrome syn of $x_{\overline{s}}$ to Bob with respect to a suitable linear error correcting code. Bob uses syn to correct the errors in $y_{\overline{s}}$ and obtains $\hat{x}_{\overline{s}}$. Let m be the bit-size of syn.

Entanglement based BB84 (Step 4)

Key distillation. Let $h(\beta) := -\beta \log \beta - (1-\beta)\log(1-\beta)$ be the binary entropy[8]. Alices chooses a random seed r for a universal hash function g with range $\{0,1\}^\ell$, where ℓ satisfies $\ell < (1 - h(\beta))\, n - k - m$ (or $\ell = 0$ if the right-hand side is not positive), and sends it to Bob. Then, Alice and Bob compute $k := g(r, x_{\overline{s}})$ and $\hat{k} := g(r, \hat{x}_{\overline{s}})$.

[8] The notation log stands for the binary logarithm.

Security of BB84 (entanglement based)

Theorem. *Consider an execution of BB84 (entanglement based) in the presence of an adversary Eve. Let* $\mathbf{K}$ *be the key obtained by Alice, and let* E *be Eve's quantum system at the end of the protocol. Let* $\tilde{\mathbf{K}}$ *be chose uniformly at random of the same bit-length as* $\mathbf{K}$*. Then, for any* δ *with* $\beta + \delta \leq \frac{1}{2}$*, we have*

$$\Delta\left(\rho_{\mathbf{K}E}, \rho_{\tilde{\mathbf{K}}E}\right) \leq \frac{1}{2} \cdot 2^{-\frac{1}{2}((1-h(\beta+\delta))n-k-m-\ell)} + 2\exp\left(-\frac{1}{6}\delta^2 k\right).$$

Sketch of the proof

Idea 1. The error estimation (step 2) in the protocol can be understood as applying a sampling and estimation strategy (described in contribution 1).

Idea 2. The state from which the raw key $x_{\overline{s}}$ is obtained is a superposition of states with bounded Hamming weight.

Idea 3. There is a lower bound of min-entropy for $x_{\overline{s}}$.

Idea 4. Apply privacy amplification in order to end the proof.

Ideas 2, 3 and 4 are standard techniques in quantum cryptography. So, we will focus on idea 1, which is the only one related to the contributions of the paper.

Idea 1

The error estimation (step 2) in the protocol is replaced by the following sampling and estimation strategy.

Let $|\psi_{ABE_0}\rangle$ be the state of Alice, Bob and Eve right after the quantum communication in the qubit distribution. The `CNOT` transformation is applied to every qubit pair A_iB_i within $|\psi_{ABE_0}\rangle$ for $i \in [n]$, such that the state

$$|\varphi_{ABE_0}\rangle = \left(U_{\text{CNOT}}^{\otimes n} \otimes \mathbb{I}_{E_0}\right) |\psi_{ABE_0}\rangle$$

is obtained.

Idea 1 (continuation)

For any $i \in [n]$, the qubit pair $A_i B_i$ of the transformed state is measured and if $\Theta_i = 0$ then the output bits are denoted W_i and Z_i, else they are denoted Z_i and W_i.

Notice that, given Θ, from W and Z, we can recover X and Y and conversely.

Idea 1 (continuation)

We assume that first the qubits that lead to the Z_i's are measured and only at some later point (after step 2 in the original protocol) the qubits leading to the W_i's are measured.

Idea 1 (continuation)

Thus, the error estimation is reduced to the following strategy, known as *pairwise one-out-of-two sampling*[9].

Consider the index set from which the subset t is chosen to be $[n] \times \{0, 1\}$. Assume that t is chosen as $t = \{(1, j_1), ..., (n, j_n)\}$, where every j_k is picked independently at random in $\{0, 1\}$. Furthermore, s is a random subset of t of size k. We define $f(t, q_t, s) = \omega(q_s)$.

[9] It is easy to check that it can be reformulated according to our general definition of strategy.

End of my presentation