COM-440, Introduction to Quantum Cryptography, Fall 2025

Exercise Solution # 9

1. **Information reconciliation based on 2-universal hash functions**

   (a) Given a pair (x, y), by using the 2-universal property, for any $x' \in X_y$ and $x' \neq x$, we have that

   $$\mathop{\mathbf{Pr}}_f(f(x') = f(x)) = \sum_{z \in \{0,1\}^k} \mathop{\mathbf{Pr}}_f(f(x') = z \wedge f(x) = z) = 2^k \cdot 2^{-2k} = 2^{-k} \ ,$$

   which means $\mathbf{Pr}_f(x' \in D) = 2^{-k}$.
   Therefore, by union bound,

   $$\begin{aligned} \mathop{\mathbf{Pr}}_f(x' \neq x) &\leq \mathop{\mathbf{Pr}}_f(\exists\, x' \in X_y, x' \neq x \wedge x' \in D) \\ &= \sum_{x' \in X_y, x' \neq x} \mathop{\mathbf{Pr}}_f(x' \in D) \\ &< |X_y| \cdot 2^{-k} \ . \end{aligned}$$

   (b) From the definition of the conditional max-entropy, for every pair (x, y) such that $P_{XY}(x, y) > 0$, $\mathbf{Pr}_f(x' \neq x) \leq 2^{H_{max}(X|Y)-k}$.
   Therefore, with probability $\mathbf{Pr}_{f,(x,y)\sim P_{XY}}(x' \neq x) \leq 2^{H_{max}(X|Y)-k}$, Bob will output the same bitstring as Alice. So the protocol is ε-correct when $2^{H_{max}(X|Y)-k} \leq \varepsilon$. We can take $k = \lceil H_{max}(X|Y) + \log(1/\varepsilon) \rceil$.

   (c) This is a bit of an open question. Technically, both (the description of) f and z are leaked in this protocol, which is $2n + k$ bits, which is large. But one can observe that the communication of the function f itself does not depend on anything, and so it does not reveal information about x or y. So, from the point of view of using the protocol for QKD, the leakage is only k bits.

   (d) For $\delta \in (0, 1)$, it is n because $P_{XY}(x, y) > 0$ for every pair (x, y).

   (e) The idea is to truncate those (x, y) that appears rarely.
   Let P'_{XY} be a renormalization of P_{XY} after excluding those (x, y) such that they differ on at least $n(\delta + \delta')$ positions for some parameter δ' that we will decide later. Then $\|P'_{XY} - P_{XY}\|_{TV} = 2 \sum_{(x,y):P'_{XY}(x,y)<P_{XY}(x,y)} (P_{XY}(x, y) - P'_{XY}(x, y))$. Since $P'_{XY}(x, y) < P_{XY}(x, y)$ only on those points we exclude, and $P'_{XY}(x, y) = 0$ on those point, we can continue the equation to get

   $$\begin{aligned} \|P'_{XY} - P_{XY}\|_{TV} &= 2 \left(\sum_{(x,y):x \text{ and } y \text{ differ on at least } n(\delta+\delta') \text{ positions}} P_{XY}(x, y) \right) \\ &= 2 \cdot \mathop{\mathbf{Pr}}_{(x,y)\sim P_{XY}(x,y)}(x \text{ and } y \text{ differ on at least } n(\delta + \delta') \text{ positions}) \\ &\leq 2 \cdot \exp(-2n(\delta')^2) \ . \qquad \text{(Hoeffding's inequality)} \end{aligned}$$

Therefore, we can set $\delta' = \sqrt{\frac{\ln(2/\varepsilon')}{2n}}$ to make $\|P'_{XY} - P_{XY}\|_{TV} \leq \varepsilon'$.

Now let's how that for P', $H_{max}(X|Y)$ is relatively small. This is because for P', X_y consists of x that differ with y on at most $n(\delta + \delta')$ positions. Therefore, for P', $|X_y| \leq \sum_{m \leq n(\delta+\delta')} \binom{n}{m}$.

We assume $\delta + \delta' < \frac{1}{2}$ (which is the case for $\delta < \frac{1}{2}$ and large enough n). Then $\sum_{m \leq n(\delta+\delta')} \binom{n}{m} \leq n\binom{n}{n(\delta+\delta')}$. By Stirling's formula, there exists a constant c such that $\binom{n}{n(\delta+\delta')} \leq c \cdot \frac{1}{\sqrt{n}} \cdot 2^{-n((\delta+\delta')\log(\delta+\delta')+(1-\delta-\delta')\log(1-\delta-\delta'))} = c \cdot \frac{1}{\sqrt{n}} \cdot 2^{n \cdot H(\delta+\delta')}$. Therefore, $|X_y| \leq c(\delta+\delta')\sqrt{n} \cdot 2^{n \cdot H(\delta+\delta')}$ and thus $H_{max}(X|Y) \leq n \cdot H(\delta+\delta') + c' \log n$ for some constant c'.

δ' goes to zero as $n \to \infty$, and $H(\delta)$ goes to zero as $\delta \to 0$, and thus we find an upper bound for $H_{max}(X|Y)/n$, $H(\delta + \delta') + c' \frac{\log n}{n}$, which goes to 0 as $\delta \to 0$ and $n \to \infty$.

(f) By part (b), if $(x, y) \sim P'_{XY}$, we would get that the protocol is ε-correct for $k = \lceil C(\delta)n + \log(1/\varepsilon) \rceil$.

Since $\|P'_{XY} - P_{XY}\|_{TV} \leq \varepsilon'$, the joint output distribution of Alice and Bob when the input $(x, y) \sim P_{XY}$ is ε'-close to the joint output distribution of Alice and Bob when the input $(x, y) \sim P'_{XY}$.

Therefore, where $(x, y) \sim P_{XY}$, the protocol is $(\varepsilon + \varepsilon')$-correct using the parameter k that would be computed from the distribution P'.