HAI709I : Fondements cryptographiques de la sécurité, Université de Montpellier, 2023

02/10/2023. Lecture 4.

1 Text compression

In the class we rediscussed the proof of the following fact.

Lemma 1. *If a set of binary words $\{c_1, \ldots, c_k\}$ is a prefix-free code, then* $\sum\limits_{i=1}^{k} 2^{-|c_i|} \le 1$.

We also discussed in which case the sum $\sum\limits_{i=1}^{k} 2^{-|c_i|}$ is strictly less than 1.

Another useful lemma claims that the inequality $\sum\limits_{i=1}^{k} 2^{-|c_i|}$ is not only necessary but also sufficient to build a prefix-free code with the given lengths of codewords.

Lemma 2. *For every set of natural number $\ell_1, \ldots, \ell_k$, if $\sum i = 1^k 2^{-\ell_i} \le 1$, then there exists a prefix-free code $\{c_1, \ldots, c_k\}$ such that $|c_i| = \ell_i$ for $i = 1, \ldots, k$.*

Reminder of the proof: First of all, we sorted the lengths ℓ_i. In what follows we assume w.l.o.g. that $\ell_1 \le \ell_2 \le \ldots \le \ell_k$. Then for each $i = 1, \ldots, k$ we choose the binary word c_i of length ℓ_i that is lexicographically first among all possible (i.e., non-extending any of the words $c_1, \ldots c_{i-1}$ fixed before). We verified that the construction works properly until $i = k$ if $\sum i = 1^k 2^{-\ell_i} \le 1$.

We used these lemmas to prove the theorem on optimal compression:

Theorem 1. *For any distribution of probabilities $(p_1, \ldots, p_k)$ there exists a prefix-free codeword $\{c_1, \ldots, c_k\}$ such that*

$$\sum_{i=1}^{k} p_c |c_i| < \sum_{i=1}^{k} p_i \log \frac{1}{p_i} + 1.$$

Idea of the proof discussed in the class: We let $\ell_i = \lceil \log \frac{1}{p_i} \rceil$. It is not difficult to verify that $\sum\limits_{i=1}^{k} 2^{-|c_i|} \le 1$. So we can use Lemma 2 and constructed a prefix-free code with $|c_i| = \ell_i$. It remains to show that with the chosen ℓ_i we have

$$\sum_{i=1}^{k} p_i \ell_i < \sum_{i=1}^{k} p_i \log \frac{1}{p_i} + 1.$$

2 Properties of Shannon's entropy

The joint distribution of a pair of random variables (X, Y) is a table of numbers p_{ij} such that

$$p_{ij} = \mathrm{Prob}[X = a_i \text{ et } Y = b_j].$$

We use the notation

$$p_{i*} = \sum_j p_{ij} = \mathrm{Prob}[X = a_i]$$

and

$$p_{*j} = \sum_i p_{ij} = \mathrm{Prob}[Y = y_j].$$

By definition of conditional probability,

$$\mathrm{Prob[Y = b_j \mid X = x_i]} = \frac{p_{ij}}{p_{i*}}.$$

In the last lecture we defined the notion of Shannon's entropy for an individual random variable,

Definition 1. For a random variable A with n possible values $a_1, \ldots, a_n$ such that $\mathrm{Prob}[A = a_i] = p_i$, we define its Shannon's entropy as

$$H(A) := \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$$

(with the usual convention $0 \cdot \log \frac{1}{0} = 0$).

Now we discuss properties of pairs of jointly distributed random variables. Given a pair of jointly distributed random variables (X, Y) we can apply the definition of Shannon's entropy three times, with three protentially different distributions: we have Shannon's entropy of the entire distribution of the pair denoted $H(X, Y)$, and the entropies of two marginal distributions X and Y, denoted $H(X)$ and $H(Y)$.

Proposition 1. *For every pair of jointly distributed random variables X and Y*

$$H(X, Y) \leq H(X) + H(Y).$$

Moreover, the equality

$$H(X, Y) = H(X) + H(Y)$$

holds if and only if A and Y are independent, i.e., for all i and j

$$\mathrm{Prob}[X = a_i \textit{ and } Y = b_j] = \mathrm{Prob}[X = a_i] \cdot \mathrm{Prob}[Y = b_j]$$

Idea of the proof: We used one more time the concavity of the function of logarithm and Jensen's inequality. □

Definition 2. Let (X, Y) be jointly distributed random variables, with

$$p_{ij} = \mathrm{Prob}[X = a_i \text{ and } Y = b_j].$$

For each value A_j with a positive probability we have the *conditional distribution* on the values of Y with probabilities

$$p'_j = \mathrm{Prob}[Y = b_i \mid X = a_i] = \frac{\mathrm{Prob}[X = a_i \text{ and } Y = b_j]}{\mathrm{Prob}[X = a_i]}.$$

This conditional distribution has its own Shannon's entropy; we denote it $H(Y \mid A = a_i)$.

Definition 3. We define the entropy of Y conditional on X as the average

$$H(Y \mid X) := \sum_i \mathrm{Prob}[X = a_i] \cdot H(Y \mid X = a_i).$$

In the class we proved the following properties of *conditional entropy*.

Proposition 2. *For all jointly distributed random variables* (X, Y)
(a) $H(X, Y) = H(X) + H(Y \mid X)$*,*
(b) $H(X \mid Y) \le H(X)$*.*
(c) Moreover, $H(X \mid Y) = H(X)$ *if and only if* X *and* Y *are independent.*

Definition 4. For a pair of jointly distributed random variables (X, Y) we define the information in X on Y as

$$I(X : Y) = H(Y) - H(Y \mid X).$$

Proposition 3. *For all jointly distributed* (X, Y)

- $I(X : Y) = I(Y : X) = H(X) + H(Y) - H(X, Y)$*,*
- *moreover,* $I(X : Y) = 0$ *is and only if* X *and* Y *are independent.*

(the proofs discussed in the class)

Exercise 1. Prove that for all jointly distributed (X, Y, Z)

$$2H(X, Y, Z) \le H(X, Y) + H(X, Z) + H(Y, Z).$$

3 Limits on compression of the secret key

The next theorem claims that we cannot make the secret key "too well-compressible" (below the threshold H(clear message)) without loosing security of the encryption scheme.

Theorem 2. *Let* (M, K, E) *(a clear message, a secret key, an encrypted message) be a triple of jointly distributed random variables satisfying two properties:*

(i) $H(M \mid K, E) = 0$ *(the clear message can be uniquely reconstructed given the secret key and the encoded message)*

(ii) $H(M \mid E) = H(M)$ *(the encrypted message gives no information on the open message).*

Then $H(K) \ge H(M)$ *(Shannon's entropy of the secret key is not less than Shannon's entropy of the clear message).*

Proof. We consider Shannon's entropy of the triple $H(M, K, E)$. On the one hand, we have

$$H(M, K, E) = H(K, E) + H(M \mid K, E) = H(K, E) + 0 \le H(K) + H(E)$$

(we used here Property (i)). On the other hand,

$$H(M, K, E) = H(M, E) + H(K \mid M, E) \ge H(M, E) = H(M \mid E) + H(E) = H(M) + H(E).$$

(this time we used Property (ii)). Combining these two observations we obtain $H(K) \ge H(M)$. □

References

[1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. Introduction to Algorithms, Second Edition. MIT Press and McGraw-Hill, 2001.

[2] V. V. Yaschenko, Cryptography: An Introduction, AMS, 2002

[3] B. Martin. Codage, cryptologie et applications. PPUR presses polytechniques, 2004