Computer Science COMP-547A
Cryptography and Data Security

Claude Crépeau

These notes are, largely, transcriptions by Anton Stiglic of class notes from the former course *Cryptography and Data Security (308-647A)* that was given by prof. Claude Crépeau at McGill University during the autumns of 1998 and 1999. These notes are updated and revised each year by Prof. Claude Crépeau.

Latest update January 7, 2009.

6 Information theory

The security of most modern cryptographic system is based on a computational assumption. In large, a system that is computationally secure relies on the fact that an attacker does not have enough **time** to break the system, while on the other hand a system that is unconditionally secure relies on the fact that an attacker will never have enough **information**. Computational security uses results from *complexity theory*, unconditional security uses *information theory* which we present in this chapter.

6.1 Notations and probability theory

Let X, Y be distributions on finite sets of events (we will only consider finite sets of probabilities), we denote $\Pr_X(x)$ as the probability of x occurring, following the distribution X (when the situation is clear, we will not explicitly write the distribution on the index).
$\Pr_{X,Y}(x, y)$ is the **mutual probability**, it is the probability that both x and y occur.
$\Pr_{X,Y}(x|y)$ is the **conditional probability**, it is the probability that x occurs knowing that y has occurred.

Definition 6.1 *X and Y are two* **independent** *distributions if*

$$\forall_{x \in X, y \in Y} \ \Pr(x, y) = \Pr(x) \Pr(y)$$

By definition, we have that $\Pr(x, y) = \Pr(x|y) \Pr(y) = \Pr(y|x) \Pr(x)$.
We can now deduce the following theorem

Theorem 6.2 *(Bayes)*

$$If \ \Pr(y) > 0 \ then \ \Pr(x|y) = \frac{\Pr(y|x) \Pr(x)}{\Pr(y)}$$

We can easily convince ourselves of the following corollary

Corollary 6.3 *X and Y are independent variables iff* $\forall_{x \in X, y \in Y}$

$$\Pr(x|y) = \Pr(x)$$

6.2 Shannon's information theory and entropy

Consider a random source $\mathcal{S} \longrightarrow 100110100111011...$ We would like to characterize its output. $\mathcal{S}$ outputs symbols with probabilities $\Pr(0), \Pr(1)$. How much uncertainty do we have about the output of this source (what is the entropy $\mathbf{H}[\mathcal{S}]$ of $\mathcal{S}$)?

Example 6.1 $\mathcal{S}_0 \longrightarrow 00000000...$ *(always outputs* 0*), then* $\mathbf{H}[\mathcal{S}_0] = 0$.
$\mathcal{S}_1 \longrightarrow 111111111...$ *(always outputs* 1*), then* $\mathbf{H}[\mathcal{S}_1] = 0$.
$\mathcal{S}_U \longrightarrow 100110100111011...$ *(Independent and Uniform), then* $\mathbf{H}[\mathcal{S}_U] = 1$.

In a more general case, how do we define $\mathbf{H}[\mathcal{S}]$?

6.2.1 Entropy

The entropy of a source $\mathcal{S}$ can be considered as the average length, in bits, needed to represent the output of $\mathcal{S}$ (this was introduced in [**?**]).

Example 6.2 *Consider a random source* $\mathcal{S} \longrightarrow a, b, c$*, where*

$$\Pr(a) = 1/2, \Pr(b) = 1/4 = \Pr(c) = 1/4$$

and the coding $a \longrightarrow 0, b \longrightarrow 10, c \longrightarrow 11$ *(a prefix coding), then the average length needed to represent the output of* S *is*

$$1/2 \cdot 1 \text{ bit } + 1/4 \cdot 2 \text{ bits } + 1/4 \cdot 2 \text{ bits } = 3/2 \text{ bits.}$$

More formally, we state the following definition of the entropy of a source $\mathcal{S}$, where the probabilities of the outputs are $p_1, p_2, ..., p_n$,

Definition 6.4 (Entropy)

$$\mathbf{H}[p_1, p_2, ..., p_n] = \sum_{i=1}^{n} p_i \lg \frac{1}{p_i} = -\sum_{i=1}^{n} p_i \lg p_i$$

where lg *represents the logarithm base 2. We assume in this definition that* $0 \lg 0 = 0$ *(the real fact is that* $\lim_{x \longrightarrow 0^+} x \lg x = 0$*).*

Another way of defining the entropy is through the following set of properties

6.2.2 Properties of entropy

$\mathbf{H}[p_1, p_2, ..., p_n]$, where $\sum_{i=1}^{n} p_i = 1$, satisfies the following:

1. $\mathbf{H}[p_1, p_2, ..., p_n]$ is maximum when $p_1 = p_2 = ... = p_n = 1/n$ *(the uniform distribution should have the greatest entropy...).*
2. $\mathbf{H}[p_1, p_2, ..., p_n] = \mathbf{H}\left[p_{\pi(1)}, p_{\pi(2)}, ..., p_{\pi(n)}\right]$ for any permutation π.
3. $\mathbf{H}[p_1, p_2, ..., p_n] \geq 0$, with $= 0$ exactly if one $p_i = 1$.
4. $\mathbf{H}[p_1, p_2, ..., p_n, 0] = \mathbf{H}[p_1, p_2, ..., p_n]$
5. $\mathbf{H}\left[\underbrace{\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n}}_{n}\right] \leq \mathbf{H}\left[\underbrace{\frac{1}{n+1}, \frac{1}{n+1}, ..., \frac{1}{n+1}}_{n+1}\right]$
6. $\mathbf{H}[*]$ is continuous

Theorem 6.5 *Let $H'(p_1, p_2, ..., p_n)$ satisfy the above properties. There exists a constant λ such that*

$$H'(p_1, p_2, ..., p_n) = \lambda \mathbf{H}[p_1, p_2, ..., p_n].$$

6.2.3 Entropy of random variables

For a random variable X such that $\Pr_X(x_1) = p_1, ..., \Pr_X(x_n) = p_n$ we abuse notation and write $\mathbf{H}[X]$ instead of $\mathbf{H}[p_1, p_2, ..., p_n]$.

Look ahead: For a cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$, we will be interested in random variables P, K, C describing the uncertainty about the plaintext, ciphertext and key, and in $\mathbf{H}[P|C]$ (uncertainty of the message given the ciphertext) and $\mathbf{H}[K|C]$ (uncertainty of the encryption key given the ciphertext).

Now, a couple more theorems and definitions...

Theorem 6.6

$$\mathbf{H}[X, Y] \leq \mathbf{H}[X] + \mathbf{H}[Y]$$

(equal iff X and Y are mutually independent)

Theorem 6.7

$$\mathbf{H}[X_1, X_2, ..., X_n] \leq \sum_{i=1}^{n} \mathbf{H}[X_i]$$

(equal iff $X_1, X_2, .., X_n$ are mutually independent)

6.3 Conditional Entropy

Definition 6.8 (Conditional entropy)

$\mathbf{H}[X|Y = y] = -\sum_{x \in X} \Pr_{X|Y=y}(x) \lg \Pr_{X|Y=y}(x)$

$\mathbf{H}[X|Y] = \sum_{y \in Y} \Pr_Y(y) \mathbf{H}[X|Y = y]$

$= -\sum_{y \in Y} \sum_{x \in X} \Pr_Y(y) \Pr_{X|Y=y}(x) \lg \Pr_{X|Y=y}(x)$

Theorem 6.9 $\mathbf{H}[X|Y] = 0$ *exactly if* $X = f(Y)$ *for some function f.*

Theorem 6.10 $\mathbf{H}[X|Y] = \mathbf{H}[X]$ *exactly if* X *and* Y *are independent.*

Theorem 6.11

$$\mathbf{H}[X, Y] = \mathbf{H}[Y] + \mathbf{H}[X|Y]$$

and thus, $\mathbf{H}[X|Y] \leq \mathbf{H}[X]$ *(using Theorem 6.6), with equality iff* X, Y *are indepenent.*

6.4 Mutual Information

Definition 6.12

$$\mathbf{I}[X; Y] = \mathbf{H}[X] - \mathbf{H}[X|Y]$$

Theorem 6.13

$$\mathbf{I}[X; Y] = \mathbf{I}[Y; X]$$

option vs vote	OUI	NON	und.
separation	30%	60%	10%
offer of partnership	39%	46%	15%
certainty of partnership	54%	35%	11%

6.5 Example:

$$\Pr[OPTION = \text{“separation”}] = 1/4$$
$$\Pr[OPTION = \text{“offer of partnership”}] = 1/2$$
$$\Pr[OPTION = \text{“certainty of partnership”}] = 1/4$$

$$\Pr[VOTE = \text{“OUI”}] = 40.5\%$$
$$\Pr[VOTE = \text{“NON”}] = 46.75\%$$
$$\Pr[VOTE = \text{“undec.”}] = 12.75\%$$

$$\mathbf{H}[VOTE] = -.405 \lg .405 - .4675 \lg .4675 - .1275 \lg .1275 = 1.4198$$
$$\mathbf{H}[OPTION] = -.25 \lg .25 - .5 \lg .5 - .25 \lg .25 = 1.5$$

$$\mathbf{H}[VOTE|OPTION = \text{“separation”}]$$
$$= -.3 \lg .3 - .6 \lg .6 - .1 \lg .1 = 1.2955$$
$$\mathbf{H}[VOTE|OPTION = \text{“offer of partnership”}]$$
$$= -.39 \lg .39 - .46 \lg .46 - .15 \lg .15 = 1.4557$$
$$\mathbf{H}[VOTE|OPTION = \text{“certainty of partnership”}]$$
$$= -.54 \lg .54 - .35 \lg .35 - .11 \lg .11 = 1.3604$$

$$\mathbf{H}[VOTE|OPTION] =$$
$$\frac{\mathbf{H}[VOTE|OPTION = \text{“separation”}]}{4} +$$
$$\frac{\mathbf{H}[VOTE|OPTION = \text{“offer of partnership”}]}{2} +$$
$$\frac{\mathbf{H}[VOTE|OPTION = \text{“certainty of partnership”}]}{4} =$$
$$\frac{1.2955}{4} + \frac{1.4557}{2} + \frac{1.3604}{4} = 1.3918$$

$$\mathbf{I}[VOTE; OPTION] = \mathbf{H}[VOTE] - \mathbf{H}[VOTE|OPTION]$$
$$= 1.4198 - 1.3918 = 0.0280$$

6.6 Jensens Lemma

The following theorem will be used to prove some more interesting statements about entropy functions. First, a preliminary definition.

Definition 6.14 *A real function f is said to be concave on an interval I if* $\forall_{x,y \in I}$

$$f\left(\frac{x+y}{2}\right) \geq \frac{f(x)+f(y)}{2}$$

Theorem 6.15 (Jensens inequality) *If f is continuous and strictly concave on I, and $\sum_{i=1}^{n} a_i = 1$, $a_i \geq 0$, $1 \leq i \leq n$, then, $\forall_{x_i \in I}$*

$$\sum_{i=1}^{n} a_i f(x_i) \leq f(\sum_{i=1}^{n} a_i x_i)$$

with equality iff $x_1 = x_2 = ... = x_n$

6.6.1 Application

Theorem 6.16 *if X takes values x_1 with probability p_1, x_2 with probability p_2, ..., x_n with probability p_n, then*

$$\mathbf{H}[X] \leq \lg\ n$$

(with equality when $p_1 = p_2 = ...p_n = 1/n$)

Proof.

$$\begin{aligned}
\mathbf{H}[X] &= -\sum_{i=1}^{n} p_i \lg p_i \\
&= \sum_{i=1}^{n} p_i \lg 1/p_i \\
&\leq \lg \sum_{i=1}^{n} p_i 1/p_i \\
&= \lg n \ \ (eq.\ when\ p_1 = p_2 = ... = p_n = 1/n)
\end{aligned}$$

6.7 Perfect Secrecy, Key Equivocation, Unicity Distance

Definition 6.17 (Perfect Secrecy) *Let P, K, C be random variables describing the uncertainty about the plaintext, ciphertext and key of a cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$. We say that a cryptosystem has perfect secrecy if* $\forall_{p\in\mathcal{P}, c\in\mathcal{C}} \Pr_{P,C}(p|c) = \Pr_P(p)$

Example 6.3 *Vernam's one-time pad has perfect secrecy. Here, $\mathcal{P} = \mathcal{K} = \mathcal{C} = \{0,1\}^n$, with K the uniform distribution on $\mathcal{K}$ (no distribution condition necessary on $\mathcal{P}$). Encryption and decryption are given by $e_k(p) = p \oplus k$, $d_k(c) = c \oplus k$.*

$$\begin{aligned} \Pr_{P,C}(p|c) &= \Pr_{P,C}(p|p \oplus k) \\ &= \Pr_P(p) \Pr_{P,C}(p \oplus k|p) / \Pr_C(p \oplus k) \ (Bayes) \end{aligned}$$

for any fixed p, $p \oplus K$ is the uniform distribution, so $\Pr_{P,C}(p \oplus k|p) = 1/2^n = \Pr_C(p \oplus k)$ and so

$$\begin{aligned} \Pr_{P,C}(p|c) &= \Pr_P(p) \frac{1}{2^n} / \frac{1}{2^n} \\ &= \Pr_P(p) \end{aligned}$$

An equivalent definition is the following

Definition 6.18 *Cryptosystem $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ has perfect secrecy iff*

$$\mathbf{H}[P|C] = \mathbf{H}[P]$$

Theorem 6.19

$$\mathbf{H}[C] \geq \mathbf{H}[P]$$

when K and P are independent.

Proof.

$$\begin{aligned} \mathbf{H}[C] &\geq \mathbf{H}[C|K] \\ &= \mathbf{H}[C, K] - \mathbf{H}[K] \\ &= \mathbf{H}[P, K] - \mathbf{H}[K] \ \text{(encryption function is bijective)} \\ &= \mathbf{H}[P] + \mathbf{H}[K|P] - \mathbf{H}[K] \ \text{(Thm 6.11)} \\ &= \mathbf{H}[P] + \mathbf{H}[K] - \mathbf{H}[K] \ \text{(we suppose } P, K \text{ independent)} \\ &= \mathbf{H}[P] \end{aligned}$$

We could also show the following

Theorem 6.20 *(when we know the key...)*

$$\mathbf{H}[C|K] = \mathbf{H}[P|K]$$

Theorem 6.21

$$\mathbf{H}[P|C] \leq \mathbf{H}[K]$$

Proof.

$$\begin{aligned} \mathbf{H}[P|C] &\leq \mathbf{H}[P,K|C] \\ &= \underbrace{\mathbf{H}[P|K,C]}_{0} + \mathbf{H}[K|C] \text{ (Bayes)} \\ &= \mathbf{H}[K|C] \\ &\leq \mathbf{H}[K] \end{aligned}$$

This shows that if there is not much doubt about the key, there can't be much doubt about the plaintext knowing the ciphertext.

Definition 6.22 (key equivocation) $\mathbf{H}[K|C]$ *is called the key equivocation, it is a measure of how much we know about the key, knowing the ciphertext.*

Theorem 6.23

$$\mathbf{H}[K|C] = \mathbf{H}[K] + \mathbf{H}[P] - \mathbf{H}[C]$$

Proof.

$$\begin{aligned} (1)\ \mathbf{H}[K,P,C] &= \underbrace{\mathbf{H}[C|K,P]}_{0} + \mathbf{H}[K,P] \\ &= \mathbf{H}[K,P] \\ &= \mathbf{H}[K] + \mathbf{H}[P] \text{ (by independence)} \\ (2)\ \mathbf{H}[K,P,C] &= \mathbf{H}[K,C] \text{ (because from } K \text{ and } C\text{, we can induce } P) \end{aligned}$$

we conclude

$$\begin{aligned}\mathbf{H}[K|C] &= \mathbf{H}[K,C] - \mathbf{H}[C] \\ &= \mathbf{H}[K,P,C] - \mathbf{H}[C] \\ &= \mathbf{H}[K] + \mathbf{H}[P] - \mathbf{H}[C]\end{aligned}$$

Let $P^n = (P_1, \ldots, P_n)$ be a distribution on $n-grams$ of $\mathcal{P}$, and $C^n = (C_1, \ldots, C_n)$ be a distribution on $n-grams$ of $\mathcal{C}$, Suppose that we are working in a cryptographic system $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ where a plaintext $x_1 x_2 \ldots x_n$ is encrypted into a ciphertext $y_1 y_2 \ldots y_n$ by some permutation function. We want to know how much ciphered letters we have to see to know what the key is. The distribution depends on the language in use.

Definition 6.24 *Rate of a Language L*

$$H_L = lim_{n \longrightarrow \infty} \frac{\mathbf{H}[P^n]}{n}$$

For english we have $1 \leq H_L \leq 1.5$

Definition 6.25 *Redondancy of language L*

$$R_L = 1 - \frac{H_L}{\lg\ |\mathcal{P}|}$$

If we take $1,25$ as an estimation of H_L in english, we have a redondancy of 0.75.

For a $c \in \mathcal{C}$, it is possible that many keys can be found such that there exists plaintexts that encrypt to c.

Definition 6.26 (Spurious Keys) *Given some $c \in C^n$, we define*

$$K(c) = \{k \in \mathcal{K} | \exists_x \in \mathcal{P}^n, \Pr_{P^n}(x) > 0 \text{ and } c = e_k(x)\}$$

Finally we can define the notion of unicity distance, which is a measure of the amount of ciphered letters we have to see to know, without doubt, the unique key that was used.

Definition 6.27 *(Unicity Distance)*

$$\overline{S_n} = \sum_{c \in \mathcal{C}^n} \Pr(c)(|K(c)| - 1)$$

We are interested in knowing for what value of n will $\overline{S_n} = 0$.

Theorem 6.28 *([?]) For large n,*

$$\overline{S_n} \geq \frac{|\mathcal{K}|}{|\mathcal{P}|^{nR_L}} - 1$$

If we take $\overline{S_n} = 0$ and solve for n we get

$$n \geq \frac{\lg |\mathcal{K}|}{R_L \ \lg |\mathcal{P}|}$$

Example 6.4

- *Shift Cipher*
  $|\mathcal{P}| = 26, |\mathcal{K}| = 26, R_L = 0.75,\ n \geq 1/R_L = 4/3$
- *Substitution Cipher*
  $|\mathcal{K}| = 26!,\ \lg 26! \approx 88.4,\ n \geq \frac{88.4}{0.75\ 4.7} \approx 25$
- *One-Time Pad*
  $|\mathcal{K}| = 26^n,\ \lg 26^n = n\ \lg 26,\ n \geq \frac{n\ \lg 26}{0.75\ \lg 26} = n/0.75$ *($n = 0$ is the only solution!).*