Théorie de l'information

D. Kohel

01/02/2021

Résumé (Conditions équivalentes pour un cryptosystème

$$\mathcal{E} : \mathcal{K} \times \mathcal{M} \longrightarrow \mathcal{C}$$

d'être parfaitement sûr.

1) $p(M|C) = p(M)$ pour tout $M \in \mathcal{M}$ et $C \in \mathrm{Supp}(\mathcal{C})$.
2) $p(C|M) = p(C)$ pour tout $M \in \mathrm{Supp}(\mathcal{M})$ et $C \in \mathcal{C}$.
3)A $H(\mathcal{M}|\mathcal{C}) = H(\mathcal{M})$.
3)B $H(\mathcal{C}|\mathcal{M}) = H(\mathcal{C})$
4) $H(\mathcal{M},\mathcal{C}) = H(\mathcal{M}) + H(\mathcal{C})$.
5) $I(\mathcal{M},\mathcal{C}) = 0$.

Définition. On appelle $H(\mathcal{K}|\mathcal{C})$ l'équivocation de la clé et $H(\mathcal{M}|\mathcal{C})$ l'équivocation du message.

Remarque. On note que le mot reçu, c.à.d. le texte chiffré, est supposer connu par tout le monde, car on l'envoi par un canal non securisé. Une attaque cryptographique vise

à déterminer soit le message clair, soit la clé. Les quantités $H(K|C)$ et $H(M|C)$ mesure le nombre de bits (car on normalise H par l'utilisation de $\log_2$) inconnus.

Rappel. $H(X|Y)$

$$= \sum_{x,y} p(x,y) \log_2(p(x|y)^{-1})$$

$$= \sum_{y} p(y) \sum_{x} p(x|y) \log_2(p(x|y)^{-1})$$

$$= \sum_{y} p(y) H(X|y) = E(H(X|y)),$$

où $p(x,y) = p(y)\, p(x|y)$.

Il suffit de connaître $p(x,y)$ ou $(p(y), p(x,y))$ pour tout $x \in X$, $y \in Y$

Propriétés de l'équivocation

Lemme. Soit $\mathcal{E}: \mathcal{K} \times \mathcal{M} \to \mathcal{C}$ donné, avec espaces de proba sur $\mathcal{K}$ et $\mathcal{M}$ donnés, $\mathcal{C}$ induite.

$$H(\mathcal{K},\mathcal{M},\mathcal{C}) = H(\mathcal{K},\mathcal{C}) = H(\mathcal{K},\mathcal{M}) = H(\mathcal{K}) + H(\mathcal{M})$$

↑ hypothèse de l'indépendance de la clé et message.

Preuve. Comme (K,M) détermine $C = \mathcal{E}(K,M) = E_K(M)$ et (K,C) détermine $M = D_K(C)$ on a $H(\mathcal{K},\mathcal{C}) = H(\mathcal{K},\mathcal{M},\mathcal{C})$

L'indépendance $= H(\mathcal{K},\mathcal{M})$.

de $\mathcal{K}$ et $\mathcal{M}$ implique: $= H(\mathcal{K}) + H(\mathcal{M})$. □

Exercice. Interpréter

$$p(C|(K,M)) = \begin{cases} 1 \text{ si } E_K(M) = C \\ 0 \text{ sinon,} \end{cases}$$

et

$$p(M|(K,C)) = \begin{cases} 1 \text{ si } D_K(C) = M \\ 0 \text{ sinon,} \end{cases}$$

en $H(\mathcal{K},\mathcal{M},\mathcal{C}) = H(\mathcal{K},\mathcal{M}) + H(\mathcal{C}|(\mathcal{K},\mathcal{M}))$ $= 0$

pour élaborer la preuve.

Lemme. $H(\mathcal{K}|\mathcal{C}) = H(\mathcal{K}) + H(\mathcal{M}) - H(\mathcal{C})$.

Preuve. $H(\mathcal{K},\mathcal{C}) = H(\mathcal{K}|\mathcal{C}) + H(\mathcal{C})$,
donc par le lemme précédent :

$$H(\mathcal{K}|\mathcal{C}) = H(\mathcal{K},\mathcal{C}) - H(\mathcal{C})$$
$$= H(\mathcal{K},\mathcal{M}) - H(\mathcal{C}) = H(\mathcal{K}) + H(\mathcal{M}) - H(\mathcal{C}). \quad \square$$

Proposition.

$$H(\mathcal{M}|\mathcal{C}) = H(\mathcal{K}) + H(\mathcal{M}) - H(\mathcal{C}) - H(\mathcal{K}|(\mathcal{M},\mathcal{C}))$$
$$= H(\mathcal{K}|\mathcal{C}) - H(\mathcal{K}|(\mathcal{M},\mathcal{C})).$$

Preuve. (i) $H(\mathcal{M}|\mathcal{C}) = H(\mathcal{M},\mathcal{C}) - H(\mathcal{C})$,
(ii) $H(\mathcal{K},\mathcal{M},\mathcal{C}) = H(\mathcal{M},\mathcal{C}) + H(\mathcal{K}|(\mathcal{M},\mathcal{C}))$, et
(iii) $H(\mathcal{K},\mathcal{M},\mathcal{C}) = H(\mathcal{K}) + H(\mathcal{M})$.

Alors $H(\mathcal{M}|\mathcal{C}) = H(\mathcal{K},\mathcal{C}) - H(\mathcal{C}) - H(\mathcal{K}|(\mathcal{M},\mathcal{C}))$
$$= H(\mathcal{K}) + H(\mathcal{M}) - H(\mathcal{C}) - H(\mathcal{K}|(\mathcal{M},\mathcal{C}))$$
$$= H(\mathcal{K}|\mathcal{C}) - H(\mathcal{K}|(\mathcal{M},\mathcal{C})). \quad \square$$

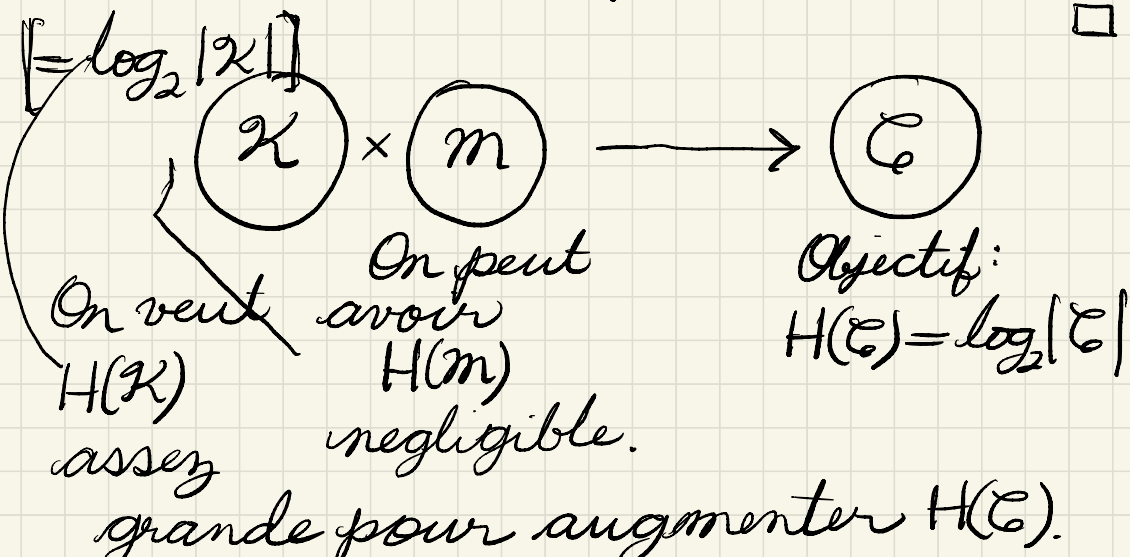
Cor. $H(\mathcal{M}|\mathcal{C}) \leq H(\mathcal{K}|\mathcal{C})$.

Remarque. Ce corollaire encode

le fait qu'une attaque cryptanalytique qui vise à déterminer la clé, déterminera aussi le message. Il peut exister une attaque qui détermine M, étant donné $C = E_K(M)$, sans déterminer K.

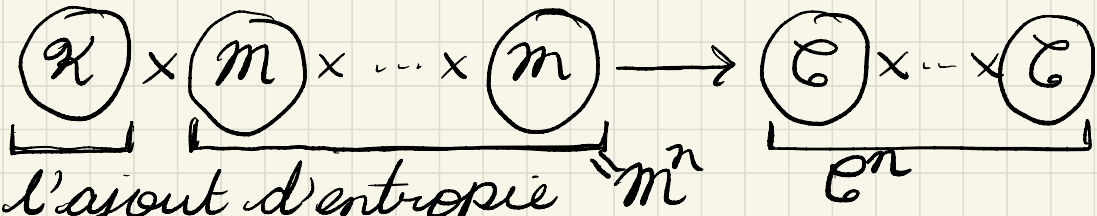
<u>Théorème</u> Si un cryptosystème est parfaitement sûr alors $H(\mathcal{K}) \geq H(\mathcal{M})$.

<u>Preuve</u> $H(\mathcal{K}) \geq H(\mathcal{K}|\mathcal{C}) \geq H(\mathcal{M}|\mathcal{C}) = H(\mathcal{M})$. □



Distance d'unicité

Shannon a aussi considéré la question de la quantité de l'information nécessaire au cryptanalyste pour retrouver la clé à partir de plusieurs textes chiffrés (avec la même clé).



l'ajout d'entropie partagé parmi plusieurs message

$H(M^n) = n\,H(M)$, si les messages sont indépendents

$\geq H(K)$ pour n suff. grand

Ici, pour $K \in \mathcal{K}$ fixé on obtient un chiffrement $\mathcal{M}^n \longrightarrow \mathcal{C}^n$ donné par $(M_1, \ldots, M_n) \longmapsto (C_1, \ldots, C_n)$ avec $C_i = E_K(M_i)$.

Défn. On appelle la distance d'unicité le plus petit entier n tel que $H(\mathcal{K}|\mathcal{C}^n) = 0$.

Remarque. Il s'agit du plus petit n tel que la connaissance de n textes chiffrés $(C_1, \dots, C_n)$, $C_i = E_K(M_i)$ ne laisse plus aucune incertitude résiduelle sur la clé.

On a :

$$\begin{aligned} H(\mathcal{K}|\mathcal{C}^n) &= H(\mathcal{K}, \mathcal{C}^n) - H(\mathcal{C}^n) \\ &= H(\mathcal{K}, \mathcal{M}^n) - H(\mathcal{C}^n) \\ &= H(\mathcal{K}) + H(\mathcal{M}^n) - H(\mathcal{C}^n), \end{aligned}$$

où on note

$$H(\mathcal{C}^n) = H(\mathcal{C}_1, \dots, \mathcal{C}_n) \text{ et}$$
$$H(\mathcal{M}^n) = H(\mathcal{M}_1, \dots, \mathcal{M}_n),$$

tel que

$$\mathcal{E}_n : \mathcal{K} \times \mathcal{M}_1 \times \cdots \times \mathcal{M}_n \longrightarrow \mathcal{C}_1 \times \cdots \times \mathcal{C}_n.$$

Si on suppose que $C_1, \dots, C_n$ sont indépendents, alors

$$H(\mathcal{C}^n) = n H(\mathcal{C}) \ (= n \log_2 |\mathcal{C}|)$$

et si la limite si $\mathcal{C}$ uniforme

$$H = \lim_{n \to \infty} \frac{1}{n} H(M_1, \dots, M_n) \text{ existe}$$

et que la suite $\left(\frac{1}{n} H(M_1, \dots, M_n)\right)$ se stabilise rapidement, on donc $d = n$ minimum tel que

$$H(K) + H(M^n) - H(\mathcal{C}^n) = 0.$$

Avec les hypothèses ci-dessus on trouve

$$H(K) + dH - d \log_2 |\mathcal{C}| = 0$$

et donc

$$d = \frac{H(K)}{\log_2 |\mathcal{C}| - H}.$$

Exemple $\quad M \quad C$

Soit $\mathcal{E} : \mathcal{K} \times A^n \longrightarrow A^n$ un chiffrement par substitution, $A = \{A, \dots, Z\}$ et $M_1, \dots, M_n$ les caractères successifs de la langue française (ou anglaise, allemande, etc.). Le calcul des fréquences des textes clairs permet d'estimer l'entropie par lettre, $H \approx 1{,}8$ bits. On note que $H < \log_2(26) \approx 4{,}7$ bits

Avec $\mathcal{K}$ = groupe symétrique $\mathfrak{S}_{26}$ (bijections $A \longrightarrow A$),

on a $$H(\mathcal{K}) = \log_2(26!)$$

et (C aléatoire)

$$H(C) = \log_2(26) \approx 4{,}7 \text{ bits.}$$

On en déduit

$$d \approx \frac{\log_2(26!)}{\log_2(26) - H} \approx 30.$$

Cela veut dire qu'on doit pouvoir recupérer la clé dès le texte chiffré dépasse une trentaine de caractères.

Remarque On doit conclure que, si on chiffre des messages avec assez de rodondance en utilisant une clé fixée, le cryptanalyste aura à sa disposition assez d'information pour retrouver la clé.

Attention. Cela ne prend pas en consideration le coût de calculer la clé.

Par exemple, en cryptographie à clé publique, il n'y a pas de securité theorique ($I(M,C)=$

$H(m) = H(c)$ ou $H(m|c) = 0$ ou $H(m,c) = H(m)$, etc). La sécurité se repose sur la difficulté computationnelle à calculer la clé privée ou le message (qui est uniquement déterminé par le texte chiffré et la clé publique).

En effet tout le monde peut construire des couples

$$(M, C = E_K(M))$$

et la clé privée est également uniquement déterminé par la clé publique.