Cryptography 26.04.2019

3.4. Vigenère cipher with running key

To avoid periodicity, use a key of the same length as the plaintext.

$$\begin{array}{rcccc} & a_1 & a_2 & \cdots & a_n \\ \oplus & \delta_1 & \delta_2 & & \delta_n \\ \hline & c_1 & c_2 & \cdots & c_n \end{array}$$

Frequency attack is possible if $(\delta_1 \cdots \delta_n)$ is from a natural language.

Model: M_i : random variable → occurence of plaintext characters.
K_i : r.v.s → occurence of key characters.
} Stoch. ind.

The most frequent characters are: $\underbrace{E, T, A, O, I, N, S}_{57\%}$

$$\mathbb{P}(M_i \in \{E, \ldots, S\}) \simeq 0.57 \simeq \mathbb{P}(K_i \in \{E, \ldots, S\})$$

$$\mathbb{P}((M_i, K_i) \in \{E, \ldots, S\} \times \{E, \ldots, S\}) \simeq 0.57^2 = \underline{\underline{0.324}}$$

⇒ Almost $1/3$ of all ciphertext characters are obtained by adding 2 of the most frequent characters.

Defense against this attack: random key stream. However, never use a key twice.

Otherwise:

$$(a_1, \ldots, a_\ell) \oplus (K_1, \ldots, K_\ell) = (c_1, \ldots, c_\ell)$$

$$(b_1, \ldots, b_\ell) \oplus (K_1, \ldots, K_\ell) = (d_1, \ldots, d_\ell)$$

Both are known to "Oscar"

$$(c_i - d_i) \bmod 26 = (a_i - b_i) \bmod 26.$$

it is possible to use the above attack.

4. Entropy and Perfect secrecy

4.1. Entropy

Consider random experiments, e.g., (0.9, 0.05, 0.05).

We aim at a measure of

$$= \begin{cases} \text{\underline{uncertainty} about the outcome (before)} \\ \\ \text{\underline{information} gained by the outcome (after)} \end{cases}$$

The measure was introduced by Shannon. (49)

Formal description:

X: discrete random variable with

finite support $\mathcal{X} = \{x_1, \dots x_m\}$

$$P(X = x_i) = p_i \ , \ i = 1, \dots, m$$

information of $X = x_i$: Positive   0 if $p_i = 1$ $+\infty$ if $p_i = 0$

candidate 1. $\dfrac{-1}{p_i - 1}$

candidate 2: $-\log p_i$

Def 4.1. Let $c > 1$ be an arbitrary constant.

$$H(X) = -\sum_{i=1}^{m} p_i \log_c p_i = -\sum_i P(X = x_i) \log_c P(X = x_i)$$

* Convention: $0 \cdot \log 0 = 0$; Omit (but fix) constant c.

Analogous definition for 2-dimensional random variables

$$(X, Y) : \quad \mathcal{X} \times \mathcal{Y} = \{x_1, \dots, x_m\} \times \{y_1, \dots, y_d\}$$

$$P(X = x_i, Y = y_j) = p_{ij}$$

Def 4.2.

a) $H(X,Y) = -\sum P(X=x_i, Y=y_j) \log P(X=x_i, Y=y_j)$

$$= -\sum_{j,i=1}^{d,m} p_{ij} \log p_{ij}$$

is called joint entropy of X, Y

b) $H(X|Y) = -\sum_{j=1}^{d} P(Y=y_j) \sum_{i=1}^{m} P(X=x_i | Y=y_j) \log P(X=x_i | Y=y_j)$

$$= -\sum_{j=1}^{d} P(Y=y_j) \sum_{i=1}^{m} P(X=x_i | Y=y_j) \log P(X=x_i | Y=y_j)$$

$$= -\sum_{i,j} P(X=x_i, Y=y_j) \log (P(X=x_i | Y=y_j)).$$

is called conditional entropy of X given Y. (equivocation)

Theorem 4.3. a) $0 \underset{(i)}{\leq} H(X) \underset{(ii)}{\leq} \log m$

(i) $\Longleftrightarrow \exists x_i : P(X=x_i) = 1$. Singleton dist.

(ii) $\Longleftrightarrow$ X is uniformly distributed. $P(X=x_i) = 1/m \quad \forall i$

b) $\underset{(i)}{0 \leq} H(X|Y) \underset{(ii)}{\leq} H(X)$

(i) $\iff P(X=x_i|Y=y_j)=1 \; \forall i,j$ with $P(X=x_i, Y=y_j)>0$

(ii) $\iff X, Y$ stoch. independent.

c) $H(X) \underset{(i)}{\leq} H(X,Y) \underset{(ii)}{\leq} H(X)+H(Y)$

(i) $\iff P(Y=y_j|X=x_i)=1 \; \forall i,j$ $P(X=x_i, Y=y_j)>0$.

(ii) $\iff X, Y$ stoch. indep.

d) (chain rule)

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

Another interpretation of $H(X)$:

* Shannon proved that $H(X)$ is a lower bound for the average codeword length of any uniquely decodable code.

- $\log m$ is the worst case average length
- $R = 1 - \frac{H(X)}{\log m}$ is called redundancy of a source

4.2 Perfect secrecy..

Cryptosystem $(\mathcal{M}, \mathcal{K}, \mathcal{C}, e, d)$ with

$\mathcal{M} = \{M_1, \ldots, M_m\}$ message space

$\mathcal{K} = \{K_1, \ldots, K_k\}$ key space

$\mathcal{C} = \{C_1, \ldots, C_n\}$ ciphertext space

$\hat{M}, \hat{K}$ indendent r.v. with support $\mathcal{M}$ and $\mathcal{K}$ resp. :

$$P(\hat{M} = M_i) = P_i$$

$$P(\hat{K} = K_j) = q_j$$

model the occurence of messages and keys.

Encryption: $e(\hat{M}, \hat{K}) = \hat{C}$

$$P(\hat{C} = C_l) = r_l = \sum_{(i,j):\, e(M_i, K_j) = C_l} P_i q_j$$

$H(\hat{M}) = -\sum p_i \log p_i$

$H(\hat{K}) = -\sum q_j \log q_j$

$H(\hat{C}) = -\sum r_l \log r_l$

$H(\hat{K} \mid \hat{C})$: Key equivocation

$H(\hat{M} \mid \hat{C})$: message equivocation

Def. 4.9. A cryptosystem $(\mathcal{M}, \mathcal{K}, \mathcal{C}, e, d)$, is said to have perfect secrecy if

$$H(\hat{M} \mid \hat{C}) = H(\hat{M}),$$

Interpretation : The knowledge of cryptogram $\hat{C}$ does not decrease uncertainty about M.

Corollary 4.11. A cryptosystem has perfect secrecy $\Longleftrightarrow$ $\hat{M}$ and $\hat{C}$ are stoch. independent

$$\Longleftrightarrow P(\hat{M} = M_i \mid \hat{C} = C_l) = P(\hat{M} = M_i)$$
$$\forall l : P(\hat{C} = C_l) > 0$$

$$\Longleftrightarrow P(\hat{C} = C_l \mid \hat{M} = M_i) = P(\hat{C} = C_l)$$
$$\forall i\ P(\hat{M} = M_i) > 0$$

The above conditions are all tedious to check.

Easy sufficient conditions ↓

Theorem 4.14. $(\mathcal{M}, \mathcal{K}, \mathcal{C}, e, d)$ has perfect secrecy <u>if</u>

(i) $P(\hat{K}=K) = \frac{1}{|K|}$ for all $K \in \mathcal{K}$

(ii) for all $M \in \mathcal{M}$ and $C \in \mathcal{C}$ there is a unique $K \in \mathcal{K}$ s.t. $e(M,K)=C$,

proof.

$$P(\hat{C}=C \mid \hat{M}=M) = \frac{P(\hat{C}=C, \hat{M}=M)}{P(\hat{M}=M)}$$

$$= \frac{P(e(M,\hat{K})=C, \hat{M}=M)}{P(\hat{M}=M)}$$

$$= P(e(M,\hat{K})=C) = \frac{1}{|\mathcal{K}|}$$

$$P(\hat{C}=C) = \sum_{M \in \mathcal{M}} P(\hat{C}=C \mid \hat{M}=M) P(\hat{M}=M)$$

$$= \frac{1}{|K|} \sum_{M \in \mathcal{M}} P(\hat{M}=M) = \frac{1}{|K|}$$

$\Rightarrow \hat{C}, \hat{M}$ are stoch. ind. $\Rightarrow$ perfect secrecy $\square$

Veram ciphers have perfect secrecy.

$\mathcal{X} = \{0, \dots, m-1\}$ $\mathcal{M}_N = \mathcal{C}_N = K_N = \mathcal{X}^N$

$e(M, K) = ((a_1+s_1) \bmod m, \dots, (a_N+s_N) \bmod m)$

$M = (a_1, \dots, a_N)$ $K = (s_1, \dots, s_N)$

$P(\hat{K} = i) = \frac{1}{m}$ $i = 0, \dots, m-1$

Theorem: The vernam cipher has perfect secrecy

Proof.

(i) $P(\hat{K}_N = K) = \prod_{i=1}^{N} P(\hat{K}_i = s_i) = \frac{1}{m^N}$

uniformly dist.

(ii) $\forall M \in M_N$, $C \in \mathcal{C}_N : \exists! K \in K_N$

$e(M, C) = K$

$K = (s_1, \dots, s_N)$: $s_j = (c_j - a_j) \bmod m$.