3.4. Vigenère cipher with running key

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

\[ a_1, a_2, \ldots, a_n \]
\[ \oplus s_1, s_2, \ldots, s_n \]
\[ c_1, c_2, \ldots, c_n \]

Frequency attack is possible if \((s_1, \ldots, s_n)\) is from a natural language.

Model: \( M_i \): random variable \( \rightarrow \) occurrence of plaintext characters.
\( K_i \): r.v. \( \rightarrow \) occurrence of key characters.

The most frequent characters are: \( E, S, I, O, N, S \) \( \underline{57\%} \).

\[ P(M_i \in \{E, S, O, N, S\}) = 0.57 \Rightarrow P(K_i \in \{E, S, O, N, S\}) \]
\[ P(M_i, K_i \in \{E, S, O, N, S\} \land M_i, K_i \in \{E, S, O, N, S\}) = 0.57^2 \approx 0.324 \]

\( \Rightarrow \) Almost \( \frac{1}{3} \) of all ciphertext characters are obtained by adding 2 of the most frequent characters.
Defense against this attack is random key stream. However, never use a key twice.

Otherwise,

\[(a_1, \ldots, a_q) \oplus (K_1, \ldots, K_q) = (c_1, \ldots, c_q)\]

\[(b_1, \ldots, b_q) \oplus (K_1, \ldots, K_q) = (d_1, \ldots, d_q)\]

Both are known to "Oscar"

\[(c_i - d_i) \mod 26 = (a_i - b_i) \mod 26.\]

It is possible to use the above attack.

4. Entropy and Perfect secrecy

4.1. Entropy

Consider random experiments, e.g., \((0.9, 0.95, 0.05)\).

We aim at a measure of uncertainty about the outcome (before)

\[
\text{information gained by the outcome (after)}
\]
The measure was introduced by Shannon. (49)

**Formal description:**

$X$: discrete random variable with finite support $X = \{x_1, \ldots, x_m\}$

$P(X = x_i) = P_i$, $i = 1, \ldots, m$

informaion: Positive $0$ if $P_i = 1$; $\infty$ if $P_i = 0$

of $X = x_i$:

- candidate 1. $\frac{-1}{\log P_i}$
- 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: $0 \cdot \log 0 = 0$; $0$mit (but fix) constant $c$

Analogous definition for 2-dimensional random variables

$(X, Y) : X \times Y = \{x_1, \ldots, x_m \times Y_1, \ldots, Y_d\}$

$P(X = x_i, Y = y_j) = P_{ij}$
Def 4.2.

a) \( H(X, Y) = -\sum_{x \in X, y \in Y} P(x = x_i, y = y_j) \log P(x = x_i, y = y_j) \)

\( \sum_{x, y} \)

\( = -\sum_{y \in Y} P(y = y_j) \sum_{x \in X} P(x = x_i | y = y_j) \log P(x = x_i | y = y_j) \)

\( \sum_{y \in Y} \sum_{x \in X} \)

\( = -\sum_{y \in Y} P(y = y_j) \sum_{x \in X} P(x = x_i | y = y_j) \log P(x = x_i | y = y_j) \)

is called joint entropy of \( X, Y \).

b) \( H(X | Y) = \frac{1}{d} \sum_{y \in Y} \sum_{i=1}^{m} P(y = y_j) \log P(x_i | y = y_j) \)

\( \frac{1}{d} \sum_{y \in Y} \sum_{i=1}^{m} \)

\( = -\sum_{i=1}^{m} P(x = x_i) \log P(x = x_i | y = y_j) \)

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

(Cequivocation)

Theorem 4.3. a) \( 0 \leq H(X) \leq \log m \)

(i) \( \leq \)

(ii) \( \geq \)

(c) \( \Leftrightarrow \exists x_i : P(x = x_i) = 1 \). Singleton dist.

(ii) \( \Leftrightarrow \) \( X \) is uniformly distributed \( \forall i \)
b) \( 0 \leq H(X|Y) \leq H(X) \)

(i) \[ \iff \mathbb{P}(X=x_i | Y=y_j) = 1 \quad \forall i,j \text{ with } \mathbb{P}(X=x_i, Y=y_j) > 0 \]

(ii) \[ \iff X, Y \text{ stoch. independent.} \]

c) \( H(X) \leq H(X,Y) \leq H(X) + H(Y) \)

(i) \[ \iff \mathbb{P}(Y = y_j | X = x_i) = 1 \quad \forall i,j \mathbb{P}(X=x_i, Y=y_j) > 0 \]

(ii) \[ \iff X, Y \text{ stoch. independent.} \]

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.
\[ \text{log } m \text{ is the worst case average length} \]

\[ R = 1 - \frac{H(x)}{\text{log } m} \text{ is called redundancy of a source} \]

4.2 Perfect secrecy

Cryptography system \((M, K, C, E, D)\) with

\[ M = \{ M_1, \ldots, M_m \} \text{ message space} \]

\[ K = \{ K_1, \ldots, K_k \} \text{ key space} \]

\[ E = \{ C_1, \ldots, C_n \} \text{ ciphertext space} \]

\(M, K\) independent r.v. with support \(M\) and \(K\) respectively \(P(C = M_i) = P_i\)

\[ P(K = K_j) = q_j \]

model the occurrence of messages and keys.

Encryption: \(E(M, K) = C\)

\[ P(C = C_k) = q_k = \sum P_i q_j \]

\(a(j,i) = E(M_i, K_j) = C_k\).
\[ H(M) = - \sum P_i \log P_i \]
\[ H(K) = - \sum q_y \log q_y \]
\[ H(C) = - \sum q_r \log q_r \]
\[ H(K|C) \text{ is key equivocation} \]
\[ H(M|C) \text{ is message equivocation} \]

**Def. 4.9.** A cryptosystem \((M, K, C, e, d)\) is said to have perfect secrecy if
\[ H(M|C) = H(M) \]

**Interpretation:** The knowledge of cryptogram \(C\) does not decrease uncertainty about \(M\).

**Corollary 4.11.** A cryptosystem has perfect secrecy \( \iff \hat{M} \) and \( \hat{C} \) are stoch. independent
\[ \iff P(C = c_i | \hat{M} = \hat{M}_i) = P(C = c_i) \quad \forall i \]
\[ \iff P(\hat{C} = c_i | \hat{M} = \hat{M}_i) = P(\hat{C} = c_i) \quad \forall i \]
\[ \iff P(C = c_i | \hat{M} = \hat{M}_i) = P(C = c_i) \quad \forall i \]
The above conditions are all tedious to check.

Easy sufficient conditions 1

Theorem 4.14. \((M, K, E, e, d)\) has perfect secrecy if

\[\Pr(K = k) = \frac{1}{|K|} \text{ for all } k \in K\]

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

Proof.

\[
\Pr(C = c | M = M) = \frac{\Pr(C = c, M = M)}{\Pr(M = M)} = \frac{\Pr(e(M, K) = c, M = M)}{\Pr(M = M)} = \frac{1}{|K|}
\]

\[
\Pr(C = c) = \sum_{M \in M} \Pr(C = c | M = M) \Pr(M = M) = \frac{1}{|K|} \sum_{M \in M} \Pr(M = M) = \frac{1}{|K|}
\]

\[
\Rightarrow e, M \text{ one-to-one match. } \text{ind. } \Rightarrow \text{Perfect secrecy.}
\]
Vernam ciphers have perfect secrecy.

\[ X = \{0, \ldots, m-1\} \quad M_N = C_N = K_N = X^N \]

\[ E(M, K) = (c_1 + s_1) \mod m, \ldots, (c_n + s_n) \mod m \]

\[ M = (a_1, \ldots, a_n) \quad K = (s_1, \ldots, s_n) \]

\[ \Pr(K = i) = \frac{1}{m} \quad i = 0, \ldots, m-1 \]

Theorem: The vernam cipher has perfect secrecy.

Proof.

\[ \Pr(K_N = K) = \prod_{i=1}^{N} \Pr(K_i = s_i) = \frac{1}{m^N} \]

uniformly dist.

\[ \forall M \in M_N, \quad C \in C_N : D(K \cdot K_N) \]

\[ E(M_0 \cdot C) = K \]

\[ K = (s_1, \ldots, s_n) : \quad 3_j = (c_j - a_j) \mod m. \]