# CS 171: Discussion Section 1 (Jan 22)

## 1. Formal definitions

Provide formal definitions of $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ for the shift, substitution, and Vigenère ciphers. (You can use $\mathbb{Z}_n$ to denote the set of numbers $\{0, 1, \ldots, n-1\}$ and identify the letters of the English alphabet with $\mathbb{Z}_{26}$.)

### Solution

**Shift Cipher.** When $m$ is a single character in $\mathbb{Z}_{26}$:

1. $\mathsf{Gen}$ : Choose a key $k$ randomly from $\mathbb{Z}_{26}$.
2. $\mathsf{Enc}(k, m)$ : Output $c = m + k \mod 26$.
3. $\mathsf{Dec}(k, c)$ : Output $m = c - k \mod 26$.

For multi-character messages and ciphertexts, apply $\mathsf{Enc}(k, \cdot)$ and $\mathsf{Dec}(k, \cdot)$ to each character, using the same $k$ for every index.

**Substitution Cipher.** When $m$ is a single character in $\mathbb{Z}_{26}$:

1. $\mathsf{Gen}$ : Choose a random permutation $f$ of $\mathbb{Z}_{26}$.
2. $\mathsf{Enc}(f, m)$ : Output $c = f(m)$.
3. $\mathsf{Dec}(f, c)$ : Output $m = f^{-1}(c)$.

For multi-character messages and ciphertexts, apply $\mathsf{Enc}(f, \cdot)$ and $\mathsf{Dec}(f, \cdot)$ to each character, using the same $f$ for every index.

**Vigenère cipher.**

1. $\mathsf{Gen}$ : Let $T$ be the maximum length of the period. Choose a random period $t$ from $\{1, \ldots, T\}$. For each $\tau \in \{1, \ldots, t\}$, choose $k_\tau$ randomly from $\mathbb{Z}_{26}$.
2. $\mathsf{Enc}(\{k_\tau\}_{\tau \in [t]}, m)$ : Chop $m$ into blocks of length $t$ i.e., parse $m$ as $m_1 m_2 \ldots m_n$ where each $m_i$ has length $t$. Let $m_{i,\tau}$ denote the $\tau$-th element of the $i$-th message block $m_i$. For $i \in [n]$ and $\tau \in [t]$, compute $c_{i,\tau} = m_{i,\tau} + k_\tau \mod 26$. Output $c = c_{1,1} c_{1,2} \ldots c_{1,t} c_{2,1} \ldots c_{n,t}$.
3. $\mathsf{Dec}(\{k_\tau\}_{\tau \in [t]}, c)$ : Parse $c$ as $c_{1,1} c_{1,2} \ldots c_{1,t} c_{2,1} \ldots c_{n,t}$. For each $i \in [n]$ and $\tau \in [t]$, compute $m_{i,\tau} = c_{i,\tau} - k_\tau \mod 26$. Output $m = m_{1,1} m_{1,2} \ldots m_{1,t} m_{2,1} \ldots m_{n,t}$.

■

## 2. Exploiting Partial Information About the Message

**Setting:** Let's say that a user wants to send one of two values over unsecured channels. For example, these could correspond to `Attack by land` and `Attack by sea`, or `abcd` and `ehgj`, or even something else. They will encrypt their message before sending it using one of the encryption schemes above and hope that an adversary who can see the ciphertext cannot figure out which message they sent.

In more general terms:

1. Assume that an adversary knows that a user's message is one of two values, $m_A$ or $m_B$.
2. Say the user encrypts their message, and the adversary sees the resulting ciphertext.
3. Can the adversary figure out which message, $m_A$ or $m_B$, was encrypted?

The answer depends on the particular values of $m_A$ and $m_B$.

**Questions:**

(a) Say the user encrypts their message with the shift cipher, and say the message is either $m_A =$ `abcd` or $m_B =$ `ehgj`. Show how the adversary can determine the user's message, or show that this is not possible.

(b) Now say the user encrypts their message with the substitution cipher. If the message is either $m_A =$ `abcd` or $m_B =$ `ehgj`, can the adversary learn the message? If so, show how the adversary can do so. If not, find different values for $m_A$ and $m_B$ such that the adversary can learn the message.

(c) Say the user encrypts their message with the Vigenère cipher, and say the message is either $m_A =$ `abcd` or $m_B =$ `ehgj`. Show how the adversary can determine the user's password, or explain why this is not possible. Consider Vigenère ciphers that use period 2, period 3, and period 4.

**Solution**

1. For the case of the shift cipher, if the password was `abcd` then the corresponding ciphertext will be 4 consecutive letters.
2. For the case of subsitution cipher, it is impossible to distinguish between encryptions of `abcd` and `ehgj`. Notice that if $f$ is the chosen permutation then the ciphertext in the first case will be $(f(0), f(1), f(2), f(3))$ and the ciphertext in the second case will be $(f(4), f(7), f(6), f(9))$. If $f$ is a random permutation then the distribution of these two ciphertexts will be the same.

   Alternatively, we can make it possible for the adversary to learn the message by choosing $m_A =$ `abcd` and $m_B =$ `aaaa`. Using the substitution cipher, the encryption of `abcd` is four different characters, whereas the encryption of `aaaa` is a single character repeated four times.
3. For the Vigenère cipher,

(a) If the period was 2, then it is not possible to distinguish since the distance between `a` and `c` (resp. `b` and `d`) and `e` and `g` (resp. `h` and `j`) are the same. Hence, the distribution of the ciphertexts in both cases will be the same. Formally, if $k_1$ and $k_2$ are the chosen keys, then the first ciphertext will be $(k_1 + 0, k_2 + 1, k_1 + 2, k_2 + 3)$ (where $+$ denotes addition modulo 26) and the second ciphertext will be $(k_1 + 4, k_2 + 7, k_1 + 6, k_2 + 9)$. When $k_1$ and $k_2$ are randomly chosen from $\mathbb{Z}_{26}$, these two distributions are the same.

(b) If the period was 3, then it is possible to distinguish since the distance between `a` and `d` is 3 whereas the distance between `e` and `j` is 5.

(c) If the period was 4, it is not possible to distinguish between the two cases. Formally, if $k_1, \ldots, k_4$ are the chosen keys then the first ciphertext will be $(k_1 + 0, k_2 + 1, k_3 + 2, k_4 + 3)$ and the second ciphertext will be $(k_1 + 4, k_2 + 7, k_3 + 6, k_4 + 9)$. When $k_1, \ldots, k_4$ are randomly chosen, the distribution of these two ciphertexts are the same.

■