CS 171: Problem Set 1

Due Date: February 1, 2024 at 8.59pm via Gradescope

1. Substitute and Shift cipher (10 points)

Consider a modification of the substitution cipher, where instead of applying only the substitution, we first apply a substitution and then apply a shift cipher on the substituted values. Give a formal description of this scheme and show how to break the substitute and shift cipher.

Solution

Substitution and Shift Cipher.

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

Attack. Observe that substitute and shift cipher is essentially a substitution cipher with the substitution defined to be $g_k(x) = f(x) + k \mod 26$. Apply the attack (frequency analysis) on the substitution cipher.

2. Double Vigenère cipher (10 points)

Consider the double Vigenère cipher, where we choose two keys k_1, k_2 with different periods $t_1, t_2 \in [t_{max}]$, and then encrypt a message in the following way: first encrypt the message using the Vigenère cipher with k_1 to obtain an intermediate ciphertext c', and then encrypt c' using the Vigenère cipher with k_2 to obtain the final ciphertext c. Show how to break the double Vigenère cipher.

Solution Let $k_1 = (k_{1,1}, \ldots, k_{1,t_1})$ where $k_{1,i}$ are random shifts from $\mathbb{Z}_{26}$. Similarly, let $k_2 = (k_{2,1}, \ldots, k_{2,t_2})$ where $k_{2,i}$ are random shifts from $\mathbb{Z}_{26}$. Observe that Double Vigenère cipher is essentially a single Vigenère cipher with period equal to $lcm(t_1, t_2)$ and for each $i \in [lcm(t_1, t_2)]$, the i-th shift corresponds to $k_{1,i \mod t_1} + k_{2,i \mod t_2} \mod 26$. Now, applying the attack against the single Vigenère cipher will break the double Vigenère.

3. Chosen Plaintext Attacks (10 points)

Assume the adversary has the ability to obtain ciphertexts for arbitrary plaintexts, i.e., it can choose a message and receive an encryption of the message without knowing the secret key. Attacks that leverage this information are known as *chosen plaintext attacks*. Using chosen plaintext attacks, show how to learn the secret key for the shift, substitution, and Vigenère ciphers. Your attacks should each use only a single plaintext. What is the smallest plaintext length that suffices for each attack?

For the Vigenère cipher, please consider two cases: (a) when the period t is known; (b) when the period t is unknown but an upper bound t_{max} on t is known. For the latter case, an asymptotic analysis of the required plaintext length suffices, i.e., an upper bound on the required plaintext length.

Solution

Shift Cipher. Choose an arbitrary message $m \in \mathbb{Z}_{26}$ and request for the ciphertext c corresponding to m. Recover the key k as $c - m \mod 26$. One character is sufficient.

Substitution Cipher. For each $m \in \{0, \ldots, 24\}$, request for the ciphertext encrypting m. Recover the permutation f of $\mathbb{Z}_{26}$ that is consistent with $\{i, c_i\}_{i \in \{0,\ldots,24\}}$ (i.e., $f(i) = c_i$) where c_i is the ciphertext corresponding to i. 25 characters are sufficient.

Vigenère cipher of known period. Let t be the period of the cipher. Choose an arbitrary message block of length t denoted by $m_1 \ldots m_t$ and obtain the ciphertext $c_1 \ldots c_t$ corresponding to this block. Recover $k_\tau = c_\tau - m_\tau$ for each $\tau \in [t]$. t characters are sufficient.

Vigenère cipher of unknown period. Let $t_{\max}$ be the maximum length of the period. Choose a message vector of length $2t_{\max}$ denoted by $m_1 \ldots m_{2t_{\max}}$ and request the ciphertext corresponding to this vector. Let the ciphertext be $c_1 \ldots c_{2t_{max}}$. For each $i \in [2t_{max}]$, recover $k_i = c_i - m_i \mod 26$. Find the minimum $t \in [t_{max}]$ such that $k_1 \ldots k_t$ is a repeating sequence in $(k_1 \ldots k_{2t_{max}})$ and output $k_1 \ldots k_t$ as the key. Notice that if for some $t' > t$, $k_1 \ldots k_{t'}$ is also a repeating sequence of $(k_1 \ldots k_{2t_{max}})$ then it must the case that $k_1 \ldots k_{gcd(t,t')}$ is also a repeating sequence of $(k_1 \ldots k_{2t_{max}})$. However, since t is the minimum period, it follows that $gcd(t, t') = t$. $2t_{\max}$ characters are sufficient.

4. (optional) Exploring Kerckhoff's Principle

In the lectures, we covered Kerckhoff's principle: cryptographic systems should be designed such that the adversary knows everything except the secret key. We will see how this principle manifests in the real world through a small exercise.

`privacytools.io` maintains a crowdsourced list of tools that adhere to high security and privacy standards. In particular, the following link[1] lists file encryption software. Select at least one of these tools and:

(a) Try out the tool. Observe that encrypted data is unintelligible and that the key payload is random bits.

(b) Confirm that the codebase is open source, i.e., all critical code is publicly auditable.

(c) Assert that standardized cryptography are used, i.e., the encryption algorithm itself is publicly known. List the cryptographic protocols used and how they are standardized (e.g., AES standardized by NIST).

[1] `https://www.privacytools.io/secure-file-encryption`

(d) Observe that a lost encryption key means lost data: there is no way to recover encrypted data without the encryption key. This is in contrast to password recovery, which is effectively key recovery because the underlying keys are derived from the password.