HW 1 - Cryptography

James Derrod

February 6, 2025

1 Question 1: Breaking the Improved Vigenère Cipher

The improved Vigenère cipher replaces shift ciphers with multiple substitution ciphers. The key consists of t random permutations of the alphabet, and each letter in positions $i, t+i, 2t+i, \ldots$ is encrypted using the i-th permutation.

Breaking the Cipher

To decrypt this cipher, proceed as follows:

1. **Partition the Ciphertext:** Since each letter in positions $i, t+i, 2t+i, \ldots$ is encrypted with the same substitution cipher, the ciphertext can be divided into t separate groups.
2. **Perform Frequency Analysis:** Each group follows a monoalphabetic substitution cipher, which is vulnerable to letter frequency analysis. By matching ciphertext letter frequencies to English letter distributions, can determine each substitution permutation independently.
3. **Recover the Key:** Once each of the t substitution ciphers is solved, the full key (set of t permutations) is reconstructed.
4. **Decrypt the Message:** With the key known, reversing the substitutions decrypts the full plaintext.

Why This Works

Monoalphabetic substitution ciphers are vulnerable to frequency analysis. Since the improved Vigenère cipher simply applies multiple independent substitutions, it can be broken by analyzing each group separately.

Question 2: Known-Plaintext Attack on Shift, Substitution, and Vigenère Ciphers

A known-plaintext attack (KPA) allows an adversary to learn plaintext-ciphertext pairs and attempt to recover the key. Show that the shift, substitution, and Vigenère ciphers are all vulnerable.

Shift Cipher

A shift cipher encrypts each letter by shifting it by a fixed amount k. Given a single known plaintext-ciphertext pair, the shift k can be determined as:

$$k = (\text{ciphertext letter index}) - (\text{plaintext letter index}) \mod 26.$$

Since the same shift is applied to all letters, knowing one plaintext letter is enough to recover k and decrypt the entire ciphertext.

Known plaintext required: 1 letter.

Substitution Cipher

A monoalphabetic substitution cipher replaces each letter with a unique mapping. If the entire alphabet appears in a known plaintext-ciphertext pair, the full substitution mapping can be directly recovered. If only a partial alphabet appears, frequency analysis can be used to infer missing letters.

Known plaintext required: At least 26 letters (if covering the full alphabet).

Vigenère Cipher

The Vigenère cipher uses a repeating keyword to apply multiple shift ciphers. If the key length t is known, the ciphertext can be split into t separate shift ciphers. Each shift can then be solved using the method for the shift cipher above.

If t is unknown, it can be determined using repeated patterns or the Kasiski examination. Once t is found, breaking the cipher reduces to solving t shift ciphers.

Known plaintext required: At least t letters (one for each shift cipher).

Conclusion

All three ciphers are vulnerable to a known-plaintext attack:

- Shift cipher: broken with **1 known letter**.
- Substitution cipher: broken with **26 known letters** (full alphabet).
- Vigenère cipher: broken with t **known letters**, assuming the period is known.

Each of these ciphers provides minimal security against a known-plaintext attack.

Question 3: Attacking an Encrypted Password

An attacker knows that a user's password is either `abcd` or `bedg` and wants to determine which one was used, given the ciphertext.

(a) Shift Cipher

A shift cipher applies the same shift k to all letters. The attacker can determine k by computing the shift for any letter:

$$k = (\text{ciphertext letter index}) - (\text{plaintext letter index}) \mod 26.$$

By encrypting both `abcd` and `bedg` under all possible shifts and comparing with the ciphertext, the attacker can uniquely determine k and identify the correct password.

Conclusion: The shift cipher is completely insecure, as the attacker can always determine the password.

(b) Vigenère Cipher

The Vigenère cipher encrypts using a repeating key. Analyze different key lengths.

Period 2

With a key length of $t = 2$, the first and second letters are encrypted with different shifts, repeating every two letters. The attacker can test all 26^2 possible two-letter keys to determine which password was used.

Conclusion: The attacker can efficiently brute-force the key.

Period 3

With $t = 3$, the attacker must test 26^3 keys, making brute-force harder. However, if enough plaintext is available, frequency analysis can be used to break each shift separately.

Conclusion: More difficult but still breakable.

Period 4

With $t = 4$, the number of possible keys increases to 26^4. This makes brute-force infeasible, but known-plaintext techniques such as frequency analysis can still be applied.

Conclusion: More secure but not unbreakable.

Conclusion

- **Shift cipher:** Completely insecure; attacker can always determine the password.
- **Vigenère cipher (Period 2):** Easily brute-forced.
- **Vigenère cipher (Period 3):** Harder but still breakable.
- **Vigenère cipher (Period 4):** More resistant but still vulnerable to analysis.

Question 4: Perfect Secrecy - Proofs and Refutations

Analyze two statements about perfect secrecy, proving or refuting them.

(a) Prove or Refute:

$$\Pr[M = m_0 \mid C = c] = \Pr[M = m_1 \mid C = c]$$

for all messages m_0, m_1 in the message space M and all ciphertexts c in the ciphertext space C.

Proof

By the definition of perfect secrecy, observing the ciphertext provides no additional information about the plaintext:

$$\Pr[M = m \mid C = c] = \Pr[M = m].$$

Applying this to two messages m_0, m_1, we get:

$$\Pr[M = m_0 \mid C = c] = \Pr[M = m_0], \quad \Pr[M = m_1 \mid C = c] = \Pr[M = m_1].$$

Since these probabilities are independent of c, they must be equal. Thus, the given statement holds.

Conclusion: The statement is **true** and equivalent to the definition of perfect secrecy.

(b) Prove or Refute:

$$\Pr[C = c_0] = \Pr[C = c_1]$$

for all ciphertexts c_0, c_1 in the ciphertext space C.

Counterexample: One-Time Pad

The one-time pad is a perfectly secret encryption scheme but does not guarantee uniform ciphertext probabilities. If messages are not uniformly distributed, some ciphertexts may occur more frequently than others.

For example, suppose $M = \{m_0, m_1\}$ with $\Pr[M = m_0] \neq \Pr[M = m_1]$. Since encryption randomly maps messages to ciphertexts, the probability of a given ciphertext depends on the distribution of messages, leading to $\Pr[C = c_0] \neq \Pr[C = c_1]$.

Conclusion: The statement is **false**; perfect secrecy does not require ciphertexts to be equally probable.

Question 5: Equivalence of Perfect Secrecy Definitions

Prove the equivalence of two definitions of perfect secrecy.

Definition 1 (Bayesian Formulation)

An encryption scheme $\Pi = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ over message space M is perfectly secret if:

$$\Pr[M = m \mid C = c] = \Pr[M = m]$$

for all messages m and ciphertexts c with $\Pr[C = c] > 0$. This means that observing c provides no additional information about m.

Definition 2 (Adversarial Indistinguishability)

An encryption scheme is perfectly secret if for every adversary A:

$$\Pr[PrivK^{eav}_{A,\Pi} = 1] = \frac{1}{2}.$$

Here, an adversary chooses two messages m_0, m_1 and receives the encryption of one of them, $\text{Enc}_k(m_b)$, where $b \in \{0, 1\}$ is randomly chosen. The adversary must guess b', and success means $b' = b$. Perfect secrecy ensures the adversary has no advantage over random guessing.

(a) Proof that Definition 1 Implies Definition 2

1. By Definition 1, knowing the ciphertext does not affect the probability of any plaintext. 2. Thus, for any messages m_0, m_1 and ciphertext c,

$$\Pr[C = c \mid M = m_0] = \Pr[C = c \mid M = m_1].$$

3. Since ciphertexts are equally likely for both messages, the adversary cannot distinguish which one was encrypted. 4. Therefore, the best strategy is random guessing, meaning:

$$\Pr[PrivK^{eav}_{A,\Pi}] = \frac{1}{2}.$$

Conclusion: Definition 1 $\Rightarrow$ Definition 2.

(b) Proof that Definition 2 Implies Definition 1

1. Assume Definition 2 holds: no adversary can distinguish encryptions of m_0 and m_1 better than random guessing. 2. This implies that for all ciphertexts c,

$$\Pr[C = c \mid M = m_0] = \Pr[C = c \mid M = m_1].$$

3. Using Bayes' theorem:

$$\Pr[M = m \mid C = c] = \frac{\Pr[C = c \mid M = m] \Pr[M = m]}{\Pr[C = c]}.$$

4. Since $\Pr[C = c \mid M = m]$ is the same for all messages, C does not affect M, satisfying Definition 1.
Conclusion: Definition 2 $\Rightarrow$ Definition 1.

Final Conclusion

Since both implications hold, can conclude that the two definitions of perfect secrecy are equivalent.