CSE350: Programming Assignment 1

Ananya Lohani (2019018), Mihir Chaturvedi (2019061)

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Vigenère cipher

- Polyalphabetic substitution cipher
- Secret keyword is used to encrypt a message by **repeating it over and over to generate a keystream**
- This keystream is then used to encrypt the message by adding each letter of the message and the corresponding letter of the keystream modulo 26 (i.e., the number of letters in the alphabet).

Constraints for choosing Inputs

- Plaintext
  - Plaintext characters must be lowercase English alphabets, i.e., {a, b, …, z}
  - The plaintext must be 567 characters long (512 + 64 for its hash).
  - The plaintext must satisfy the property `p = (s, Hash(s))`.

- Key
  - Length of key must be 4
  - Key characters must be lowercase English alphabets, i.e., {a, b, …, z}

Encryption

1. Convert the key into a sequence of numerical values representing the ASCII codes of each letter in the key. The key is repeated as many times as needed to encrypt the entire plaintext.
2. For each letter in the plaintext, do the following:
    a. If the letter is a lowercase letter:
        i. Add the ASCII values of the current letter and the next letter from the key, modulo 26 (the number of letters in the alphabet).
        ii. Convert the result back to a character after adding the ASCII value of 'a'.
        iii. Add the result to the ciphertext string.
    b. If the letter is not a lowercase letter, simply add the letter to the ciphertext string without modification.
3. Return the ciphertext string as the encrypted ciphertext.

- Time complexity: **O(n)**
    - where n is the length of the plaintext
    - as it needs to iterate through each letter of the plaintext once.

Decryption

1. Convert the key into a sequence of numerical values representing the ASCII codes of each letter in the key. The key is repeated as many times as needed to decrypt the entire ciphertext.
2. For each letter in the ciphertext, perform the following:
    a. If the letter is a lowercase letter, perform the following steps to decrypt it:
        i. Get the next numerical value from the sequence of ASCII values obtained in step 1.
        ii. Subtract the numerical value of the key letter from the numerical value of the ciphertext letter, modulo 26.
        iii. Convert the result back into a character after adding the ASCII value of 'a'.
        iv. Append the resulting character to the plaintext.
    b. If the letter is not a lowercase letter, append it to the plaintext without decryption.
3. Return the plaintext as the output.

- Time complexity: **O(n)**
    - where n is the length of the ciphertext
    - This is because the function performs a constant amount of work for each letter in the ciphertext.

Hashing

We use this hash function to construct plaintexts that are **recognizable**, i.e, those that satisfy the property: `p = (s, Hash(s))`.

1. Divide the input plaintext into blocks of N-character segments, where N is a constant specified in the implementation – in our implementation, **N = 8.**
2. For each block of characters, do the following:
    a. Take the first character of the block and initialize it as the initial hash value for that block.
    b. For each subsequent character in the block, perform the following:
        i. Add the ASCII values of the current hash value and the new character, modulo 26 (the number of letters in the alphabet).
        ii. Convert the result back to a character after adding the ASCII value of 'a'.
        iii. Assign this character as the new hash value for that block.
3. Concatenate all the hash values for each block to form the final hash value.

Verification

We need the plaintext to be recognizable in order to verify whether a key correctly decrypts a given ciphertext in our brute force algorithm.

The verification function `is_recognizable` takes a plaintext input and determines whether it is a valid hash value. The function does the following:

1. Divide the plaintext input into two parts: the first part is used as the original plaintext, and the second part is the expected hash value.
2. Calculate the hash value of the original plaintext by calling the hashing function.
3. Compare the calculated hash value with the expected hash value.
4. If the two values are the same, return "True", indicating that the input is a valid hash value. If the two values are different, return "False", indicating that the input is not a valid hash value.

Brute-force Solution

Since we know the key length to be 4, we can directly perform a brute-force attack by iterating through **all possible combinations of lowercase English alphabets of length 4**.

Total number of possible combinations = **26^4 = 4,56,976**.

For a modern computer, iterating through all possible combinations should complete within a few minutes.

- At 100it/s: 1.27 hours
- At 1000it/s: 7.62 minutes
- At 4000it/s: 1.92 minutes [local benchmark]

The asymptotic time complexity of discovering the key via brute force is $\mathbf{O(26^k n)}$, where k is the length of the key and n is the length of the plaintext. This is because for each key, we call the decrypt function which has a time complexity of O(n).