Cryptology 11^{th} August 2015

Lecture 3: Classical Ciphers III; Kerckhoff's Principle; Definition of Security

Instructor: Goutam Paul *Scribe: Kaushik Nath*

1 Attacking the Substitution cipher

A frequency attack on substitution cipher can be performed using the mono-gram, bi-gram or in general the n-gram frequencies. Size of the keyspace for the n-gram frequencies is $(26)^n!$. We can sort the frequencies in $(26)^n! \log((26)^n!) = n(26)^n \log(26)$ time, the complexity of which is much better than the exhaustive search.

There are two kind of substitution ciphers, namely *mono-alphabetic* and *poly-alphabetic*.

2 Permutation(Transposition) cipher

In permutation cipher parameter is m = block length of the plain text. Permutation is defined on the position of the character within each block. Size of the keyspace is $m!$. For small block length sizes complexity of exhaustive search for a attack is tolerable, but increment of block length makes exhaustive search for a attack intractable. For example if $m = 64$, then the size of the search space is (64)! which is much greater than 2^{64}.

In known plain text model it is easy to break the permutation cipher because with the help of a pair of plain text block and cipher text block one can easily find the encryption map.

3 Vigenère cipher

In this case also the parameter is m = block length of the plain text. For example, let us consider the block divisions of length m for the plaintext,key and ciphertext given below, where the symbol '|' is used as a block seperator.

Plaintext = $x_1x_2x_3\cdots x_m|x_{m+1}x_{m+2}x_{m+3}\cdots x_{2m}|x_{2m+1}x_{2m+2}x_{2m+3}\cdots x_{3m}|x_{3m+1}\cdots$

Key = $k_1k_2k_3\cdots k_m|k_1k_2k_3\cdots k_m|k_1k_2k_3\cdots k_m|k_1\cdots$

Ciphertext = $y_1y_2y_3\cdots y_m|y_{m+1}y_{m+2}y_{m+3}\cdots y_{2m}|y_{2m+1}y_{2m+2}y_{2m+3}\cdots y_{3m}|y_{3m+1}\cdots$

The encryption map for the cipher is

$$y_i = (x_i + k_i) \bmod 26 \ , \ 1 \leq i \leq m$$

which in general can be expressed as

$$y_{i+\lambda m} = (x_{i+\lambda m} + k_i) \bmod 26 \ , \ 1 \leq i \leq m \ , \ \lambda > 0$$

3.1 Attacking the Vigenère cipher when the block length is known

Keyspace $K = \{(k_1, k_2, k_3 \cdots, k_m) : k_i \in \{0, 1, 2 \cdots, n-1\}\}$ and $|K| = n^m$. For English alphabet $n = 26$ and $|K| = 26^m$. Hence, the complexity for the attack using exhaustive search is high.

3.1.1 A method of frequency attack

For a better method when the block length m is known, one can gather the cipher text characters $y_1, y_{m+1}, y_{2m+1}, \cdots$ which are encrypted by the same key k_1. So, one can apply a frequency attack to find the key k_1. The same process can be repeated to find the keys of other positions till the size of the block length exhausts.

3.1.2 Alternate way of frequency attack

Let p_i and q_i to be the probability of the i^{th} character in the plain text and cipher text respectively. Then $\sum_{i=0}^{25} p_i^2$ serves as a signature for the entire plain text. For English alphabet $\sum_{i=0}^{25} p_i^2 \approx 0.065$. If the shift is k then, we should have

$$q_{i+k} = p_i$$

Define

$$I_k = \sum_{i=0}^{25} q_{i+k} p_i$$

We now calculate I_k for each k, and if the guess is correct then using $q_{i+k} = p_i$ we have

$$I_k = \sum_{i=0}^{25} q_{i+k} p_i = \sum_{i=0}^{25} p_i^2 \approx 0.065$$

3.2 Attacking the Vigenère cipher when the block length is unknown

Determination of the block length is the primary issue in this situation. There are methods to find the block length.

3.2.1 Kasiski's method

In this method we look for the 2-length or the 3-length character sequences in a large enough cipher text and calculate the gap lengths. The gap lengths can be assumed to be multiples of the block length. We can then find the block length by computing the gcd of the gap lengths. This method might fail if the assumption that the gap lengths is a multiple of the block lengths is wrong.

3.2.2 Index of coincidence

In this method we initially fix τ the block length. After that we collect $y_1, y_{1+\tau}, y_{1+2\tau}, \cdots$ and compute q_0^τ to q_{25}^τ from the cipher text chuck. Define

$$IC_\tau = \sum_{i=0}^{25} (q_i^\tau)^2$$

IC_τ is equal to 0.065 if $\tau = n$ and much less than 0.065 if $\tau \neq n$. We now check out for different values of τ starting from $\tau = 1$. Iteratively, we can find the block length the complexity of which would be in the order of the unknown block length after finding which the algorithm stops. After finding the block length we can continue with the methods discussed before.

4 Kerckhoffs' principle

The security of a crypto-system should depend only on the secrecy of the key and not on the secrecy of the algorithm

If the algorithm is also kept secret then such ciphers are known as *classified ciphers*. *Classified ciphers* are used in defense and military organizations, the usage and practice of which are not recommended in general.

4.1 Rationale behind Kerckhoffs' principle

a) It is easy to hide a short key than a long algorithm.

b) Algorithm can be easily reverse engineered.

c) It is easy to replace a compromised key than a compromised algorithm, if required.

d) If many parties communicate with different algorithms, then the complexity of maintenance matters.

e) If the algorithm is public, it is available for analysis by experts worldwide, and if it survives we gain high confidence on the algorithm.

f) A flaw in the algorithm can be fixed easily, if it is in public domain.

g) Keeping algorithm open helps in standardization.

5 How to define security

a) An obvious notion from the Kerckhoffs' principle follows that, the key should not be leaked from the cipher text. But it is not sufficient, because for known plain text attack model this notion might fail.

b) Plain text should not be leaked from the cipher text. This may also be insufficient. For example, if we consider a situation where 80% of the plain text is leaked and the remaining 20% is secured then the amount of leaked plain text may be containing sufficient data for a successful attack.

c) No plain text character should be leaked from the cipher text. This even may be insufficient, as for example in bank transactions even if the exact amount of account balance is not revealed from the cipher text, but leakage of any extra information about the account balance would be undesirable.

d) No "*meaningful information*" about the plain text should be revealed from the cipher text. For example, in bank transactions even if the exact balance of an account related to a transaction is not revealed, but any kind of information, for example like, "*account balance is more than one lakh*" might come out to be a meaningful information leading to insecurity of the account.