Network and Information Security
Lecture 9

B.Tech. Computer Engineering
Sem. VI.

Prof. Mrudang T. Mehta
Associate Professor
Computer Engineering Department
Faculty of Technology,
Dharmsinh Desai University, Nadiad

Cryptanalysis

Two parts:

1. Finding the length of the key
2. Finding the key itself

- For 1st, there are several methods, one such method is 'kasiski test'
  - Cryptanalyst searches for repeated text segments, of at least three characters in the cipher text.
  - Suppose that two of these segments are found and the distance between them is d.

- The cryptanalyst assumes that d / m ,ie. d divides m
- Where m = key length
- If more repeated segments are found with distances (d_1, d_2,...d_n), then take, gcd (d_1, d_2,...d_n) / m
- This assumption is logical because if the two characters are same and are (k x m) (k=1,2,...) characters apart in the plaintext, they are same and (k x m) characters apart from the ciphertext.
- Cryptanalyst uses segments of at least three characters to avoid the cases where the characters in the key are not distinct.

- The index of coincidence (IC) method is used to confirm the m value determined by the kasiski test.
- Definition:
- The index of coincidence of $x = x_1,x_2,...x_n$, which is a string of length n formed by the alphabets A, B, ..., Z is defined as probability that the random elements of x are the same.
- Frequencies of A, B, C,....,Z in x are denoted by the $f_0,f_1,...f_{25}$
- $I_c(x) = \sum_{fi}C_2 \ / \ _nC_2$

  $= \sum f_i \times (f_i -1) / n \times (n -1) = \sum(f_i/n)^2$

- The index of coincidence (IC) is an invariant for any shift cipher.
- This is because in a shift cipher, the individual probabilities will get permuted but the sum of the squares of the probabilities will remain constant.
- For standard english language text, the value of IC is approximately (0.065).
- However, if all the letters are equally likely then the IC value is 0.038.

n=length {there are 26 alphabets and each is appearing nearly equal number of times.}

$P_i = (n/26)/n \quad = 1/26$

$$IC(x) = \sum P_i^2$$

$$= \sum_{i=0}^{i=25} (1/26)^2$$

$$= 26\,(1/26)^2$$

$$= 1/26 = 0.038$$

- Since, these two values are quite far apart, the IC serves as an important tool to “distinguish” between English text and a random string of English alphabets.
- How to verify the value of m?
- Arrange the given alphabetic string $Y = Y_1....Y_n$, into m substring as follows:

$$Y_1 = Y_1 Y_{m+1} Y_{2m+1}\cdots\cdots$$

$$Y_2 = Y_2\ Y_{m+2} Y_{2m+2}\cdots\cdots$$

$$Y_m = Y_m\ Y_{2m}\ Y_{3m}\cdots\cdots$$

- If the value of m reported by Kasiski test is correct, each substring Y_i, 1<= i <=m is a shift cipher which has been shifted by a key K_i.
- Hence, the expected value of $I_c(Y_i)$ is about 0.065.
- However, if the guess of m is incorrect, each substring is a random string and thus the IC value is about 0.038.
- Thus we can confirm the value of m reported by the Kasiski test.

- Next we investigate a method to actually determine the key $K = (k_1, k_2, \ldots k_m)$

Mutual Index of Coincidence (MI) between two alphabetic strings x and y.

Definition:

Suppose, $x = x_1 x_2 \ldots\ldots x_n$ and $y = y_1 y_2 \ldots y_{n'}$ are two alphabetic strings

- The mutual index of coincidence between x and y is the probability that a random element of x is equal to that of y.

- Thus if the probabilities of A, B.....are $f_0, f_1, \ldots. f_{25}$ and $f'_0, f'_1, \ldots. f'_{25}$ respectively in x and y, then

  Length of x =n, Length of y=n'

- $MI_c(x,y) = \sum_{i=0}^{i=25} f_i f'_i \;/nn'$

- For string x (shift by K_i)

Letter	A	B	C	Z
Probability	p_0	p_1	p_2	p_{25}

- Shift by key k_i

$A + k_i$	$B + k_i$	$Z + k_i$
p_0	p_1	p_{25}

- To find which out of $A + k_i$ $B + k_i$ $Z + k_i$ is mapped to A.

- Consider that a letter denoted by a number j between 0 to 25 in the unencrypted text thus becomes

  $j + k_i = 0 \ (\mathrm{mod}\ 26)$

  $j = -k_i \ (\mathrm{mod}\ 26)$

- Hence, corresponding probability of A in the encrypted text is $p_j = p_{-ki}$

  Suffix values are modulo 26 (e.g. $p_3 \equiv p_{-23}$)

- Thus if we consider two strings x and y, which have been shifted by k_i and k_j respectively, the probability that both characters in x and y are A is $p_{-ki}p_{-kj}$
- Similarly for B,

  $(j + k_i) = 1 \pmod{26}$

  $j = (1 - k_i) \pmod{26}$
- Likewise, the probability that both the characters are B is $p_{1-ki}p_{1-kj}$ and so on.

- Total probability that randomly selected characters are same from X and Y

= sum of all such probabilities

= Both are A or Both are B or ...Both are Z

Since all the events are Mutually Exclusive, It can be written as sum.

=P(both are A's) + P(both are B's) +...+P(both are Z's)

$= p_{-ki}\, p_{-kj} + p_{1-ki}\, p_{1-kj} + \ldots + p_{25-ki}\, p_{25-kj}$

$$MI_c\ (x,y) = \sum_{h=0}^{h=25} p_{h-ki}\, p_{h-kj}$$

- $h' = h - k_i \Rightarrow h = h' + k_i$

$$MI_c(x,y) = \sum_{h=0}^{h=25} p_{h-ki}\, p_{h-kj}$$

$$MI_c(x,y) = \sum_{h=-k_i}^{h=25-k_i} p_{h'}\, p_{h'+ki-kj}$$

- $h' = -k_i$ to $(25 - k_i)$ is equivalent to $h' = 0$ to 25

$$MI_c(x,y) = \sum_{h'=0}^{h'=25} p_{h'}\, p_{h'+ki-kj}$$

$$MI_c(x,y) = \sum_{h=0}^{h=25} p_h\, p_{h+ki-kj}$$

If $k_i = k_j$ then $MI(X,Y) = \sum_{h=0}^{h=25} p_h^2 \quad =0.065$