Cryptanalysis

- Assumption: (Kerckhoffs’ principle) the cryptosystem used is known to the opponent.
  - designer should not assume what system used can remain secret.
  - attack models: kind of information available to the adversary.
Attack models

- $x=$plaintext, $y=$ciphertext.
  1. ciphertext only attack: only $y$ is known. (weakest type of attack models)
  2. known plaintext attack: some $(x,y)$ where $x \rightarrow y$ is known.
  3. chosen plaintext attack: temporary access to $e_k(x)$ [encryption machine]
  4. chosen ciphertext attack: temporary access to $d_k(y)$ [decryption machine]
Cryptanalysis using statistical properties

- statistical analysis is useful for most of mono-alphabetic cryptosystems of English text.

- basic idea:
  - relative frequency of 26 letters are quite different. (e.g. “E” vs. “Z”).
  - there are popular digrams (e.g. “TH”) and trigrams (e.g. “ING”).
### Table 1.1. Letter Frequency

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.082</td>
<td>N</td>
<td>0.067</td>
</tr>
<tr>
<td>B</td>
<td>0.015</td>
<td>O</td>
<td>0.075</td>
</tr>
<tr>
<td>C</td>
<td>0.028</td>
<td>P</td>
<td>0.019</td>
</tr>
<tr>
<td>D</td>
<td>0.043</td>
<td>Q</td>
<td>0.001</td>
</tr>
<tr>
<td>E</td>
<td>0.127</td>
<td>R</td>
<td>0.060</td>
</tr>
<tr>
<td>F</td>
<td>0.022</td>
<td>S</td>
<td>0.063</td>
</tr>
<tr>
<td>G</td>
<td>0.020</td>
<td>T</td>
<td>0.091</td>
</tr>
<tr>
<td>H</td>
<td>0.061</td>
<td>U</td>
<td>0.028</td>
</tr>
<tr>
<td>I</td>
<td>0.070</td>
<td>V</td>
<td>0.010</td>
</tr>
<tr>
<td>J</td>
<td>0.002</td>
<td>W</td>
<td>0.023</td>
</tr>
<tr>
<td>K</td>
<td>0.008</td>
<td>X</td>
<td>0.001</td>
</tr>
<tr>
<td>L</td>
<td>0.040</td>
<td>Y</td>
<td>0.020</td>
</tr>
<tr>
<td>M</td>
<td>0.024</td>
<td>Z</td>
<td>0.001</td>
</tr>
<tr>
<td>E</td>
<td>0.127</td>
<td>M</td>
<td>0.024</td>
</tr>
<tr>
<td>T</td>
<td>0.091</td>
<td>W</td>
<td>0.023</td>
</tr>
<tr>
<td>A</td>
<td>0.082</td>
<td>F</td>
<td>0.022</td>
</tr>
<tr>
<td>O</td>
<td>0.075</td>
<td>G</td>
<td>0.020</td>
</tr>
<tr>
<td>I</td>
<td>0.070</td>
<td>Y</td>
<td>0.020</td>
</tr>
<tr>
<td>N</td>
<td>0.067</td>
<td>P</td>
<td>0.019</td>
</tr>
<tr>
<td>S</td>
<td>0.063</td>
<td>B</td>
<td>0.015</td>
</tr>
<tr>
<td>H</td>
<td>0.061</td>
<td>V</td>
<td>0.010</td>
</tr>
<tr>
<td>R</td>
<td>0.060</td>
<td>K</td>
<td>0.008</td>
</tr>
<tr>
<td>D</td>
<td>0.043</td>
<td>J</td>
<td>0.002</td>
</tr>
<tr>
<td>L</td>
<td>0.040</td>
<td>Q</td>
<td>0.001</td>
</tr>
<tr>
<td>C</td>
<td>0.028</td>
<td>X</td>
<td>0.001</td>
</tr>
<tr>
<td>U</td>
<td>0.028</td>
<td>Z</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Common Digram and Trigram

- **Common Digrams:**
  - TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, TI, IS, ET, IT, AR, TE, SE, HI, OF

- **Common Trigram:**
  - THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH
Cryptanalysis: affine cipher

- Encryption $e_k(x) = a \times x + b \mod 26$.
  - “a” and “b” are unknown.
- Some known ciphertext as in e.g. 1.10: FMXVEDKAPHERBNDKRXRSRE...
  - Frequency table in Table 1.2 (page 28).
- Matching most popular letters between Table 1.1 and Table 1.2 can be useful to solve “a” and “b”. (two unknowns and two equations)
## Table 1.1 vs. Table 1.2

<table>
<thead>
<tr>
<th></th>
<th>0.127</th>
<th>M</th>
<th>0.024</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.091</td>
<td>W</td>
<td>0.023</td>
</tr>
<tr>
<td>A</td>
<td>0.082</td>
<td>F</td>
<td>0.022</td>
</tr>
<tr>
<td>O</td>
<td>0.075</td>
<td>G</td>
<td>0.020</td>
</tr>
<tr>
<td>I</td>
<td>0.070</td>
<td>Y</td>
<td>0.020</td>
</tr>
<tr>
<td>N</td>
<td>0.067</td>
<td>P</td>
<td>0.019</td>
</tr>
<tr>
<td>S</td>
<td>0.063</td>
<td>B</td>
<td>0.015</td>
</tr>
<tr>
<td>H</td>
<td>0.061</td>
<td>V</td>
<td>0.010</td>
</tr>
<tr>
<td>R</td>
<td>0.060</td>
<td>K</td>
<td>0.008</td>
</tr>
<tr>
<td>D</td>
<td>0.043</td>
<td>J</td>
<td>0.002</td>
</tr>
<tr>
<td>L</td>
<td>0.040</td>
<td>Q</td>
<td>0.001</td>
</tr>
<tr>
<td>C</td>
<td>0.028</td>
<td>X</td>
<td>0.001</td>
</tr>
<tr>
<td>U</td>
<td>0.028</td>
<td>Z</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>8</th>
<th>X</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>7</td>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>H</td>
<td>5</td>
<td>O</td>
<td>1</td>
</tr>
<tr>
<td>K</td>
<td>5</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>4</td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td>4</td>
<td>G</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>3</td>
<td>I</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>J</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>2</td>
<td>Q</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>2</td>
<td>T</td>
<td>0</td>
</tr>
<tr>
<td>P</td>
<td>2</td>
<td>W</td>
<td>0</td>
</tr>
<tr>
<td>U</td>
<td>2</td>
<td>Z</td>
<td>0</td>
</tr>
</tbody>
</table>
Cryptanalysis: affine cipher

- Encryption $e_k(x) = ax + b \mod 26$.  
  - “$a$” and “$b$” are unknown.
- Matching Table 1.1 and Table 1.2 can reduce the number of ways to solve “$a$” and “$b$”. (read e.g. 1.10, page 28-29).
- NOTE: since the key space is small (how many ?), we can easily solve by an exhaustive search program.
Cryptanalysis: substitution cipher

- The key space is $26!$, too big for exhaustive search.
- We can use frequency table approach, along with digrams, and trigrams to solve.
  - For a more complicated example, read e.g. 1.11, page 29-32.
Cryptanalysis: Vigenere cipher

- keyword of length m: $k = (k_1, k_2, ..., k_m)$
  - m = key word length.
- $y = (y_1, y_2, ..., y_n)$ is observed
  - n = (large) ciphertext length. Assume $m | n$.
- $e_k(x_i) = x_i + k_i \pmod{26}$, $i = 1, ..., m$ is the "position" in each block of size m.
- Cryptanalysis: need to find $m$ and $k$. 
Cryptanalysis: Vigenere cipher

- e.g. 1.12 (page 34) ciphertext:
  - CHREEVOAHMAERATBIAXXWTNXBE...
- Q: how to find m and keyword k?
- A: Kasiski test.

- NOTE: CHR appeared five times at position 1, 166, 236, 276, and 286.
  - “distances” are multiple of 5. Hence m=5.
  - Other systematic method?
I_c(x): index of coincidence

- x = (x_1, x_2, ..., x_n)
- Let f_0, f_1, ..., f_25 be the frequency counts of letters 'A', 'B', ..., 'Z' in x.
- Q: Randomly choosing two letters from x, what is the probability of being identical letter? [denoted as I_c(x)]
- A: I_c(x) = \sum [f_i (f_i -1)]/[n(n-1)]. (why?)
Using $I_c(x)$ to find $m$

- $y = (y_1, y_2, \ldots, y_n)$ be the ciphertext.
- Divide $y$ into $m$ (guess) sub-strings as
  - $y_1 = y_1, y_{m+1}, y_{2m+1}, \ldots$
  - $y_2 = y_2, y_{m+2}, y_{2m+2}, \ldots$
  - $\ldots$
  - $y_m = y_m, y_{2m}, y_{3m}, \ldots$
- Compute $I_c(y_i), i=1, 2, \ldots, m$.
  - for correct $m$, the values of $I_c(y_i) \approx 0.065$?
  - for incorrect $m$, the values of $I_c(y_i) \approx 0.038$?
Block length m determination

- Recall $I_c(x) = \sum \frac{f_i(f_i - 1)}{n(n-1)}$
  - If $x$ is a regular English text, $I_c(x) \approx \sum p_i^2 = 0.065$.
    - $p_i =$ the relative frequency in Table 1.1. (i=0,1,...,25)
    - Note: $I_c(x)$ remains unchanged with permutation.
  - If $x$ is a random text, $I_c(x) \approx \sum \left(\frac{1}{26}\right)^2 = 0.038$.

- For m indices $I_c(y_i), i=1, 2, .., m$.
  - if m is correct, $y_i$ is a sub-string of regular English text, and the values of $I_c(y_i) \approx 0.065$
  - if m is incorrect, $y_i$ is a sub-string of random test, and values of $I_c(y_i) \approx 0.038$
Example 1.12. Find m.

- e.g. 1.12 (page 34) ciphertext:
  - CHREEVOAHMAERATBIAXXWTNXBE...
- if m=1, only one string, \(I_c(y) = 0.045\)
- if m=2, \(I_c(y_1) = 0.046\), \(I_c(y_2) = 0.041\)
- if m=3, \(I_c(y_i) = 0.043, 0.050, 0.047\)
- if m=4, \(I_c(y_i) = 0.042, 0.039, 0.045, 0.040\)
- if m=5, \(I_c(y_i) = 0.063, 0.068, 0.069, 0.0061, 0.072. (\approx 0.065)\)
Find key $k$

- Divide $y$ into $m$ sub-strings as
  - $y_1 = y_1, y_{m+1}, y_{2m+1}, \ldots$
  - $y_2 = y_2, y_{m+2}, y_{2m+2}, \ldots$
  - $\ldots$
  - $y_m = y_m, y_{2m}, y_{3m}, \ldots$

- Note: Each letter in $y_i$ has been shifted by the same amount $g=k_i$. We search for $g$ such that $M_g = \sum p_j Q_{j+g} = \sum p_j^2 \approx 0.065$. [why ?]
  - $Q_{j+g}$ is the relative letter frequency in $y_i$. 
Example 1.12. Find key $k$

- e.g. 1.12 (page 34) ciphertext:
  - CHREEVOAHMAERATBIAXXWTNXBE...
- $m=5$. Divide the ciphertext into 5 substrings. $y_1, y_2, \ldots, y_5$.
- For each $g=0,1,2,\ldots,25$ compute $M_g(y_i)$ as shown in Table 1.4 (page 35)
- The correct key index $g$ are boxed.
  - $k=(9,0,13,4,19)=\text{JANET.}$ (show!)
TABLE 1.4
Values of $M_g$

<table>
<thead>
<tr>
<th>i</th>
<th>0.035</th>
<th>0.031</th>
<th>0.036</th>
<th>0.037</th>
<th>0.035</th>
<th>0.039</th>
<th>0.028</th>
<th>0.028</th>
<th>0.018</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.061</td>
<td>0.039</td>
<td>0.032</td>
<td>0.040</td>
<td>0.038</td>
<td>0.038</td>
<td>0.045</td>
<td>0.036</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>0.042</td>
<td>0.043</td>
<td>0.036</td>
<td>0.033</td>
<td>0.049</td>
<td>0.043</td>
<td>0.042</td>
<td>0.036</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.069</td>
<td>0.044</td>
<td>0.032</td>
<td>0.035</td>
<td>0.044</td>
<td>0.034</td>
<td>0.036</td>
<td>0.033</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>0.031</td>
<td>0.042</td>
<td>0.045</td>
<td>0.040</td>
<td>0.045</td>
<td>0.046</td>
<td>0.042</td>
<td>0.037</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>0.034</td>
<td>0.037</td>
<td>0.032</td>
<td>0.034</td>
<td>0.043</td>
<td>0.032</td>
<td>0.026</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.048</td>
<td>0.029</td>
<td>0.042</td>
<td>0.043</td>
<td>0.044</td>
<td>0.034</td>
<td>0.038</td>
<td>0.025</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>0.049</td>
<td>0.035</td>
<td>0.031</td>
<td>0.035</td>
<td>0.066</td>
<td>0.035</td>
<td>0.038</td>
<td>0.036</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>0.027</td>
<td>0.035</td>
<td>0.034</td>
<td>0.034</td>
<td>0.036</td>
<td>0.035</td>
<td>0.046</td>
<td>0.040</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.045</td>
<td>0.032</td>
<td>0.033</td>
<td>0.038</td>
<td>0.060</td>
<td>0.034</td>
<td>0.034</td>
<td>0.034</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>0.033</td>
<td>0.033</td>
<td>0.043</td>
<td>0.040</td>
<td>0.033</td>
<td>0.029</td>
<td>0.036</td>
<td>0.040</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>0.037</td>
<td>0.050</td>
<td>0.034</td>
<td>0.034</td>
<td>0.039</td>
<td>0.044</td>
<td>0.038</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.034</td>
<td>0.031</td>
<td>0.035</td>
<td>0.044</td>
<td>0.047</td>
<td>0.037</td>
<td>0.043</td>
<td>0.038</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>0.037</td>
<td>0.033</td>
<td>0.032</td>
<td>0.036</td>
<td>0.037</td>
<td>0.036</td>
<td>0.045</td>
<td>0.032</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>0.044</td>
<td>0.072</td>
<td>0.037</td>
<td>0.027</td>
<td>0.031</td>
<td>0.048</td>
<td>0.036</td>
<td>0.037</td>
<td></td>
</tr>
</tbody>
</table>
Hill cipher

- $P = C = (Z_{26})^m$
- $K = (Z_{26})^{mxm}$, key $K$ is a $mxm$ matrix.
- plaintext: $x = (x_1, x_2, ..., x_m)$
- ciphertext: $y = (y_1, y_2, ..., y_m)$
- $e_k(x) = x K \pmod{26}$,
- $d_k(y) = y K^{-1} \pmod{26}$. 
Cryptanalysis: Hill cipher

- Can be hard to break with ciphertext only.
  - statistical frequency analysis is not useful. Why not?
- However, it is quite simple to break under known plaintext attack.
  - Collect at least m pairs of \((x_i, y_i)\) and solve a mxm matrix equation. (how?)
Break Hill cipher

- For \( i=1,2, \ldots, m \)
  - i-th plaintext: \( \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im}) \)
  - i-th ciphertext: \( \mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im}) \)
  - \( \mathbf{y}_i = \mathbf{x}_i \mathbf{K} \pmod{26}, \mathbf{K} \) unknown.
- Q: how find \( \mathbf{K} \) (and therefore \( \mathbf{K}^{-1} \))?
- A: stack \( \mathbf{x}_i \) together as matrix \( \mathbf{X} \), stack \( \mathbf{y}_i \) together as matrix \( \mathbf{Y} \). We can solve

\[
\mathbf{Y} = \mathbf{X} \mathbf{K} \text{ by } \mathbf{Y} \mathbf{X}^{-1} = \mathbf{K} \pmod{26}.
\]
Example 1.13

- Suppose friday → PQCFKU using Hill cipher with m=2. Find the key matrix.

1. fr → PQ: [5, 17] = [15,16] K
2. id → CF: [8, 3] = [2,5] K
3. ay → KU: [0, 24] = [10, 20] K

- From first two equations, we can solve a 2x2 matrix equation: (show!!)
  \[ Y = X K \]

- Q: what if we don’t know m?
LFSR key stream cipher

- \( z_{i+m} = c_0 z_i + c_1 z_{i+1} + \ldots + c_{m-1} z_{i+m} \mod 2 \)
  - \( c_0 = 1, \) other \( c_i \) is 0 or 1.
  - Max period = \( 2^m - 1. \)
  - How to choose “keys” \( c_i \) ?

- We can break the cryptosystem with a partial sequence (with length \( 2^m \)) of \( z_i \).

- Q: How?
Cryptanalysis: LFSR stream cipher

- $z_{m+1} = c_0 z_1 + c_1 z_2 + \ldots + c_{m-1} z_m$
- $z_{m+2} = c_0 z_2 + c_1 z_3 + \ldots + c_{m-1} z_{m+1}$
- $z_{m+3} = c_0 z_3 + c_1 z_4 + \ldots + c_{m-1} z_{m+2}$
- ...
- $z_{2m} = c_0 z_m + c_1 z_{m+1} + \ldots + c_{m-1} z_{2m-1}$

re-written as (column) $z = M c$, $c = M^{-1} z$.
- $m$ equations, $m$ unknowns.
LFSR Cryptoanalysis Example

- e.g. 1.14 (page 38). Assume m=5 is known.
- Given a pair of \((x,y)\) for \(x \rightarrow y = x+z \mod 2\).
- key stream (LFSR) is \(z = x+y \mod 2\). (why?)
- we can find the key stream generator.

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(z = x+y \mod 2)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Example 1.14 (page 38)

\[ \mathbf{z} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \end{pmatrix}' \]

\[ \mathbf{c} = \mathbf{M}^{-1} \mathbf{z} \]

\[ \mathbf{z}_{i+5} = (\mathbf{z}_i + \mathbf{z}_{i+3}) \mod 2 \]
Summary and HW2

- What we have learned so far?
- Crypto-analysis for various ciphers?
- Mathematical tools?

- HW2: Write a program to solve 1.21 or to verify Table 1.4.
Chapter Review: Modulus and Matrix operations

- Modulus operations
  - \((-a) \mod m, a^{-1} \mod m\).

- Matrix Multiplication

- Matrix Inverse
  - 2x2 matrix
  - m \times m matrix
  - Matrix Inverse \mod m.
Chapter Review: Euler totient function $\phi(n)$

- $\phi(n)$: number of integers between 1 and $n$ that are relative prime to $n$.

- Computation of $\phi(n)$:
  1. $\phi(p^e) = p^{e-1} (p-1)$
  2. $\phi(P \cdot Q) = \phi(P) \cdot \phi(Q)$, if $\gcd(P,Q)=1$.

- E.g.
  - $\phi(20) = \#\{1, 3, 7, 9, 11, 13, 17, 19\} = 8$
  - $\phi(5) = \#\{1, 2, 3, 4\} = 4$
  - $\phi(4) = \#\{1, 3\} = 2$
The Use of Encryption

- DES and AES (Ch 3)
- Cryptographic Hash Functions (Ch 4)
- Digital Signatures (Ch 7)
- Certificates (Ch 9)
- Key Exchange/Distribution (Ch 10)