# Modern Cryptography

## Cryptanalysis of Historical Ciphers

Dr Shashank Singh

# 1 Cryptanalysis of Historical Ciphers

# Cryptanalysis of Historical Ciphers



- Kerckhoffs' Principle.
- Ciphertext-only attack (known ciphertext attack).
- Message space is ordinary English text, without punctuation or spaces.
- Statistical properties of the English language: e.g., letter freq.

| **Letter** | **Prob ($p_i$)** | **Letter** | **Prob** | **Letter** | **Prob** | **Letter** | **Prob** |
|---|---|---|---|---|---|---|---|
| A | .082 | H | .061 | O | .075 | V | .010 |
| B | .015 | I | .070 | P | .019 | W | .023 |
| C | .028 | J | .002 | Q | .001 | X | .001 |
| D | .043 | K | .008 | R | .060 | Y | .020 |
| E | .127 | L | .040 | S | .063 | Z | .001 |
| F | .022 | M | .024 | T | .091 | | |
| G | .020 | N | .067 | U | .028 | | |

# Cryptanalysis of Substitution Cipher



Consider the following ciphertext obtained from a substitution cipher:

YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

- ***Recall a substitution cipher is a mono-alphabetic cipher.***

# Frequency of Letters in the ciphertext



| letter | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| freq. | 0 | 1 | 15 | 13 | 7 | 11 | 1 | 4 | 5 | 11 | 1 | 0 | 16 |

| letter | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| freq. | 9 | 0 | 1 | 4 | 10 | 3 | 2 | 5 | 5 | 8 | 6 | 10 | 20 |

What symbol does encrypt to the letter ‘Z’?

# Cryptanalysis of Substitution Cipher..



Since ‘Z’ occurs significantly more often than any other ciphertext char, we might conjecture that $\mathrm{Dec}_k(Z) = e$.

- The characters C, D, F, J, M, R, and Y occur at least ten times.
- We might expect that these letters are encryptions of (a subset of) t, a, o, i, n, s, h, r.
- But the frequencies do not vary enough to tell us what the correspondence might be.

# Cryptanalysis of Substitution Cipher..



- Look for the *digrams*, especially those of the form −Z or Z−, since we conjecture that Z decrypts to e.
  E.g., DZ, ZW ( 4-times each ), NZ, ZU ( 3-times each), and RZ, HZ, XZ, FZ, ZR, ZV, ZC, ZD, and ZJ ( twice each).
- Since ZW occurs four times and WZ not at all, and W occurs less often than many other characters, we might guess that $\mathrm{Dec}_k(W) = d$.

# Cryptanalysis of Substitution Cipher..

- Using the frequencies of *digrams* and *trigrams*, We can decrypt many characters.

```
------end---------e----ned---e------------
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

--------e----e---------n--d---en----e----e
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

-e---n------n------ed---e---e--ne-nd-e-e--
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

-ed-----n-----------e----ed-------d---e--n
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

## Cryptanalysis of Substitution Cipher..



- Eliminating the wrong options, we eventually end up getting the complete plaintext

```
o-r-riend-ro--arise-a-inedhise--t---ass-it
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

hs-r-riseasi-e-a-orationhadta-en--ace-hi-e
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

he-asnt-oo-in-i-o-redso-e-ore-ineandhesett
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

-ed-ac-inhischair-aceti-ted--to-ardsthes-n
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

# 2 Cryptanalysis of Vigenère Cipher

# Cryptanalysis of Vigenère Cipher



- The first step is determining the keyword length $m$.
- There is two ways to do that
    - **Kasiski test**, described by Friedrich Kasiski in 1863;
    - By using the **index of coincidence**.

# Cryptanalysis of Vigenère Cipher..

**Kasiski test**:

- Observe that two identical segments of plaintext will be encrypted to the same ciphertext whenever their occurrence in the plaintext is δ positions apart, where $\delta \equiv 0 \pmod m$.
- If we obtain several such distances, say $\delta_1, \delta_2, \ldots$, then we would conjecture that $m$ divides all of the $\delta_i$'s and hence m divides the greatest common divisor of the $\delta_i$'s.

# Analysis of Vigenère Cipher..

- This concept of the index of coincidence was defined by William Friedman in 1920.

**Definition** (Index of Coincidence):
Suppose $x = x_1 x_2 \ldots x_n$ is a string of $n$ alphabetic characters. The index of coincidence of $x$, denoted $I_c(x)$, is defined as the probability that two random elements of $x$ are identical.

- If we denote the frequencies of $A, B, C, \ldots Z$ in $x$ by $f_0, f_1, \ldots, f_{25}$ respectively,

# Analysis of Vigenère Cipher.. (ii)



$$I_{c(x)} = \frac{\sum_{i=0}^{25} {}^{f_i}C_2}{{}^{n}C_2} = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)}$$

- If $x$ is English language text, $I_{c(x)} \approx \sum p_i^2 = 0.065$
- The same applies when $x$ is a ciphertext string obtained using any mono-alphabetic cipher. (why?)
- Suppose we start with a ciphertext string $\alpha = y_1 y_2 ... y_n$ constructed using a Vigenère Cipher.

$$\begin{aligned} \alpha_1 &= y_1 \;\; y_{m+1} \;\; y_{2m+1} \;\; \cdots, \\ \alpha_2 &= y_2 \;\; y_{m+2} \;\; y_{2m+2} \;\; \cdots, \\ &\vdots \\ \alpha_m &= y_m \;\; y_{2m} \;\; y_{3m} \;\; \cdots, \end{aligned}$$

# Analysis of Vigenère Cipher.. (iii)



- If $\alpha_1, \alpha_2, \ldots, \alpha_m$ are constructed in this way, and $m$ is indeed the keyword length, then each value $I_c(\alpha_i)$ should be roughly equal to 0.065.
- If $m$ is not the keyword length, then the substrings $\alpha_i$ will look random, and $I_c \approx \sum \left(\frac{1}{26}\right)^2 = 26\left(\frac{1}{26}\right)^2 = \frac{1}{26} = 0.038$.

# Cryptanalysis of Vigenère Cipher ..

Consider the ciphertext obtained from the Vigenère cipher

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWI
AKLXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJEL
XVRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJTAM
RVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBIPEE
WEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHPW
QAIIWXNRMGWOIIFKEE

# Cryptanalysis of Vigenère Cipher ..



Consider the ciphertext obtained from the Vigenère cipher

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWI
AKLXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJEL
XVRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJTAM
RVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBIPEE
WEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHPW
QAIIWXNRMGWOIIFKEE

- The distances from the first occurrence of CHR to the other four occurrences are 165, 235, 275 and 285, respectively.
- The greatest common divisor of these four integers is 5, so it is very likely that $m = 5$.

# Cryptanalysis of Vigenère Cipher .. (ii)

**Using index of coincidence**:

| $m$ | $(I_c(\alpha_i) : i = 1, 2, ..., m)$ |
|---|---|
| 1 | $(0.045)$ |
| 2 | $(0.046, 0.041)$ |
| 3 | $(0.043, 0.050, 0.047)$ |
| 4 | $(0.042, 0.039, 0.045, 0.040)$ |
| 5 | $(0.063, 0.068, 0.069, 0.061, 0.072)$ |

- Thus, we have good enough evidence that $m = 5$.
- Remains to figure out the **'keyword'** i.e. $K = (k_1, k_2, ..., k_m)$?

# Cryptanalysis of Vigenère Cipher ..



Let $f_0, \ldots, f_{25}$ denote the frequencies of $A, B, \ldots, Z$, respectively, in the string $\alpha_i$ and let $n' = \frac{n}{m}$ be the length of string $\alpha_i$.

- The probability distribution of the 26 letters in $\alpha_i$ is

$$\left(\frac{f_0}{n'}, \frac{f_1}{n'}, \ldots, \frac{f_{25}}{n'}\right),$$

- Since the sub-string $\alpha_i$ is obtained by using a shift $k_i$, we would hope that the shifted probability distribution

$$\left(\frac{f_{k_i+0 \pmod{25}}}{n'}, \frac{f_{k_i+1 \pmod{25}}}{n'}, \ldots, \frac{f_{k_i+25 \pmod{25}}}{n'}\right),$$

  would be "close to" the ideal probability distribution $p_0, \ldots, p_{25}$ of the English language.

# Cryptanalysis of Vigenère Cipher ..



Suppose that $0 \leq k \leq 25$, and define the quantity

$$M_k = \sum_{i=0}^{25} p_i \frac{f_{i+k}}{n'}$$

If $k = k_i$, then we would expect that

$$M_k = \sum_{i=0}^{25} p_i^2 = 0.065$$

- If $k \neq k_i$, then $M_k$ will usually be significantly smaller than 0.065.
- This technique will allow us to determine the correct value of $k_i$ for each value of $i$.

# Cryptanalysis of Vigenère Cipher ..



| $i$ | value of $M_k(\mathbf{y}_i)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | .035 | .031 | .036 | .037 | .035 | .039 | .028 | .028 | .048 |
| | **.061** | .039 | .032 | .040 | .038 | .038 | .045 | .036 | .030 |
| | .042 | .043 | .036 | .033 | .049 | .043 | .042 | .036 | |
| 2 | **.069** | .044 | .032 | .035 | .044 | .034 | .036 | .033 | .029 |
| | .031 | .042 | .045 | .040 | .045 | .046 | .042 | .037 | .032 |
| | .034 | .037 | .032 | .034 | .043 | .032 | .026 | .047 | |
| 3 | .048 | .029 | .042 | .043 | .044 | .034 | .038 | .035 | .032 |
| | .049 | .035 | .031 | .035 | **.066** | .035 | .038 | .036 | .045 |
| | .027 | .035 | .034 | .034 | .036 | .035 | .046 | .040 | |
| 4 | .045 | .032 | .033 | .038 | **.060** | .034 | .034 | .034 | .050 |
| | .033 | .033 | .043 | .040 | .033 | .029 | .036 | .040 | .044 |
| | .037 | .050 | .034 | .034 | .039 | .044 | .038 | .035 | |
| 5 | .034 | .031 | .035 | .044 | .047 | .037 | .043 | .038 | .042 |
| | .037 | .033 | .032 | .036 | .037 | .036 | .045 | .032 | .029 |
| | .044 | **.072** | .037 | .027 | .031 | .048 | .036 | .037 | |

# Cryptanalysis of Vigenère Cipher .. (ii)



- Can we guess the key now? $K = (k_1, k_2, ..., k_m) = ?$