# 古典密码学

清华大学计算机系
于红波
2023年3月1日

# 提纲

- 置换密码 Transposition (permutation) cipher
- 移位密码 Shift Cipher （Caesar Cipher )
- 代换密码 Substitution Cipher
- 频率分析 frequency cryptanalysis
- 维吉尼亚密码 Vigenere cipher
- 一次一密（One-time Pad)

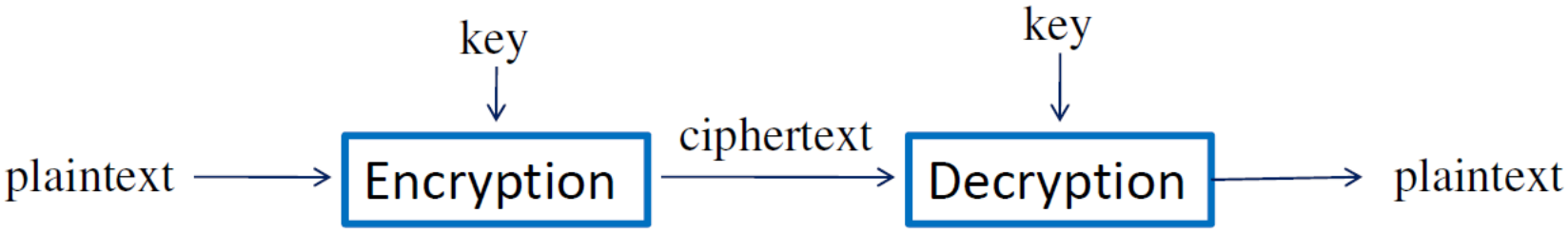
# 推荐文献

- 密码学原理与实践
  - 第一章
- 维基百科 Wikipedia
  - 密码学的发展史
    http://en.wikipedia.org/wiki/History_of_cryptography
  - 古典密码 classical cipher
    http://en.wikipedia.org/wiki/Classical_cipher

# 密码学的分支

- 密码编码学
  - 研究安全高效的信息加密算法和信息认证算法的设计理论与技术
- 密码分析学
  - 研究密码破译的理论与技术
- 密钥管理学
  - 密钥的生成、配送、更新、保存、销毁

# 加解密系统




- 密码体制（Cipher）：用于加解密的算法或装置
- 明文：被加密的消息
- 密文：加密后的消息
- 密钥：密码体制的保密信息
- 加密：将明文消息在密钥的作用下变成密码消息
- 解密：加密的逆过程

# 密码分析学

- 根据敌手掌握的信息类型的不同
  - 唯密文攻击: 攻击者 除了截获的密文，没有其他可用信息
  - 已知明文攻击：攻击者仅知道当前密钥下一些明密文对
  - 选择明文攻击：攻击者能够获得当前密钥下的一些特定的明文对应的密文

    适应性选择明文

    非适应性选择明文
  - 选择密文攻击：攻击者能够获得当前密钥下的一些特定的密文对应的明文

# Transposition(Permution) Cipher

- 置换密码（Permutation Cipher）：保持明文的所有字母不变，只是利用置换打乱了明文字母的位置和次序；对明文字符的位置进行重新排列的一种密码。
- 也称易位密码、换位密码、移位密码

# 置换密码

- 例：设$m$=6，密钥为如下置换：

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\pi(x)$ | 3 | 5 | 1 | 6 | 4 | 2 |

- 上表第一行是关于$x(1 \leqslant x \leqslant 6)$值的列表，第二行是相应置换$\pi(x)$。逆置换为：

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\pi^{-1}(x)$ | 3 | 6 | 1 | 5 | 2 | 4 |

# 置换密码

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\pi(x)$ | 3 | 5 | 1 | 6 | 4 | 2 |

- 假设明文是：*shesellsseashellsbytheseashore*
- 将明文字母分为每六个一组：

  *shesel* | *lsseas* | *hellsb* | *ythese* | *ashore*

- 对每组6个字母使用加密变换可得：

  EESLSH | SALSES | LSHBLE | HSYEET | HRAEOS

- 故密文如下：

  EESLSHSALSESLSHBLEHSYEETHRAEOS

# 置换密码分析方法

- 置换密码的缺点
  - 明文字符的形态不变，导致置换密码信息泄露
  - 一个密文字符的出现次数也是该字符在明文中出现的次数
- 攻击方法
  - 已知明文攻击（直接破译）
  - 唯密文攻击 （查字典）

# 移位密码（凯撒密码）

- 密钥
  - 整数; 1≤K ≤25 （26个英文字母）
- 加密
  - 明文P中的每个字母被它之后的第K个字母替代
- 解密
  - 密文C中的每个字母被它之前的第K个字母替代

# 移位密码（凯撒密码）

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

加密：$c_i = (p_i + K) \bmod 26$
解密：$p_i = (c_i - K) \bmod 26$

当 K=3时，该密码体制成为凯撒密码（Caesar Cipher）在古罗马的战争（公元前54年）中使用

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
D E F G H I J K L M N O P Q R S T U V W X Y Z A B C

# 移位密码（凯撒密码）

- 安全性
  - 对于不知道加密算法的人似乎很难破解
    - 一个密码系统的安全性不在于对加密算法进行保密，而仅在于对密钥的保密
    - Kerckhoffs Principle：密码学的基本假设
  - 对于知道加密方法的人
    - 密码只有25种可能 （暴力破解，brute force）

# 代换密码（Substitution Cipher）

- 密钥K是所有的26个数字0,1,2…,25的一个置换，即
  - 加密：$c_i=S(p_i)$
  - 解密：$p_i=S^{-1}(c_i)$
- 例
  - 密码表S （S-box) 如下

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

X N Y A H P O G Z Q W B T S F L R C V M U E K J D I

明文：C R Y P T O G R A P H Y

密文：Y C D L M F O C X L G D

# 代换密码

- 关键词组：JULIUS CAESAR（JULISCAER）
  - 密码表

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| J | U | L | I | S | C | A | E | R | T | V | W | X | Y | Z | B | D | F | G | H | K | M | N | O | P | Q |

- 安全性分析
  - 暴力搜索所有的密钥不可行
    - 密钥空间（可能的代换表）大小：
    - 26！≈ $4*10^{26} \approx 2^{88.4}$
    - 每秒搜索1billion ($2^{30}$)，需要13 billion年
  - 公元前1个千年里认为是无法破译的

# 代换密码频率分析

- 公元9世纪，阿拉伯科学家al-Kindi 发明
- 主要思想
  - 代换密码没有掩藏密文字母出现的频率
  - 计算密文中字母出现的频率，与（英文）字母统计表相比较，很容易确定代换表（密码表）。

# 代换密码频率分析

- 26个英文字母出现的概率

| 字母 | 概率 | 字母 | 概率 |
|---|---|---|---|
| A | 0.082 | N | 0.067 |
| B | 0.015 | O | 0.075 |
| C | 0.028 | P | 0.019 |
| D | 0.043 | Q | 0.001 |
| E | 0.127 | R | 0.060 |
| F | 0.022 | S | 0.063 |
| G | 0.020 | T | 0.091 |
| H | 0.061 | U | 0.028 |
| I | 0.070 | V | 0.010 |
| J | 0.002 | W | 0.023 |
| K | 0.008 | X | 0.001 |
| L | 0.040 | Y | 0.020 |
| M | 0.024 | Z | 0.001 |

# 代换密码频率分析

- 两个连续的字母（diagrams）出现的概率

| | | |
|---|---|---|
| th 1.52% | en 0.55% | ng 0.18% |
| he 1.28% | ed 0.53% | of 0.16% |
| in 0.94% | to 0.52% | al 0.09% |
| er 0.94% | it 0.50% | de 0.09% |
| an 0.82% | ou 0.50% | se 0.08% |
| re 0.68% | ea 0.47% | le 0.08% |
| nd 0.63% | hi 0.46% | sa 0.06% |
| at 0.59% | is 0.46% | si 0.05% |
| on 0.57% | or 0.43% | ar 0.04% |
| nt 0.56% | ti 0.34% | ve 0.04% |
| ha 0.56% | as 0.33% | ra 0.04% |
| es 0.56% | te 0.27% | ld 0.02% |
| st 0.55% | et 0.19% | ur 0.02% |

# 代换密码频率分析

- 英文中最常用的3字母

  the, ing, and, her, ere, ent,

  tha, nth, was, eth, for, dht

# 代换密码频率分析

- 例：利用代换密码获得如下密文，如何恢复明文？

```
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

在下面的分析中，明文用小写字母表示，密文用大写字母表示

# 代换密码频率分析

Step1：将密文字母出现的频率 与英文字母表中字母出现的频率作比较

| 字母 | 频数 | 字母 | 频数 |
|---|---|---|---|
| A | 0 | N | 9 |
| B | 1 | O | 0 |
| C | 15 | P | 1 |
| D | 13 | Q | 4 |
| E | 7 | R | 10 |
| F | 11 | S | 3 |
| G | 1 | T | 2 |
| H | 4 | U | 5 |
| I | 5 | V | 5 |
| J | 11 | W | 8 |
| K | 1 | X | 6 |
| L | 0 | Y | 10 |
| M | 16 | Z | 20 |

| 字母 | 概率 | 字母 | 概率 |
|---|---|---|---|
| A | 0.082 | N | 0.067 |
| B | 0.015 | O | 0.075 |
| C | 0.028 | P | 0.019 |
| D | 0.043 | Q | 0.001 |
| E | 0.127 | R | 0.060 |
| F | 0.022 | S | 0.063 |
| G | 0.020 | T | 0.091 |
| H | 0.061 | U | 0.028 |
| I | 0.070 | V | 0.010 |
| J | 0.002 | W | 0.023 |
| K | 0.008 | X | 0.001 |
| L | 0.040 | Y | 0.020 |
| M | 0.024 | Z | 0.001 |

Z出现次数最多，可能 S(e)=Z

# 代换密码频率

## Step2:考虑双字出现频率 假设S(e)=Z

| 字母 | 概率 | 字母 | 概率 |
|---|---|---|---|
| A | 0.082 | N | 0.067 |
| B | 0.015 | O | 0.075 |
| C | 0.028 | P | 0.019 |
| D | 0.043 | Q | 0.001 |
| E | 0.127 | R | 0.060 |
| F | 0.022 | S | 0.063 |
| G | 0.020 | T | 0.091 |
| H | 0.061 | U | 0.028 |
| I | 0.070 | V | 0.010 |
| J | 0.002 | W | 0.023 |
| K | 0.008 | X | 0.001 |
| L | 0.040 | Y | 0.020 |
| M | 0.024 | Z | 0.001 |

| | | |
|---|---|---|
| th 1.52% | en 0.55% | ng 0.18% |
| he 1.28% | ed 0.53% | of 0.16% |
| in 0.94% | to 0.52% | al 0.09% |
| er 0.94% | it 0.50% | de 0.09% |
| an 0.82% | ou 0.50% | se 0.08% |
| re 0.68% | ea 0.47% | le 0.08% |
| nd 0.63% | hi 0.46% | sa 0.06% |
| at 0.59% | is 0.46% | si 0.05% |
| on 0.57% | or 0.43% | ar 0.04% |
| nt 0.56% | ti 0.34% | ve 0.04% |
| ha 0.56% | as 0.33% | ra 0.04% |
| es 0.56% | te 0.27% | ld 0.02% |
| st 0.55% | et 0.19% | ur 0.02% |

1） ZW 出现4次

W可能是r,s,n,d,a

2) WZ没有出现

W 不是 r

3) W出现8次（0.047）

所以 S(d)=W

# 代换密码频率分析

Step3：考虑双字

- 假设S(d)=W
- RW出现2次
  - R可能是e,n
    e的密文是Z
  - R可能是n
    S(n)=R

| | | |
|---|---|---|
| th 1.52% | en 0.55% | ng 0.18% |
| he 1.28% | ed 0.53% | of 0.16% |
| in 0.94% | to 0.52% | al 0.09% |
| er 0.94% | it 0.50% | de 0.09% |
| an 0.82% | ou 0.50% | se 0.08% |
| re 0.68% | ea 0.47% | le 0.08% |
| nd 0.63% | hi 0.46% | sa 0.06% |
| at 0.59% | is 0.46% | si 0.05% |
| on 0.57% | or 0.43% | ar 0.04% |
| nt 0.56% | ti 0.34% | ve 0.04% |
| ha 0.56% | as 0.33% | ra 0.04% |
| es 0.56% | te 0.27% | id 0.02% |
| st 0.55% | et 0.19% | ur 0.02% |

# 代换密码频率分析

- 通过以上三步的分析

```
------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
```

# 代换密码频率分析

Step4:双字母

假设S(e)=Z

1)NZ出现3次

N可能是h,r,t

2)ZN没有出现

N可能不是r,t

故 可能 S(h)=N

| | | |
|---|---|---|
| th 1.52% | en 0.55% | ng 0.18% |
| he 1.28% | ed 0.53% | of 0.16% |
| in 0.94% | to 0.52% | al 0.09% |
| er 0.94% | it 0.50% | de 0.09% |
| an 0.82% | ou 0.50% | se 0.08% |
| re 0.68% | ea 0.47% | le 0.08% |
| nd 0.63% | hi 0.46% | sa 0.06% |
| at 0.59% | is 0.46% | si 0.05% |
| on 0.57% | or 0.43% | ar 0.04% |
| nt 0.56% | ti 0.34% | ve 0.04% |
| ha 0.56% | as 0.33% | ra 0.04% |
| es 0.56% | te 0.27% | ld 0.02% |
| st 0.55% | et 0.19% | ur 0.02% |

# 代换密码频率分析

- 考虑3个字母
  - 明文中有 ne-ndhe, 其中 ‘–’ 对应密文C
- 从3个字母组的分布看 and 出现的频率较高
  - 猜测S(a)=C

# 代换密码频率分析

- 进一步

```
------end-----a---e-a--nedh--e------a-----
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

h-------ea---e-a---a---nhad-a-en--a-e-h--e
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

he-a-n------n------ed---e---e--neandhe-e--
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

-ed-a---nh---ha---a-e----ed-----a-d--he--n
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

# 代换密码频率分析

- 确定M
  - M是密文中出现频率次高的字母
    - M可能是 t, a, o , i, n, s, h, r
  - NRM解密成nh-
    - h-可能是一个词的开头
    - M应该是一个元音 o,i
  - CM出现在密文中
    - ai, ao
    - 猜测S(i)=M

| 字母 | 概率 | 字母 | 概率 |
|---|---|---|---|
| A | 0.082 | N | 0.067 |
| B | 0.015 | O | 0.075 |
| C | 0.028 | P | 0.019 |
| D | 0.043 | Q | 0.001 |
| E | 0.127 | R | 0.060 |
| F | 0.022 | S | 0.063 |
| G | 0.020 | T | 0.091 |
| H | 0.061 | U | 0.028 |
| I | 0.070 | V | 0.010 |
| J | 0.002 | W | 0.023 |
| K | 0.008 | X | 0.001 |
| L | 0.040 | Y | 0.020 |
| M | 0.024 | Z | 0.001 |

# 代换密码频率分析

```
-----iend-----a-i-e-a-inedhi-e------a---i-
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

h-----i-ea-i-e-a---a-i-nhad-a-en--a-e-hi-e
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

he-a-n-----in-i----ed---e---e-ineandhe-e--
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

-ed-a--inhi--hai--a-e-i--ed-----a-d--he--n
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

# 代换密码频率分析

- 确定J
  - 密文中JN出现两次，对应的明文-
    - th出现频率高
    - 猜测S(t)=J
- 确定Y
  - 密文中JY出现一次，对应明文t-
  - 猜测S(o)=Y

| | | |
|---|---|---|
| th 1.52% | en 0.55% | ng 0.18% |
| he 1.28% | ed 0.53% | of 0.16% |
| in 0.94% | to 0.52% | al 0.09% |
| er 0.94% | it 0.50% | de 0.09% |
| an 0.82% | ou 0.50% | se 0.08% |
| re 0.68% | ea 0.47% | le 0.08% |
| nd 0.63% | hi 0.46% | sa 0.06% |
| at 0.59% | is 0.46% | si 0.05% |
| on 0.57% | or 0.43% | ar 0.04% |
| nt 0.56% | ti 0.34% | ve 0.04% |
| ha 0.56% | as 0.33% | ra 0.04% |
| es 0.56% | te 0.27% | ld 0.02% |
| st 0.55% | et 0.19% | ur 0.02% |

# 代换密码频率分析

- 确定D
  - MD出现4次，i-
  - D可能对应n,t,s
  - 猜测S（s)=D
- 确定H
  - 密文HNCMF
  - 明文chai-
  - 猜测S(r)=F

| | | |
|---|---|---|
| th 1.52% | en 0.55% | ng 0.18% |
| he 1.28% | ed 0.53% | of 0.16% |
| in 0.94% | to 0.52% | al 0.09% |
| er 0.94% | it 0.50% | de 0.09% |
| an 0.82% | ou 0.50% | se 0.08% |
| re 0.68% | ea 0.47% | le 0.08% |
| nd 0.63% | hi 0.46% | sa 0.06% |
| at 0.59% | is 0.46% | si 0.05% |
| on 0.57% | or 0.43% | ar 0.04% |
| nt 0.56% | ti 0.34% | ve 0.04% |
| ha 0.56% | as 0.33% | ra 0.04% |
| es 0.56% | te 0.27% | ld 0.02% |
| st 0.55% | et 0.19% | ur 0.02% |

# 代换密码频率分析

```
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
```

# 代换密码频率分析

- 进一步猜测，得到密文

Our friend from Paris examined his empty glass with surprise, as if evaporation had taken place while he wasn't looking. I poured some more wine and he settled back in his chair, face tilted up towards the sun.

# 单表代换

- 问题
  - 明文中字母发生的频率没有被随机化，每个字母被加密成唯一的另外的一个字母
- 如何掩盖加密后密文的统计规律
  - 多表替换（Polyalphabetic substitution）
    - Vigenere密码
    - http://en.wikipedia.org/wiki/Vigenere_cipher
  - 加密多字母（Polygraphic substitution）
    - Playfair 密码
    - http://en.wikipedia.org/wiki/Playfair_cipher
    - Hill密码

# Vigenere Cipher

- 使用多个移位密码 (shift cipher)
- Giovan Battista Bellaso, 1553
- 加密

明文: A T T A C K A T D A W N

密钥: L E MO NL E MON L E

密文: L X F O PV E F R N H R

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
| B | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A |
| C | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B |
| D | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C |
| E | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D |
| F | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E |
| G | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F |
| H | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G |
| I | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H |
| J | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I |
| K | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J |
| L | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K |
| M | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L |
| N | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M |
| O | O | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
| P | P | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| Q | Q | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
| R | R | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
| S | S | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
| T | T | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
| U | U | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
| V | V | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
| W | W | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V |
| X | X | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W |
| Y | Y | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X |
| Z | Z | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y |

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

# Vigenere Cipher

- 定义
  - 明文： $P=(Z_{26})^n$
  - 密文： $C=(Z_{26})^n$
  - 密钥： $K=(Z_{26})^m$ (K由m个字母组成）
  - 加密： $C_i=(P_i+K_{i \bmod m})\bmod 26$
  - 解密： $P_i=(C_i-K_{i \bmod m})\bmod 26$

# Playfair Cipher

- 由Charles Wheatstone发明，1854
- Lyon Playfair提倡在英国军队和政府使用

- 双字代换（digram substitution)
  - Step1:产生密钥表 5*5的矩阵
  - Step2:加密消息

# Playfair Cipher

- 产生密码表
  - 5*5=25的密钥表
    - I和J看成一个字母
    - 第一行（列）是密钥, 密钥是一个单词或词组，去掉重复字母。
      例如：Playfair example
    - 其余按照字母顺序

P L A Y F A
I R E X A M PLE A
B C D EF G H I=J
K LM N O P Q R S
T U V W XY Z

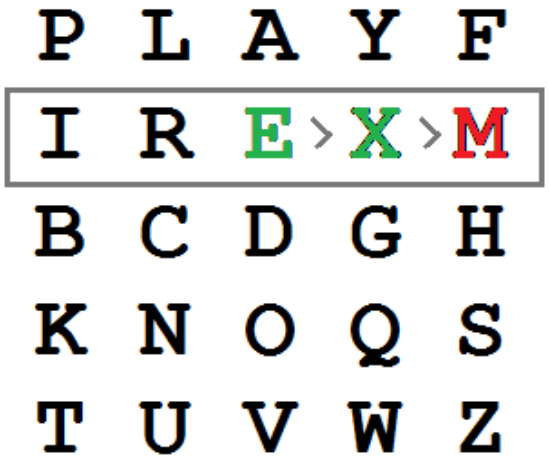
| | | | | |
|---|---|---|---|---|
| P | L | A | Y | F |
| I | R | E | X | M |
| B | C | D | G | H |
| K | N | O | Q | S |
| T | U | V | W | Z |

# Playfair Cipher

- 消息分组
  - 将消息分成两字母一组。如果成对后有两个相同字母紧挨或最后一个字母是单个的，就插入一个字母X
    - 例： communist，应成为co,mx,mu,ni,st

# Playfair Cipher

- 加密消息
  - 若明文$p_1p_2$在同一行，对应密文$c_1c_2$分别是紧靠$p_1p_2$右端的字母。其中第一列被看做是最后一列的右方。
    - 如，按照前表，EX→XM



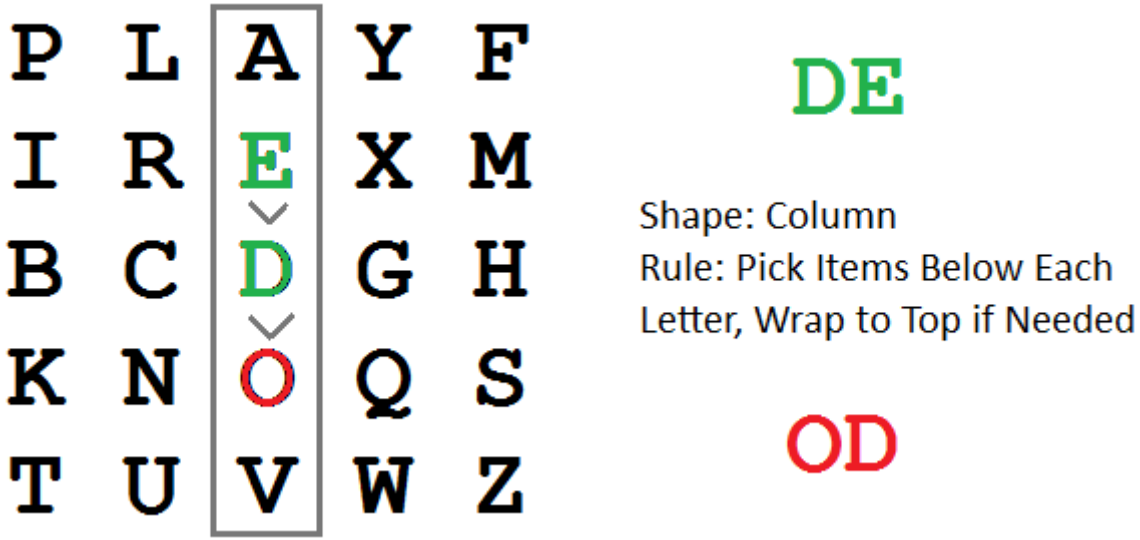

EX

Shape: Row
Rule: Pick Items to Right of Each Letter, Wrap to Left if Needed

XM

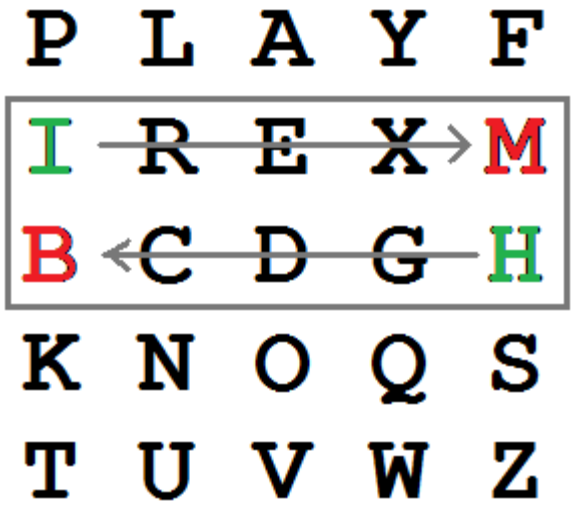
# Playfair Cipher

- 加密消息
  - 若$p_1$ $p_2$在同一列，对应密文$c_1$ $c_2$分别是紧靠$p_1$ $p_2$ 下方的字母。其中第一行被看做是最后一行的下方。
    - 例如 DE→OD



# Playfair Cipher

- 加密消息
  - 若$p_1$ $p_2$不在同一行，不在同一列，则$c_1$ $c_2$是由$p_1$ $p_2$确定的矩形的其他两角的字母
    - 如：HI→BM

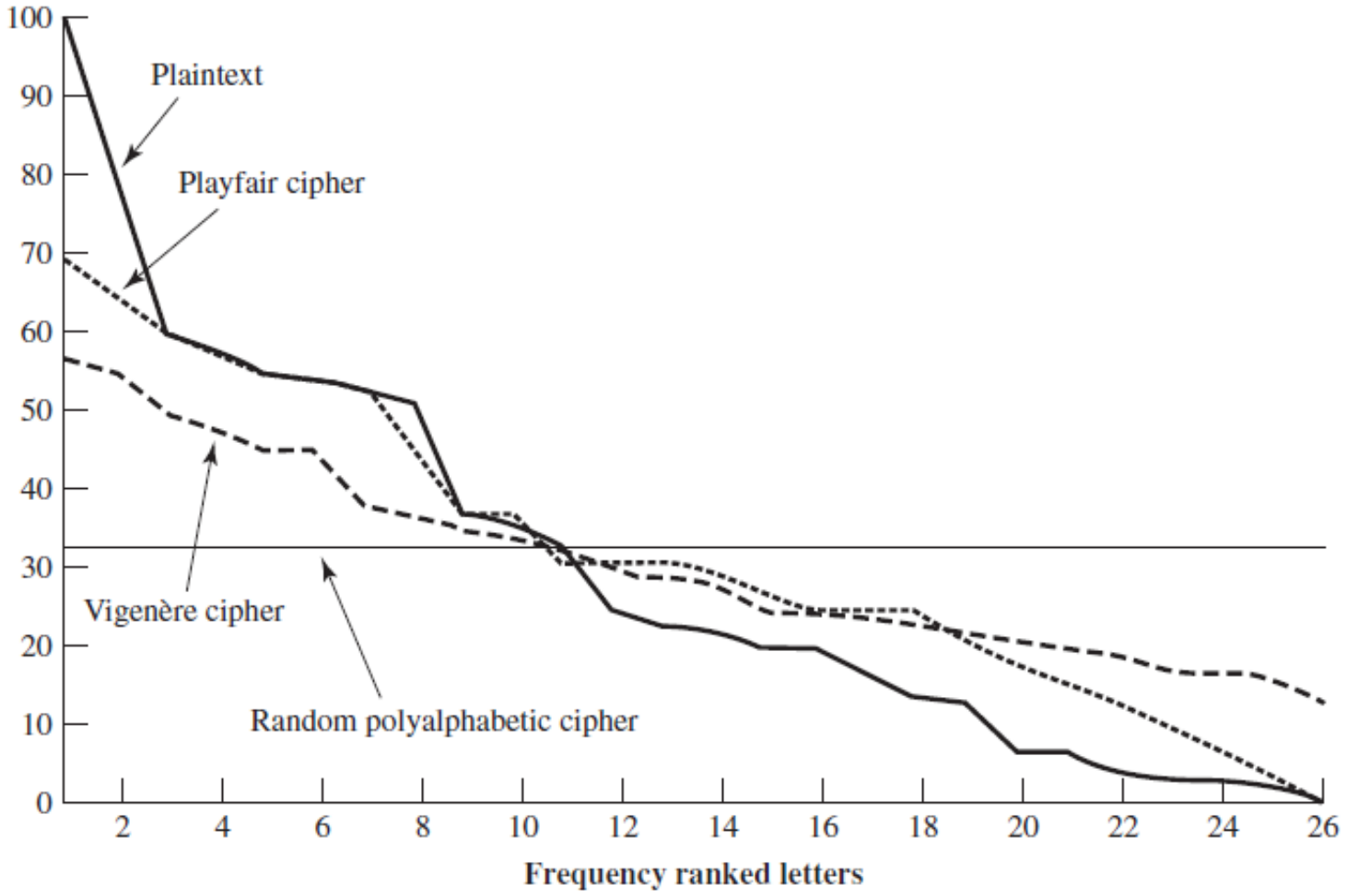



HI

Shape: Rectangle
Rule: Pick Same Rows,
Opposite Corners

BM

# 频率特征


100
90
80
70
60
50
40
30
20
10
0
Plaintext
Playfair cipher
Vigenère cipher
Random polyalphabetic cipher
2
4
6
8
10
12
14
16
18
20
22
24
26
Frequency ranked letters

# Vigenere Cipher的分析

- 分析方法
  - 密钥空间大小为$26^m$，若$m$很大，使用计算机穷尽密钥搜索也需要很长时间
  - 寻找密钥长度，将问题变成简单的移位密码
- 如何寻找m？
  - 两种方法
    - Kasiski测试法，1863
    - 重合指数法分析， 1920

# Vigenere Cipher的分析

- Kasiski test（卡西斯基测试）
  - 基于以下事实
    - 两段相同的明文段将被加密成相同的密文段，则他们的位置间距为m的倍数
  - 算法
    - 搜索长度至少为3的相同的密文段
    - 记下离起始密码段的距离$d_1,d_2,d_3\ldots$
    - 则m整除$\gcd(d_1,d_2,d_3\ldots)$

# Vigenere Cipher的分析

- Kasiski test
  - 例

明文: ***CRYPTO***ISSHORTFOR***CRYPTO***GRAPHY

密钥: ABCDABCDABCDABCDABCDABCDABCDABCD

密文: ***CSASTP***KVSIQUTGQU***CSASTP***IUAQJB

距离为16， 则m 可能为 4, 8, 16

# Vigenere Cipher的分析

- 重合指数法（Index of coincidence）
  - 定义：设**x**=$x_1x_2x_3\ldots,x_n$是一条长度为n的串，**x**的重合指数$I_c(\mathbf{x})$定义为**x**中两个随机元素相同的概率
    - 设$f_0,f_1,\ldots f_{25}$分别表示A,B,…,Z在串x中出现的频数，则
    - 若x是英语文本串， 则 $I_c(x)=\sum_{i=0}^{25} p_i^2 = 0.065$
    - 若x是一个完全的随机串，则 $I_c(x)=\frac{1}{26}=0.038$

$$I_c(x)=\frac{\sum_{i=0}^{25}\binom{f_i}{2}}{\binom{n}{2}}\approx\frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)}$$

# Vigenere Cipher的分析

- 重合指数法
  - 若m猜对，则每一条字符串的重合指数接近于0.065
  - 若m猜错，则每一条字符串的重合指数接近于0.038

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX
VRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
AMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBI
PEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHP
WQAIIWXNRMGWOIIFKEE

# Vigenere Cipher的分析

- $y=y_1y_2y_3\cdots y_n$, 将y分割为m个长度相等的子串

$\mathbf{y_1}=y_1\ y_{m+1}\ y_{2m+1}\cdots$

$\mathbf{y_2}=y_2\ y_{m+2}\ y_{2m+2}\cdots$

$\cdots \qquad\qquad \cdots$

$\mathbf{y_m}=y_m\ y_{2m}\ y_{3m}\cdots$

# Vigenere Cipher的分析

- 重合指数
  - m=1, $I_c$=0.045
  - m=2, $I_c$=0.046,0.041
  - m=3, $I_c$=0.045,0.050,0.047
  - m=4, $I_c$=0.042,0.039,0.045,0.040
  - m=5, $I_c$=0.063,0.068,0.069.0.061,0.072
- Kasiski 测试
  - CHR出现5次，位置1,166,236,276,286,
  - m=gcd(165,235,275,285)=5

# Vigenere密码的弱点

- 密钥长度<明文长度
  - 密钥扩展（repeated）后对每个明文字母使用移位密码加密
- 分析方法
  - 寻找密钥长度m
    - Kasiski test
    - Index of Coincidence
  - 对每个移位密码使用频率分析

# One-Time Pad (OTP)

- 如何加强Vigenere密码
  - 密钥产生
    - 密钥长度与消息长度相同
    - 密钥是随机的
  - 加密
    - 每个密钥仅加密一个消息
- 上面的密码体制被称为一次一密（OTP）

# One-Time Pad (OTP)

- 1917年 ( World War I )
- Gilbert Vernam：AT&T Bell Labs 工程师
- Joseph Mauborgne：时为美军上尉(Captain)后为少校(Major), 曾在1914年首次发表对Playfair密码的解决方案。

# 一次一密（one-time pad）

- 一次一密乱码本：一个**大**的**不重复**的**真随机**密钥**字母集**被写在几张纸上并粘在一起成为一个乱码本。
- 发方：用乱码本中的每一密钥字母加密一个明文字符(明文与密钥模26加)，每个密钥仅对一个消息使用一次，加密后销毁乱码本中用过的部分。
- 收方：有一个同样的乱码本，并依次使用每个密钥去解密密文的每个字符。收方解密后也同样销毁乱码本中用过的部分。
- **新的消息**用乱码本**新的密钥**加密，不能重复使用。
- 所以叫做“one-time pad”。

# One-time Pad(OTP)

- 例：

明文 H E L L O　　L A T E R

密钥 X M C K L　　T Q U R A

密文 E Q N V Z　　E Q N V Z

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

H→7, X→23, 7+23(mod 26)=4(E)

# One-time Pad

- 现代的OTP处理二进制数据（bit 序列）
  - 使用“模2加”(XOR)代替“模26加”
  - (a+b) mod 2 ， a XOR b,$a \oplus b$ ， a^b
- 例：

明文　1010011000

密钥 $\oplus$ 0110101110

密文　1100110110

# 香农简介

- 1949年《保密系统的通信理论》
  *"Communication Theory of Secrecy Systems"*
- 利用信息熵在理论上**证明了一次一密是无法被破译的，同时证明了一个无法被破译的密码系统必须具备与一次一密相同的条件**，即密钥必须有以下特征：
  - 完全随机
  - 不能重复使用
  - 保密
  - 和明文一样长
- 使保密通信由艺术变成科学
- 密码设计的新思想，对现代密码体制设计非常重要。

# OTP的完善保密性

- 一个密码体制的完善保密性是指已知的密文不会泄露明文的任何信息
- 定义：一个密码体制具有完善保密性， 如果对任意的明文p和任意的密文c, 都有

$$Pr(P=p|C=c)=Pr(P=p)$$

  - Pr(P=p|C=c)是已知密文c时明文p的后验概率
  - Pr(P=p)是明文p的后验概率
- 即使知道密文后，攻击者也不能以更高的概率猜测出明文

# OTP的完善保密性

- One-time pad
  - $P=C=K=\{0,1\}^n$
  - K随机产生
    - $\Pr(K=k)=1/2^n$
  - 证明$\Pr(P=p|C=c)=P(P=p)$
- 证明

$$\Pr[C=c \mid P=p]=\Pr[K=p\oplus c]=1/2^n$$

$$\Pr[C=c]=\sum_{p\in P}\Pr[P=p]\Pr[C=c \mid P=p]$$

$$=1/2^n\sum_{p\in P}\Pr[P=p]=1/2^n$$

$$\therefore \Pr(P=p \mid C=c)=\frac{\Pr[C=c \mid P=p]\Pr[P=p]}{\Pr[C=c]}=\Pr[P=p]$$

# 密码体制的安全性

- 无条件安全性(unconditional security)即完善保密性(perfect security)
  - 即使攻击者有无限的计算资源也不可能攻破密码体制，则该密码体制是无条件安全的
  - OTP是无条件安全的

- 计算安全性(computational security)
  - 破译一个密码体制所做的计算上的努力
  - 如果使用最好的算法破译一个密码体制至少需要N次操作（N是一个特定的非常大的数字），定义该密码体制是计算安全的
- 可证明安全性(provable security)
  - 通过规约的方式为安全性提供证据
  - 如果可以破译密码体制A，则就可以解决一个数学难题B（分解因子问题，离散对数问题）

# 课后预习

- 二战Enigma密码
  - 密码机的原理
  - 加密过程
  - 秘钥的产生和传递（日密码，通信密码）
  - Enigma的破译

Enigma密码阅读

https://www.ciphermachinesandcryptology.com/index.htm

Enigma密码破译：电影《模仿游戏》


# 谢　谢！