1 传统密码简介

例题如果用数字 (0-25) 分别和字母 a, b, c, d,..., y, z 相对应, 即

明文	a	b	c	d	e	f	g	h	i	j	k	l	m
数字	0	1	2	3	4	5	6	7	8	9	10	11	12
$k=3$ 密文	D	E	F	G	H	I	J	K	L	M	N	O	P
明文	n	o	p	q	r	s	t	u	v	w	x	y	z
数字	13	14	15	16	17	18	19	20	21	22	23	24	25
$k=3$ 密文	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

则密文字母 ϕ 可以用明文字母 θ 表示如下:

$$\phi = \theta + 3(\mathrm{mod}26) \tag{1}$$

若明文字母为 y, 即 $\theta = 24$ 时,

$$\phi \equiv 24 + 3(\mathrm{mod}26) \equiv 27(\mathrm{mod}26) \equiv 1(\mathrm{mod}26)$$

因此, 密文字母为 E.

- 公式 (1) 是凯撒密码的数学形式, 也表示一种算法. 式中密钥为 3, 实际表示倒退 3 步.
- 此例中, 密钥可为 1 25 之间的任意一个数字, 但若是选择 0 为密钥, 则密钥 = 明文, 实际上没有加密. 故式子 (1) 可以推广成

$$\phi = \theta + k(\mathrm{mod}26)$$

这里 $k \in K, K = 1, 2, 3, ..., 24, 25$ 是密钥集合或密钥空间.

- 密码系统的两个基本单元是算法和密钥.
  - 算法—是一些公式、法则或者程序, 规定着明文和密钥之间的变换方法. 算法相对是稳定的, 可是为常量.
  - 密钥—可看成式算法中的参数, 是一个变量. 可以根据事前的规定好的安排, 或用过若干次后改变一个密钥, 或每过一段时间更换一次密钥, 等等. 为密码系统的安全, 频繁更换密钥是必要的.
- 一般来讲, 算法往往不能保密, 真正需要保密的是密钥.
- 传统密码可以分为两大类: 换位密码和代替密码, 有时可能是二者的组合

2 换位密码

- 换位 (置换) 密码在编制时, 是把明文中的字母重新排列, 但是字母本身不变, 只改变其位置.
- 最简单的换位密码是把明文顺序倒过来, 然后裁成固定长度的字母作为密文.

  例如明文为 this cryptosystem is not secure. (27 个字符)

  密文为 ERUC ESTO NSIM ETSY SOTP YRCS IHT
- 另一种换位密码是把明文按某一顺序排成一个矩阵, 然后按某一顺序选出矩阵中的字母已形成密文, 最后截成固定长度的字母组.

  例如明文为 this cryptosystem is not secure. (27 个字符)

  解 (1) 排成 5*6 矩阵

t	h	i	s	c	r
y	p	t	o	s	y
s	t	e	m	i	s
n	o	t	s	e	c
u	r	e			

  (2) 选出顺序, 排列 (3) 按字符长度 5 写出密文 TYSNU HPTOR ITETE SOMSC SIERY SC
- 可见, 改变矩阵的大小和选出顺序可以得到不同形式的密码.
- 换位密码比较简单, 经不起" 已知明文的攻击". 但是, 将其与其密码相结合, 可以得到十分有效的密码.

  例如明文 can you understand, 使用 type 为密钥进行列换位加密. 解 将明文写成如下 4*4 矩阵

密钥	t	y	p	e
顺序	3	4	2	1
明文矩阵	c o d t	a u e a	n u r n	y n s d

例如明文为 can you understand, 试写出周期为 4 的周期换位密文. 密钥是 $i = 1, 2, 3, 4$ 的置换 $f(i) = 3, 4, 2, 1$

将明文写成如下 4*4 矩阵

置换 $f(i)$	3	4	2	1
明文矩阵	c o d t	a u e a	n u r n	y n s d

按照密钥所确定的顺序, 按行置换该矩阵的字母并依次写出, 即得到密文

YNC ANU OUSRDEDNTA

3 代替密码

代替密码共有 4 种: 单表代替密码 (monalphabetic substitution cipher)、同音代替密码 (homophnic substitution cipher)、多表代替密码 (polyalphabetic substitution cipher) 和多字母代替密码 (polygram substitution cipher)

- 单表代替密码: 将明文中的每个字母用密文中对应的字母取代, 在全部信息加密过程中, 明文字母与密文字母一一映射.
- 同音代替密码: 明文字符与密文字符是一对多映射, 每个明文字符可以变换成不同的密文字符.
- 多表代替密码: 用多个映射把明文字符转换为密文字符, 其中每个映射相当于单表代替密码中的一一映射.
- 多字母代替密码: 是最一般的, 可以对一组字符进行任何方式的代替.

3.1 单表代替密码

构造一个密文字母表, 然后用密文字母表中的字母或字母表中来代替明文字母表的字母或字母组, 各字母或字母组的相对位置不变, 但其本身改变了, 这样的编成的密码称为单表代替密码.

设 $A = a_0, a_1, a_2, ..., a_{n-1}$ 为含有 n 个字母的明文字母表,

$B = b_0, b_1, b_2, ..., b_{n-1}$ 是含有 n 个字母的密文字母表
定义一个由 A 到 B 的映射:

$$f : A \to B, f(a_i) = b_i$$

若明文为: $M = (m_0, m_1, m_2, ..., m_{c-1})$,
则相应的密文为: $C = (f(m_0), f(m_1), f(m_2), ..., f(m_{c-1}))$.
可见, 单表代替密码的密钥就是映射 f 或密文字母表 B.

3.1.1 几种典型的单表代替密码

- 加法密码定义映射

$$f(a_i) = a_j, j \equiv i + k(\mathrm{mod}n)0 < k < n$$

加密算法实际上是每一个字母向前推移 k 位, 不同的 k 可得到不同的密文, 若令 26 个字母分别对应于整数 025, 如下表

明文	a	b	c	d	e	f	g	h	i	j	k	l	m
数字	0	1	2	3	4	5	6	7	8	9	10	11	12
明文	n	o	p	q	r	s	t	u	v	w	x	y	z
数字	13	14	15	16	17	18	19	20	21	22	23	24	25

则加法密码辩护那实际是

$$c \equiv m + k(\mathrm{mod}26), 0 < k < 26$$

m 是明文对应的明文数据, c 是密文对应的密文数据, k 是加密用的参数, 也称为密钥

例题明文为 data security 解对应的数据序列为 3 0 19 0 18 4 2 20 17 8 19 24 k=5 时的密文数据序列为 8 5 24 5 23 9 7 25 22 13 24 3 对应的

密文为 I F Y F X J H Z W N Y D 密文数据的计算如下

$$c_1 = m_1 + 5(\mathrm{mod}26) = 8(\mathrm{mod}26) = 8$$
$$c_2 = m_2 + 5(\mathrm{mod}26) = 5(\mathrm{mod}26) = 5$$
$$c_3 = m_3 + 5(\mathrm{mod}26) = 24(\mathrm{mod}26) = 24$$
$$c_4 = m_4 + 5(\mathrm{mod}26) = 5(\mathrm{mod}26) = 5$$
$$c_5 = m_5 + 5(\mathrm{mod}26) = 23(\mathrm{mod}26) = 23$$
$$c_6 = m_6 + 5(\mathrm{mod}26) = 9(\mathrm{mod}26) = 9$$
$$c_7 = m_7 + 5(\mathrm{mod}26) = 7(\mathrm{mod}26) = 7$$
$$c_8 = m_8 + 5(\mathrm{mod}26) = 25(\mathrm{mod}26) = 25$$
$$c_9 = m_9 + 5(\mathrm{mod}26) = 22(\mathrm{mod}26) = 22$$
$$c_{10} = m_{10} + 5(\mathrm{mod}26) = 13(\mathrm{mod}26) = 13$$
$$c_{11} = m_{11} + 5(\mathrm{mod}26) = 24(\mathrm{mod}26) = 24$$
$$c_{12} = m_{12} + 5(\mathrm{mod}26) = 29(\mathrm{mod}26) = 3$$

- 乘法密码映射

$$f(a_i) = a_j, j \equiv ik(\mathrm{mod}n)(k, n) = 1$$

若令 26 个字母与整数 0 25 对应, 则乘法密码变换实际是

$$c \equiv (\mathrm{mod}26)0 < k < 26$$

m 是明文对应的明文数据, c 是密文对应的密文数据, k 是加密用的参数, 也称为密钥

例子明文为 data security 解对应的数据序列为 3 0 19 0 18 4 2 20 17 8 19 24 k=5 时对应的密文数据序列 15 0 17 12 20 10 22 7 14 17 16 对应的密文为 P A R A M U K W H O R Q 密文数据的计算如下:

$$c_1 = 3 \times 5 = 15(\mathrm{mod}26) = 15$$
$$c_2 = 0 \times 5 = 15(\mathrm{mod}26) = 0$$
$$c_3 = 19 \times 5 = 95(\mathrm{mod}26) = 17$$
$$c_4 = 0 \times 5 = 0(\mathrm{mod}26) = 0$$
$$c_5 = 18 \times 5 = 15(\mathrm{mod}26) = 15$$
$$c_6 = 4 \times 5 = 20(\mathrm{mod}26) = 20$$
$$c_7 = 2 \times 5 = 10(\mathrm{mod}26) = 10$$
$$c_8 = 20 \times 5 = 100(\mathrm{mod}26) = 22$$
$$c_9 = 17 \times 5 = 85(\mathrm{mod}26) = 7$$
$$c_10 = 8 \times 5 = 40(\mathrm{mod}26) = 14$$
$$c_11 = 19 \times 5 = 95(\mathrm{mod}26) = 17$$
$$c_12 = 24 \times 5 = 120(\mathrm{mod}26) = 16$$

- 仿射密码乘法密码与加法密码相结合便构成仿射密码

  映射 $f(a_i)a_j, j \equiv ik_1 + k_0(\mathrm{mod}n)(k,n) = 1, 0 < k_0 < n$

  若令 26 个字母与整数 0 25 相对应, 则仿射密码变换实际是

  $$c \equiv mk_1 + k_0(\mathrm{mod}26)(k_1, 26) = 10 < k_0 < 26$$

  , 其中 $(k_1, 26) = 1$ 表示 k_1 与 26 的最大公约数为 1

  m 是明文对应的明文数据, c 是密文对应的密文数据, k_1, k_0 是加密的参数, 也称为密钥.

  例如明文为 data security

  解对应的数据序列为 3 0 19 0 18 4 2 20 17 8 19 24

  $\left.\begin{array}{l} k_0 = 3 \\ k_1 = 5 \end{array}\right\}$ 时的密文数据序列为 18 3 20 3 15 2313 25 10 17 20 19 对应的密文为 S D U D P X N Z K R U T

密文数据计算如下

$c_1 = m_1 \times k_1 + k_0(\mathrm{mod}26) = 3 \times 5 + 3(\mathrm{mod}26) = 18(\mathrm{mod}26) = 18$

$c_2 = m_2 \times k_1 + k_0(\mathrm{mod}26) = 0 \times 5 + 3(\mathrm{mod}26) = 3(\mathrm{mod}26) = 3$

$c_3 = m_3 \times k_1 + k_0(\mathrm{mod}26) = 19 \times 5 + 3(\mathrm{mod}26) = 98(\mathrm{mod}26) = 20$

$c_4 = m_4 \times k_1 + k_0(\mathrm{mod}26) = 0 \times 5 + 3(\mathrm{mod}26) = 3(\mathrm{mod}26) = 3$

$c_5 = m_5 \times k_1 + k_0(\mathrm{mod}26) = 18 \times 5 + 3(\mathrm{mod}26) = 93(\mathrm{mod}26)15$

$c_6 = m_6 \times k_1 + k_0(\mathrm{mod}26) = 4 \times 5 + 3(\mathrm{mod}26) = 23(\mathrm{mod}26) = 23$

$c_7 = m_7 \times k_1 + k_0(\mathrm{mod}26) = 2 \times 5 + 3(\mathrm{mod}26) = 13(\mathrm{mod}26) = 13$

$c_8 = m_8 \times k_1 + k_0(\mathrm{mod}26) = 20 \times 5 + 3(\mathrm{mod}26) = 103(\mathrm{mod}26) = 25$

$c_9 = m_9 \times k_1 + k_0(\mathrm{mod}26) = 17 \times 5 + 3(\mathrm{mod}26) = 88(\mathrm{mod}26) = 10$

$c_10 = m_10 \times k_1 + k_0(\mathrm{mod}26) = 8 \times 5 + 3(\mathrm{mod}26) = 43(\mathrm{mod}26) = 17$

$c_11 = m_11 \times k_1 + k_0(\mathrm{mod}26) = 19 \times 5 + 3(\mathrm{mod}26) = 98(\mathrm{mod}26) = 20$

$c_12 = m_1 \times k_1 + k_0(\mathrm{mod}26) = 24 \times 5 + 3(\mathrm{mod}26) = 123(\mathrm{mod}26) = 19$

- 多项式密码映射

$$f(a_i) = a_j, j \equiv i^t k_t + i^{t-1} k_{t-1} + \cdots + i k_1 + k_0(\mathrm{mod}n)$$

其中 $(k_i, n) = 1, i = 1, \cdots, t, 0 < k_0 < n$.

若令 26 个字母与整数 025 对应, 则多项式密码变换实际是

$$c \equiv m^t k_t + m^{t-1} k_{t-1} + \cdots + k_1 m + k_0(\mathrm{mod}26)$$

$(k_i, 26) = 1, i = 1, 2, \cdots, t, 0 < k_0 < 26$

m 是明文对应的明文数据, c 是与密文对应的密文数据, $k_t, \cdots, k_1, k_0$ 是加密用的参数, 也称为密钥.

- 密钥词组代替密码用一个词组或者短语作密钥, 去掉密钥中的重复字母, 把结果作为矩阵的第一行, 其次从明文字母表中补入其余字母, 最后按某一顺序从矩阵中取出字母构成密文字母表.

如, 设密钥词组为 red star, 明文字母表和密文字母表均由 26 个英文字母构成, 则构成矩阵如下:

r	e	d	s	t	a
b	c	f	g	h	i
j	k	l	m	n	o
p	q	u	v	w	x
y	z				

若选出顺序按列, 则得到明文字母和密文字母对应如下:

明文字母	a	b	c	d	e	f	g	h	i	j	k	l	m
密文字母	R	B	J	P	Y	E	C	K	Q	Z	D	F	L
明文字母	n	o	p	q	r	s	t	u	v	w	x	y	z
密文字母	U	S	G	M	V	T	H	N	W	A	I	O	X

例如若明文 data security

则对应的密文为 PRHR TYJNVQHO

- 单表代替密码的统计分析
  - 加法密码和乘法密码的密钥量比较小, 可利用穷举密钥的方法进行译.
  - 仿射密码和多项式密码的密钥量可以百、千位计, 可利用计算机进行穷举密钥的方法来破译
  - 单表代替密码的密文字母表实质上就是明文字母表的一种排列. 对以英文字母作为明文的情况, 密文字母表可能的排列 $26! \approx 4 \times 10^{26}$, 所以即使使用计算机穷举密钥的方法破译也是不可能的
  - 但是, 穷举密钥的方法不是攻击密码的唯一方法. 因为, 一旦信息足够长, 密码分析者便可利用统计分析的方法进行攻击.
  - 任何自然语言都有许多固有的统计特性, 如果明文语言的这种统计特性在密文中有所反映, 则密码分析便可以通过分析明文和密文的统计规律破译密码
  - 当所统计的文献篇幅足够长, 且不特别专业化, 可发现各个英文字母出现的相对频率十分稳定.
    - 极高频率字母组: e
    - 次高频率字母组: t a o n i r s h
    - 中等频率字母组: d l u c m

* 低频率字母组: p f y w g b v
* 甚低频率字母组: j k q x z

单个英文字母出现频率的顺序为:

序号	1	2	3	4	5	6	7	8	9	10	11	12	13
字母	e	t	a	o	n	r	i	s	h	d	l	f	c
序号	14	15	16	17	18	19	20	21	22	23	24	25	26
字母	m	u	g	p	y	w	b	v	k	x	j	q	z

分析过程如下:

1. 统计密文中单字母频数
2. 按出现的频率分组 (密文和明文都要统计: 1 8 5 7 5)
3. 分析字母组
   - 单字母单词
   - 双字母单词
   - 三字母单词
   - 四字母单词
4. 判断并翻译

- 同音代替密码
  - 在同音代替密码中, 一个明文字母表中的字母 a, 可以变换为若干个密文字母 $f(a)$, 称为同音字母. 从明文到密文的映射的形式是 $f : A \to 2^C$, 其中 A 和 C 分别是明文字母表和密文字母表.

    例如假定明文同音代替密码的密钥是一段短文

    Canada's large land mass and scattered populations makes efficient communication a neccessity. Extensive railway, road and other traspotation systems, as well as telephone, telegraph, and cable networks, have helped to link communities and have played a vital part in the country's development.

    将该短文机器中的各个单词编号如下:

1	2	3	4	5	6	7	
Canada's	large	land	mass	and	scattered	populations	
9	10	11	12	13	14	15	
efficient	communication	a	neccessity	extensive	railway	road	
17	18	19	20	21	22	23	
other	traspotation	systems	as	well	as	telephone	
25	26	27	28	29	30	31	
and	cable	networks	have	helped	to	link	co
33	34	35	36	37	38	39	
and	have	played	a	vital	part	in	
41	42						
country	development						

表中每个单词的首字母对应一个数字, 加密时, 可以用与某个字母相对应的任何一个数字来代替该字母

3.2 多表代替密码

- 单表代替密码容易被攻破, 明文中的一个字母与密文中的一个字母一一对应, 明文中字母的统计特性在密文中可以反映出来.
- 多表代替密码采用多个密文字母表 (提高代替密码强度), 使明文中的每一个字母都有多种可能的代替.

  构造由 d 个字符组成的密文字母表

$$B_j = (b_{j0}, b_{j1}, b_{j2}, \cdots, b_{jn-1}), j = 0, 1, \cdots, d-1$$

  定义 d 个映射

$$f_j : A \to B, f_j(a_i) = b_{ji}$$

  设明文

$$M = (m_0, m_1, \cdots, m_{d-1}, m_d, \cdots)$$

  则相应的密文:

$$C = [f_0(m_0), f_1(m_1), f_2(m_2), \cdots, f_{d-1}(m_{d-1}), f_d(m_d)]$$

.

- 多表代替密码的密钥就是 d 个映射或者密文字母表
- 多表代替密码的种类有很多, 最著名的多表代替密码是 16 世纪的法国密码学者 Vigenere 使用过的 Vigenere 密码. 此外, 还有游动钥密码 (running-key cipher)
- 弗吉尼亚密码简介

1. 设密钥 $K = k_1k_2\cdots k_n$, 明文 $M = m_1m_2\cdots m_n$, 加密变换为

$$E_k(M) = c_1c_2\cdots c_n$$

其中 $c_i \equiv (m_i + k_i) \bmod 26, i = 1, 2, \cdots, n$.
维吉尼亚密码的密钥可以周期性延长, 周而复始, 以至无穷. 即令

$$K = k_1k_2k_3\cdots$$

维吉尼亚方阵可以用来加密和脱密
例如 $M = datasecurity, k = best$, 试着写出对应的密文

1 将 M 分解成长度为 4 的序列
data secu rity

2 加密过程, 以密钥字母 $k_j(j = 0, 1, 2, 3)$ 为行号, 明文字母 $m_i(i = 0, 1, 2, \cdots, 11)$, 在维吉尼亚方阵中依次得到对应密文字母.

$$c = E_k(M) = EELTTIUNSMLR$$

即维吉尼亚方阵中, b 行 d 列的 E 为密文的第一个字母, e 行 a 列的 E 唯密文的第二个字母, s 行 t 列为密文的第三个字母, $\cdots$. 或者利用公式 $c_i = (m_i + k_i) \bmod 26, i = 1, 2, \cdots, n$.

$$c_1 = (m_1 + k_1) \bmod 26 = (1 + 3) \bmod 26 = 4 = E$$
$$c_2 = (m_2 + k_2) \bmod 26 = (0 + 4) \bmod 26 = 4 = E$$
$$\cdots\cdots$$
$$c_{10} = (m_{10} + k_{10}) \bmod 26 = (1 + 3) \bmod 26 = 4 = E$$
$$c_{11} = (m_{11} + k_{11}) \bmod 26 = (8 + 4) \bmod 26 = 12 = M$$
$$c_{12} = (m_{12} + k_{12}) \bmod 26 = (24 + 19) \bmod 26 = 17 = R$$

3 脱密过程: 以密钥字母 $k_j(j=0,1,2,3)$ 为行号, 在方阵中该行的密文字母 $c_i(i=0,1,2,\cdots,11)$ 所在的列的列号即为明文字母.

$$m_1=(c_1-k_1) \bmod 26=(4-1) \bmod 26=3=D$$
$$m_2=(c_2-k_2) \bmod 26=(4-4) \bmod 26=0=A$$
$$\cdots\cdots$$
$$m_{10}=(c_{10}-k_{10}) \bmod 26=(4-1) \bmod 26=3=E$$
$$m_{11}=(c_{11}-k_{11}) \bmod 26=(12-8) \bmod 26=4=M$$
$$m_{12}=(c_{12}-k_{12}) \bmod 26=(17-24) \bmod 26=19=R$$

2. 游动钥密码

– 游动钥密码是一种非周期性的 Vigenere 密码, 其密钥和明文信息一样长, 且不重复.
例如设 $m=6$, 密钥字为 cipher, 密钥 k 相对应的数字为 $k=(2,8,15,7,4,1,7)$, 设明文是字符串 this cryptosystem, 其加密过程如下

明文数据	19	7	8	18	2	17	24	15	19	14	18	24	18	19	4	12	
密钥数据	2	8	15	7	4	1	17	2	8	15	7	4	17	2	8	15	7
密文数据	21	15	23	25	6	8	0	23	8	21	22	15	20	1	19	19	
密文字符串	V	P	X	Z	G	I	A	X	I	V	W	P	U	B	T	T	

3. 对弗吉尼亚密码的分析

– 对弗吉尼亚密码进行分析, 首先要确定密钥的长度, 记作 m; 其次需要确定密钥字符串. 一般用到两个技术: 一是 Kaisiski 测试, 二是重合指数验证

∗ Kaisiski 测试: 在密文中, 搜索长度至少为 2 的相同一对密文, 记下这两段开始点的距离; 如果可以获得几个这样距离 $d_1,d_2,\cdots$, 则可以假设 m 是整出这些 d_i 的最大公因数
—Kaisiski 测试的基本出发点: 即两个相同的明文段将加密成相同的密文段, 如果在密文中观察到两个相同的长度至少为 3 的密文段, 则它们对应相同的明文串, 可以作为对该密文进行攻击的切入点.

* 重合指数验证 (所获得的 m 值)
  - 重合指数: 假设 $C = c_1c_2c_3\cdots c_n$ 是 n 个字符的串, C 重合指数记为 $IC(C)$, 定义为 C 中两个随即元素相同的概率.

    假设用 $n_A, n_B, \cdots, n_Z$ 分别表示为 $A, B, \cdots, Z$ 在 C 中出现的频数; 以 C_n^2 种方式选择 C 中的两个元素, 其中 $C_{n_A}^2$ 种方式选择两个元素都是 A, $C_{n_B}^2$ 种方式选择两个元素都是 B, $\cdots$, $C_{n_Z}^2$ 种方式选择两个元素都是 Z.

    则有

    $$IC(C) = \frac{\sum_{\xi=A}^{Z} n_\xi(n_\xi - 1)}{n(n-1)} \tag{2}$$

    其中 $n_A, n_B, n_C, \cdots, n_Z$ 分别表示 $A, B, C, \cdots, Z$ 在 C 中出现的频数

    随机英文字母序列 Y 的重合指数: 在 Y 中两个随机元素出现相同字母的概率为 $\frac{1}{26}$, 即:

    $$IC(Y) = \frac{1}{26} = 0.0385$$

    文献的 X 的重合指数: 设英文论文为 X, 其中两个元素都是 a 的概率为 $(0.0856)^2 = 0.0073274$, 两个元素都是 b 的概率为 $(0.0139)^2 = 0.001932, \cdots$, 两个元素都是 z 的概率为 $(0.0008)^2 = 0.00000064$. 若用 p_1 表示 X 中 a 出现的概率, 用 p_2 表示 X 中出现 b 的概率, $\cdots$, 用 p_{26} 表示 X 中出现 z 的概率, 则

    $$IC(X) = \sum_{i=1}^{26} p_i^2 = 0.0687$$

    重合指数验证方法: 设密文 C 为用 Vigenere 密码加密的得到的密文串 $C = c_1c_2c_3\cdots c_n$. 假定使用的密钥词组长度为 m, 使用的密钥词组长度为 m, 即 $K = k_1k_2k_3\cdots k_m$, 而且各位由不同的字母组成.
  - 将密文 C 分成 m 行, 则每一行是单表代替密码, 不同的行由不同的密钥加密
  - 如果 m 是实际密钥的长度, 则同一行两个选定位置上有相同字母的概率为 0.0687

· 如果 m 不是实际密钥的长度, 则同一行两个选定位置上有相同字母的概率为 0.0385
例如从 Vigenere 密码中获得密文如下:
UFQUIUDDWFHGLZARIHWLLWYYFSYYQATJP FK-MUXSSWWCSVFAEVWWGQCMVVSWFKUtBLLG ZFVI-TYOEIPASJWGGSJEPNSUETPTMPOPHZSF DCX-EPLZQWKDWFXWTHASPWIUOVSSSFKWWL CCEZWEUE-HGVGLRLLGWOFKWLUWSHEVWSTT UARCWH-WBVTGNITJRWWKCOTFGMILRQESKW VMPFZMVEG-EEPFODJQCHZIUZZMXKZBGJOTZ AXCCMUMRSSJW
[1] 统计得, 字符数共 280 个, 其中每个字母出现的频率如下表

A	B	C	D	E	F	G	H	I	J	K	L	M
9	4	10	7	14	15	14	10	11	7	9	13	10
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3	7	12	9	8	20	12	14	12	27	5	6	12

此时, $IC = 0.0431$
[2] 作 Kasiski 试验, 列出重复的字符所在的位置及其间距:

3.3 代数密码 (Vernam 加密算法)

- 代数密码的加密方式如下:

$$M = m_1 m_2 m_3 \cdots, m_i = 0(or1), i = 1, 2, \cdots$$
$$K = k_1 k_2 k_3 \cdots, k_j = 0(or1), j = 1, 2, \cdots$$

即明文和密钥均用二元数组序列 (二进制) 表示

加密过程如下:

$$c_i = m_i \oplus k_i, i = 0, 1, \cdots$$

即 $c_i \equiv (m_i + k_i) \bmod 2, i = 1, 2, 3 \cdots$

- 编制代数密码

需要先把明文和密钥表示成二元序列, 再把它们按位模 2 相加即可.

- 把 M 和 K 转化为二元序列, 可按如下表所示 (也可以按照其他方式转化).

a	00000	j	01001	s	10010
b	00001	k	01010	t	10011
c	00010	l	01011	u	10111
d	00011	m	01100	v	10101
e	00100	n	01101	w	10110
f	00101	o	01110	x	10111
g	00110	p	01111	y	11000
h	00111	q	10000	z	11001
i	01000	r	10001		

  解密过程为:

$$m_i = c_i \oplus k_i i = 0, 1, 2 \cdots$$

- 可见, 代数密码的加密和解密非常简单, 特别适合计算机和通信系统的应用

  例如明文 cat, 密钥 key, 用代数密码加密.

  解:

$$M = 000100000010011$$

$$K = 0101000100011000$$

  则 $C = M \oplus K = 011000010001011$

  密文为 IEL.

- 代数密码属于序列密码, 其突出优点是加密变换和解密变换相同, 故使加密和解密软硬件实现极为简单, 加密和解密可共用一个软件模块或硬件电路.

- 若同用一个密钥重复使用或密钥本身包含重复, 则代数密码经不起" 已知明文的攻击". 为增强代数密码的强度, 应该避免密钥重复. 可采用以下方式:

  – 密钥是真正的随机序列.

  – 密钥的长度大于或等于明文的长度.

  – 一个密钥只用一次.

3.4 Hill 密码

- 基本思想: 将 l 个明文字母通过线性变换把它们转换为 l 个密文字母. 解密时, 只要做一次逆变换即可. 在该算法中, 密钥即是变换举证

(1) Hill 加密算法设明文为: $M = m_1 m_2 \cdots m_l$, 密钥为 $K = (k_{ij})_{l \times l}$

$$C = KM(\mathrm{mod} n)$$

式中:

$$C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_l \end{bmatrix}, M = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_l \end{bmatrix}, K = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1l} \\ 0 & k_{22} & \cdots & k_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_{ll} \end{bmatrix} \tag{3}$$

即加密变换中:

$$E_k(M) = c_1 c_2 \cdots c_l = C$$

其中

$$\begin{cases} c_1 = k_{11} m_1 + k_{12} m_2 + k_{13} m_3 + \cdots + k_{1l} m_l (\mathrm{mod} n) \\ c_2 = k_{21} m_1 + k_{22} m_2 + k_{23} m_3 + \cdots + k_{2l} m_l (\mathrm{mod} n) \\ \cdots\cdots \\ c_3 = k_{l1} m_1 + k_{l2} m_2 + k_{l3} m_3 + \cdots + k_{ll} m_l (\mathrm{mod} n) \end{cases} \tag{4}$$

一般取 $n = 26$

例如明文为 hill, 采用 $a \sim z$ 的 26 字母与数字 $0 \sim 25$ 对应. 若取 $l = 4, n = 26$.

密钥 K 为:

$$K = \begin{bmatrix} 8 & 6 & 9 & 5 \\ 6 & 9 & 5 & 10 \\ 5 & 8 & 4 & 9 \\ 10 & 6 & 11 & 4 \end{bmatrix}$$

解: 明文数字序列为 7 8 11 11, 即 $M = (781111)$

于是有

$$\begin{aligned}
c_1 &= k_{11}m_1 + k_{12}m_2 + k_{13}m_3 + k_{14}m_4 \\
&= 8\times 7 + 6\times 8 + 9\times 11 + 5\times 11 = 258 \equiv 24(\mathrm{mod}26) \\
c_2 &= k_{21}m_1 + k_{22}m_2 + k_{23}m_3 + k_{24}m_4 \\
&= 6\times 7 + 9\times 8 + 5\times 11 + 10\times 11 = 179 \equiv 19(\mathrm{mod}26) \\
c_3 &= k_{31}m_1 + k_{32}m_2 + k_{33}m_3 + k_{34}m_4 \\
&= 5\times 7 + 8\times 8 + 4\times 11 + 9\times 11 = 242 \equiv 8(\mathrm{mod}26) \\
c_4 &= k_{41}m_1 + k_{42}m_2 + k_{43}m_3 + k_{44}m_4 \\
&= 10\times 7 + 6\times 8 + 11\times 11 + 4\times 11 = 283 \equiv 23(\mathrm{mod}26)
\end{aligned} \tag{5}$$

所以有 C=(24 19 8 23) 故密文为 Y T I X

例如假设 Hill 密码加密使用密钥 $K = \begin{bmatrix} 6 & 7 \\ 3 & 8 \end{bmatrix}$, 试着对明文 good 加密.

解: 明文数字序列为 (6, 14, 14, 3), 把明文划分为两组 (6, 14) 和 (14, 3), 加密过程如下:

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = K\begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} 6 & 7 \\ 3 & 8 \end{bmatrix}\begin{bmatrix} 6 \\ 14 \end{bmatrix} = \begin{bmatrix} 134 \\ 130 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}(\mathrm{mod}26)$$

$$\begin{bmatrix} c_3 \\ c_4 \end{bmatrix} = K\begin{bmatrix} m_3 \\ m_4 \end{bmatrix} = \begin{bmatrix} 6 & 7 \\ 3 & 8 \end{bmatrix}\begin{bmatrix} 14 \\ 3 \end{bmatrix} = \begin{bmatrix} 105 \\ 66 \end{bmatrix} = \begin{bmatrix} 1 \\ 14 \end{bmatrix}(\mathrm{mod}26)$$

密文数据序列为 (4 0 1 14), 则密文为 EABO.

(2) Hill 解密算法

– 一般来说, Hill 密码能较好地抵抗频率分析. 采用" 唯密文分析" 很难攻破它, 但采用" 已知明文的攻击" 易被攻破.

Hill 解密算法为:

$$M = K^{-1}C(\mathrm{mod}n)$$

一般取 $n=26$, 如果 K 有逆矩阵, 解密才是可能的.

– 一个实矩阵 K 有逆元的条件是: 当且仅当其行列式非零.

– ∵Hill 解密是在 Z_{26} 中进行的

∴ 当且仅当 $gcd\{detK, 26\}=1$ 时, 矩阵 K 有模 26 的逆元, 记为 $(detK)^{-1}$.

注: 若用记号 Z_m 表示集合, 则 $Z_m=\{0,1,\cdots,|m|-1\}$.

– K^{-1} 的表达式为:

$$K^{-1}=(detK)^{-1}K^*$$

式中 K^* 称为 K 的伴随矩阵, 其 (i,j) 单元为 A_{ji}.

$$K^*=\begin{bmatrix} A_{11} & A_{21} & \cdots & A_{l1} \\ A_{12} & A_{22} & \cdots & A_{l2} \\ A_{13} & A_{23} & \cdots & A_{l3} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1l} & A_{2l} & \cdots & A_{ll} \end{bmatrix}, A_{ij}=(-1)^{i+j}detK_{ij} \quad (6)$$

$K_{ij}(1\le i\le l, 1\le j\le l)$ 是从矩阵 K 中删除第 i 行和第 j 列后得到的矩阵.

– 对于 2×2 矩阵求逆有

定理: 假设 $A=(a_{ij})_{2\times 2}$, 且 $detA$ 在 Z_{26} 中是可逆的, 则

$$A^{-1}=(detA)^{-1}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

例如设 $A=\begin{bmatrix} 6 & 7 \\ 3 & 8 \end{bmatrix}$, 求 A^{-1}

解: $detA=6\times 8-3\times 7=27\equiv 1(\mathrm{mod}26)$

$$\because 1\times 1\equiv 1(\mathrm{mod}26) \therefore (detA)^{-1}=1$$

$$\therefore \begin{aligned} A^{-1} &= \begin{bmatrix} 8 & -7 \\ -3 & 6 \end{bmatrix}(\mathrm{mod}26) \\ &= \begin{bmatrix} 8 & 19 \\ 23 & 6 \end{bmatrix} \end{aligned}$$

例如假设 Hill 密码加密使用密钥 $K=\begin{bmatrix}6&7\\3&8\end{bmatrix}$, 使用 Hill 密码对密文 EABO 解密

解: 密文数据序列为 (4, 0, 1, 14) $\because l=2,\therefore$ 把密文划分为两组 (4, 0) 和 (1, 14)

求出 K 的逆矩阵 $K^{-1}=\begin{bmatrix}8&19\\23&6\end{bmatrix}$, 则解密如下:

$$\begin{bmatrix}m_1\\m_2\end{bmatrix}=K^{-1}\begin{bmatrix}c_1\\c_2\end{bmatrix}=\begin{bmatrix}8&19\\23&6\end{bmatrix}\begin{bmatrix}4\\0\end{bmatrix}=\begin{bmatrix}32\\92\end{bmatrix}$$
$$=\begin{bmatrix}6\\14\end{bmatrix}(\mathrm{mod}26)$$

$$\begin{bmatrix}m_3\\m_4\end{bmatrix}=K^{-1}\begin{bmatrix}c_3\\c_4\end{bmatrix}=\begin{bmatrix}8&19\\23&6\end{bmatrix}\begin{bmatrix}1\\14\end{bmatrix}=\begin{bmatrix}274\\107\end{bmatrix}$$
$$=\begin{bmatrix}6\\14\end{bmatrix}(\mathrm{mod}26)$$

得到明文数字序列为 (6, 14, 14, 3), 明文为 good.

(3) 关于 Hill 密码的已知明文攻击

– 假定分析者已知正在使用的 l 值, 且掌握了至少 l 个不同的 l 元组, 即 $M_i=\begin{bmatrix}m_{1i}\\m_{2i}\\\vdots\\m_{li}\end{bmatrix}$ 和 $C_i=\begin{bmatrix}c_{1i}\\c_{2i}\\\vdots\\c_{li}\end{bmatrix}$

满足

$$C_i=E(M_i,K)=e_K(M_i),1\le i\le l$$

若定义两个 $l\times l$ 矩阵: $M=(m_{ij})_{l\times l},C=(c_{ij})_{l\times l}$, 则有矩阵方程

$$C\equiv KM(\mathrm{mod}26)$$

式中 K 是未知密钥 K. 若 M 是可逆的, 则能计算出

$$K\equiv CM^{-1}(\mathrm{mod}26)$$

从而破译改密码体制. 但是, 若 M 不可逆, 则必须再试另外 l 个—密文对

例如假设明文 worker 利用 $l = 2$ 的 Hill 密码加密, 得到密文为 QIHRYB, 求密钥.

解: 明文 worker 对应的数据序列为: (22, 14, 17, 10, 4, 17);

密文 QIHRYB 对应的数据序列 (16, 8, 7, 17, 24, 1).

因为 l=2, 故将明文、密文按顺序分成三组, 即分别对应如下:

$$\begin{pmatrix} 16 \\ 8 \end{pmatrix} = K \begin{pmatrix} 22 \\ 14 \end{pmatrix} (\mathrm{mod}26)$$

$$\begin{pmatrix} 7 \\ 17 \end{pmatrix} = K \begin{pmatrix} 17 \\ 10 \end{pmatrix} (\mathrm{mod}26)$$

$$\begin{pmatrix} 24 \\ 1 \end{pmatrix} = K \begin{pmatrix} 4 \\ 17 \end{pmatrix} (\mathrm{mod}26)$$

从前面两个明文—密文对可以得到:

$$\begin{bmatrix} 16 & 7 \\ 8 & 17 \end{bmatrix} = K \begin{bmatrix} 22 & 17 \\ 14 & 10 \end{bmatrix} (\mathrm{mod}26)$$

$\because \begin{bmatrix} 22 & 17 \\ 14 & 10 \end{bmatrix} = -18, (-18, 26) \neq 1 \therefore$ 重新考虑第二、第三组明文—密文对, 此时的到如下矩阵方程:

$$\begin{bmatrix} 7 & 24 \\ 17 & 1 \end{bmatrix} = K \begin{bmatrix} 17 & 4 \\ 10 & 17 \end{bmatrix}$$

$\because \begin{vmatrix} 17 & 4 \\ 10 & 17 \end{vmatrix} = 249, 249 \times 7 \equiv 1(\mathrm{mod}26) \therefore (detA)^{-1} = 7$ 容易算出

$$\begin{bmatrix} 17 & -4 \\ -10 & 17 \end{bmatrix} = 7 \times \begin{bmatrix} 17 & -4 \\ -10 & 17 \end{bmatrix} = \begin{bmatrix} 119 & 28 \\ -70 & 119 \end{bmatrix}$$

$$\equiv \begin{bmatrix} 15 & 24 \\ 8 & 15 \end{bmatrix} (\mathrm{mod}26)$$

$$\therefore \quad K = \begin{bmatrix} 7 & 24 \\ 17 & 1 \end{bmatrix} \begin{bmatrix} 15 & 24 \\ 8 & 15 \end{bmatrix} = \begin{bmatrix} 297 & 528 \\ 263 & 423 \end{bmatrix}$$

$$\equiv \begin{bmatrix} 11 & 8 \\ 3 & 7 \end{bmatrix} (\mathrm{mod} 26)$$

所得密钥 K, 可以通过已知的明文—密文对进行验证.