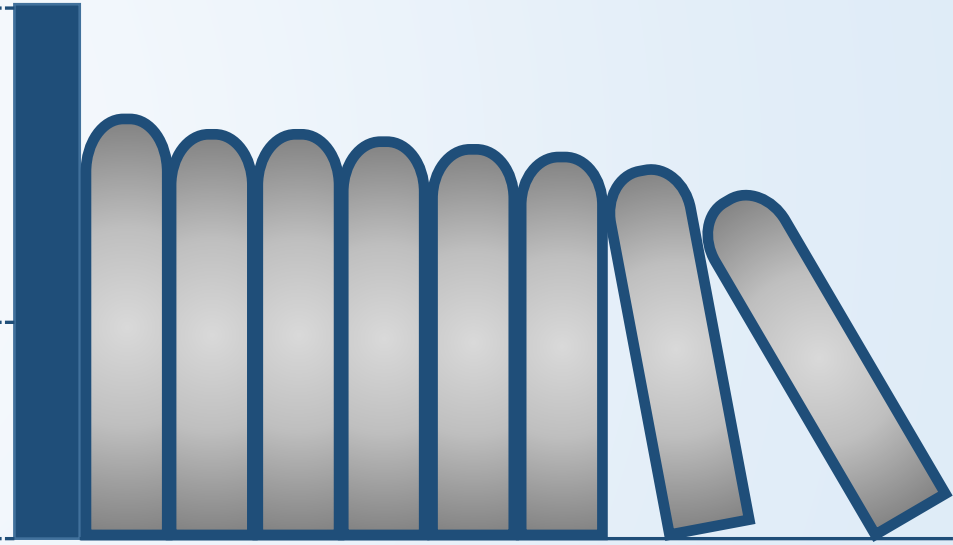
哈爾濱工業大學(深圳)

HARBIN INSTITUTE OF TECHNOLOGY, SHENZHEN

密码学基础实验课程

主讲教师：蒋琳

实验教师：苏婷



实验课程安排与考核标准

实验课程共16个学时，7个实验项目，总成绩为30分（30%）。

实验项目

项目编号	实验一	实验二	实验三	实验四	实验五	实验6
学时数	4	2	2	4	2	2
实验项目	古典密码算法实验 DES密码算法实验	AES密码算法实验	流密码算法实验	RSA算法实验	SHA-1/MD5算法实验	数字签名算法实验
分数数	6	4	3	5	4	5

考核方式

- 源代码和结果截图：每次课程均需提交实验程序源代码，以及程序的运行结果截图。
- 实验报告(3分)：最后一次课程需提交实验报告，参照提供的报告模板，于实验课结束一周内提交电子版实验报告。
- 附加题累加到总分中计算，1-2分不等。

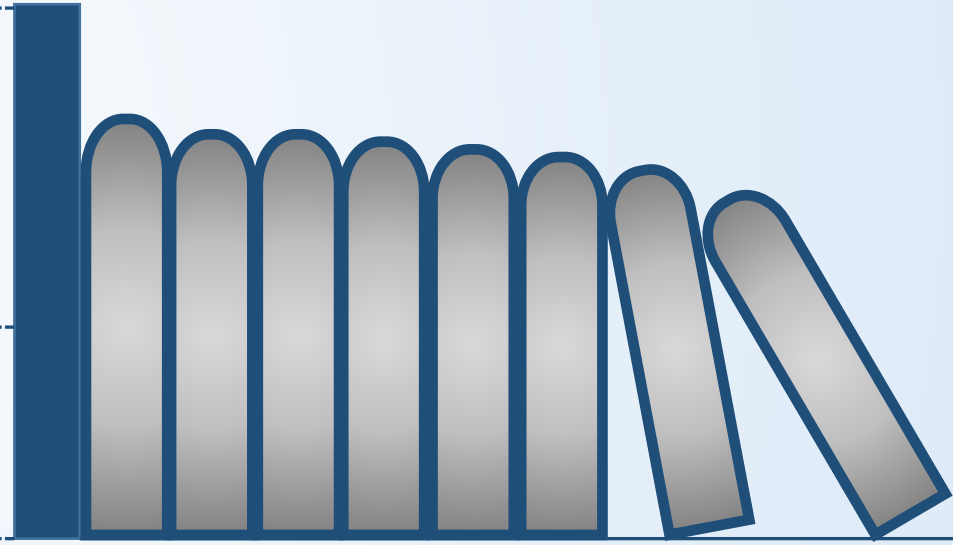
禁止抄袭，发现雷同，本次实验双方都是0分。

哈爾濱工業大學(深圳)
HARBIN INSTITUTE OF TECHNOLOGY, SHENZHEN

古典密码算法实验

主讲教师：蒋琳

实验教师：苏婷



2020年11月3日星期二

实验目的

- 理解明文、密文、加密密钥、解密密钥、加密算法、解密算法
- 掌握古典密码学代换的基本加密原理
- 能够实现仿射密码和维吉尼亚密码算法
- 掌握Kasiski测试法、重合指数测试法和交互重合指数测试法等密码分析方法

实验环境

- **密码算法实验系统**

  http://10.251.129.2/index/check.html

1、登录

Step 1

Step 2

Step 3

实验环境


首页
古典密码
序列密码
分组密码
公钥密码
摘要算法
认证算法
抗量子密码算法
个人中心
退出
当前用户
姓名：苏婷测试
编号：20188197s
专业：
班级：
管理菜单
用户管理
课程管理
选择课程
课程管理
3、确认选课内容和提交截止时间
2019年春季学期
序号
实验项目
实验目的
上传实验报告(支持pdf,doc,docx,zip,rar,7z格式，不超过20M,重复上传将被覆盖)
报告提交截止时间
1
DES密码算法实验
1．了解DES算法的安全参数。
2．掌握DES算法的密钥生成方法。
选择文件
未选择任何文件
提交
2019-04-26
2
Vigenere密码算法实验
1．理解Vigenere密码算法的原理。
2．掌握Vigenere密码算法加解密方
选择文件
未选择任何文件
提交
2019-04-08


首页
古典密码
序列密码
分组密码
公钥密码
摘要算法
认证算法
抗量子密码算法
个人中心
退出
Ceaser密码算法
置换密码
移位密码算法
Vigenere密码算法
仿射密码算法
4、可下载程序查看运行结果
置换密码
程序运行
置换密码算法是不改变明文字符,而是按照某一规则重新排列消息中的比特或字符顺序，从而实现明文信息的加密。将明文中的字母按照给定的顺序存放在一个矩阵中，然后用根据密钥提供的顺序重新组合矩阵中的字母，从而形成密文。其解密过程是根据密钥的字母数作为列数，将密文按照列、行的顺序写出，再根据密钥给出的矩阵置换产生新的矩阵，从而恢复明文。
算法原理
置换密码（Permutation Cipher），又称换位密码。算法实施时，明文的字母保持相同，但顺序会被打乱。

实验内容

- **编码**
  - **实现仿射密码算法的加解密方案；**
  - **实现维吉尼亚密码算法的加解密方案。**
- **分析**
  - **对仿射密码算法做暴力破解；**

附加题：利用Kasiski测试法和重合指数法对维吉尼亚密码算法进行统计分析，并破解。

实验原理

- **仿射密码加解密**

$$c = E_{a,b}(m) \equiv am + b \pmod{26}$$

$$m = D_{a,b}(c) \equiv a^{-1}(c - b) \pmod{26}$$

其中a,b是满足 $0 \le a, b \le 25$ 和 $\gcd(a, 26) = 1$ 的整数。 其中 $\gcd(a, 26)$ 表示a和26的最大公因子,表示a和26是互素的a^{-1}表示a的逆元，即 $a^{-1} \cdot a \equiv 1 \bmod 26$。

m为明文，c 为密文;

a,b 为加密密钥， a^{-1},b为解密密钥;

$E_{a,b}(m)$为加密算法，$D_{a,b}(c)$为解密算法。

实验原理

➢ 仿射密码分析

- **密钥空间**

  a与26互素，根据欧拉定理可知 $\varphi(26) = \varphi(2) * \varphi(13) = 12$**, a的取值空间为12；**

  b的偏移空间为26；

  那么密钥空间为：　　a*b = 12 * 26 = 312

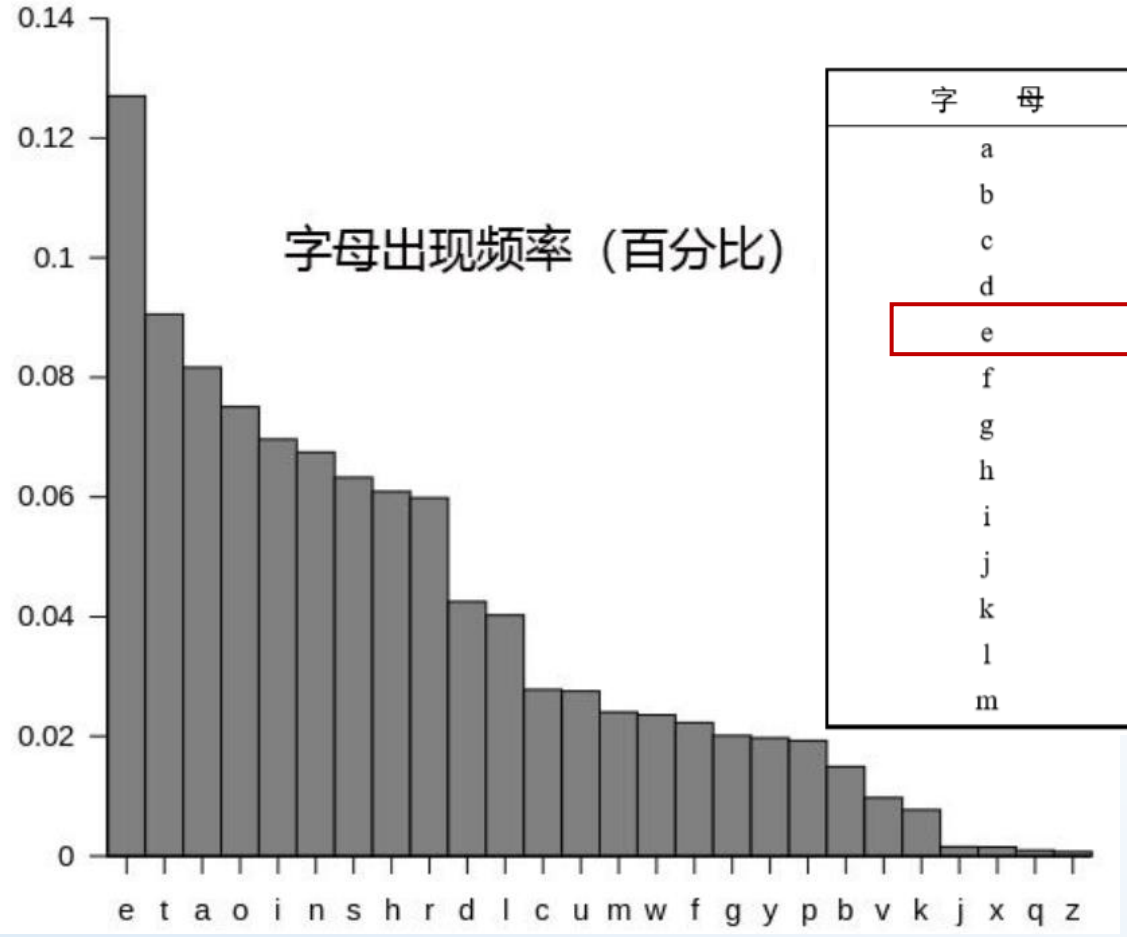
  密钥空间较小，可进行穷举密钥搜索的方法破解。

- **频率统计**

  因为明文空间和密文空间都为26个英文字母，根据英文字母的出现的频率统计规律，猜测密文对应的明文，可以提高破解的效率。

实验原理

➢ 英文字母出现频率



26 个英文字母的出现频率

字　母	频　率	字　母	频　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	0.090
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

注：在实际统计中可能会遇到统计密文的结果有几个字母的频率相同，需要进行验证筛选出正确的对应关系。

实验原理

➢ 英文字母出现频率

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

出现频率最高的30个双字母（按频率从大到小排列）

the	ing	and	her	ere	ent	tha	nth	was	eth
for	dth	hat	she	ion	int	his	sth	ers	ver

出现频率最高的20个三字母（按频率从大到小排列）

实验原理

➢ 欧几里得距离计算相似度

```
double fstd[26] = {0.082, 0.015, 0.028, 0.043,
0.127, 0.022, 0.020, 0.061, 0.070, 0.002,
0.008, 0.040, 0.024, 0.067, 0.075, 0.019,
0.001, 0.060, 0.063, 0.090, 0.028, 0.010,
0.023, 0.001, 0.020, 0.001};

double cal_dis(double* f)
{
        double sum = 0;
        for (int i = 0; i < 26; ++i)
        {
                sum += (f[i] - fstd[i]) * (f[i] - fstd[i]);
        }
        return sum;
}
```

自然语言频率

解密结果和自然语言频率欧几里得距离的平方

数值越小越相似

实验原理

➢ 维吉尼亚密码加解密

$$C_i = E_{ki}(m_i) = (m_i + k_i \bmod l) \bmod 26$$

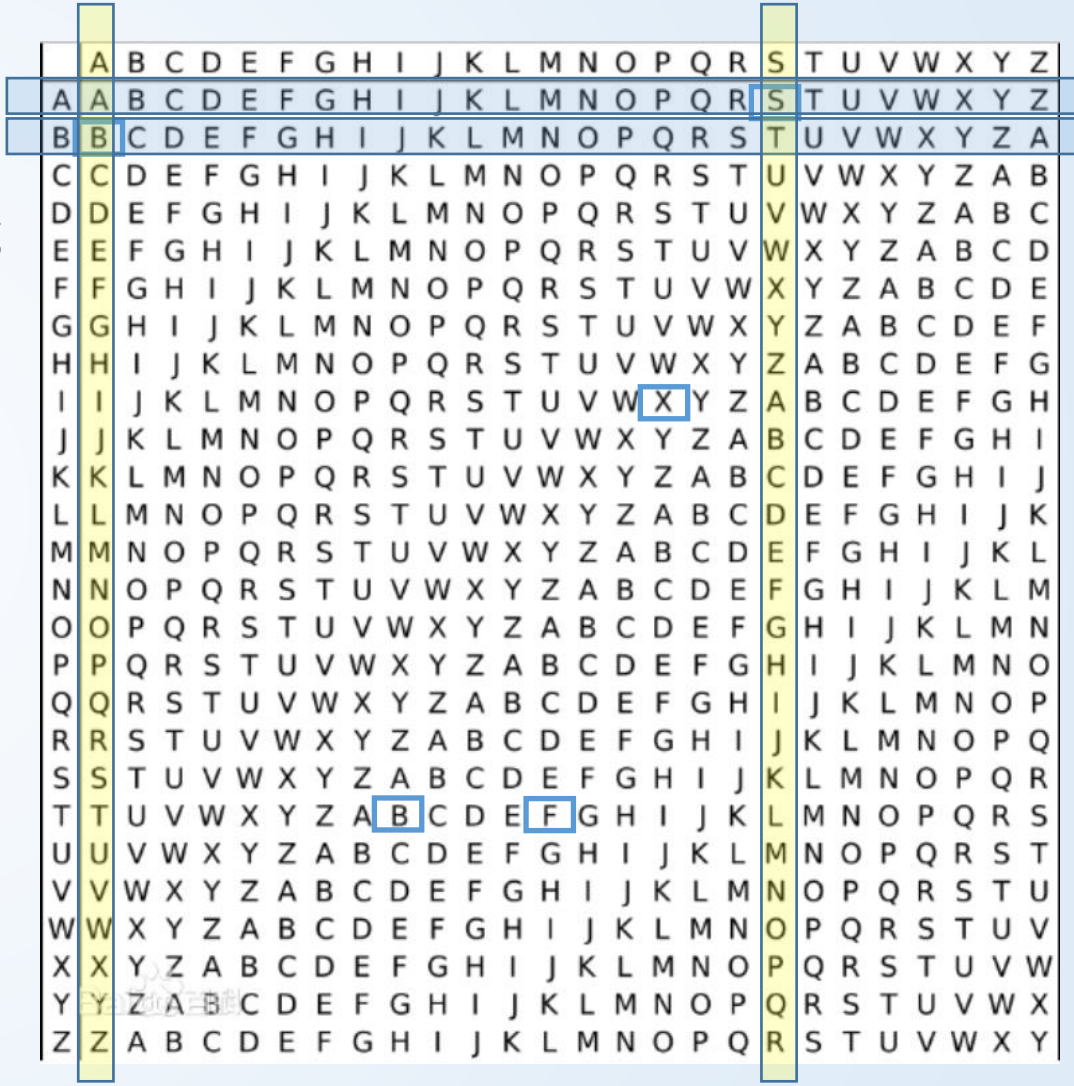
$$M_i = D_{ki}(c_i) = (C_i - k_i \bmod l) \bmod 26$$

m 为明文，c 为密文；
k为密钥，l为密钥长度；

$E_{ki}(m_i)$为加密算法，$D_{ki}(c_i)$为解密算法。

举个例子1：明文 a simple example
密钥 battista

Plaintext	a	s	i	m	p	l	e	e	x	a	m	p	l	e
Keystream	b	a	t	t	i	s	t	a	b	a	t	t	i	s
Ciphertext	B	S	B	F	X	D	X	E	Y	A	F	I	T	W



实验要求

仿射变换算法：从终端或者文件读取明文/密文，使用随机生成的一对密钥，实现加密和解密操作，并使用暴力破击方法破译一段信息；

1、待加密的明文如包含非大小写字母的情况，比如空格和标点符号，可以直接保留，或者删除。

2、密钥对随机生成，将生成的密钥对存放到文件affine_cipher.txt中的第一行，用于加解密。

3、从待加密/待解密文件（affine_data.txt/affine_cipher.txt）读取明文/密文，将加密和解密后的结果分别写入affine_cipher.txt和affine_decipher_result.txt文件中。也可以从终端读取或者显示在终端。

4、用暴力破解方式解密，待破译内容存放在文件A_ciphertext.txt中。将最终结果截图放在待提交的结果文件中。

实验要求

Vigenere算法：从文件读取明文/密文，密钥统一用security，完成加解密算法，并破译一段密文；

1、从待加密/待解密文件（vigenere_data.txt/ vigenere _cipher.txt）读取明文/密文，将加密和解密后的结果分别写入vigenere _data_result.txt和vigenere _cipher_result.txt文件中。

2、待加密的明文只考虑全部小写字母这一种情况，不考虑空格和标点符号。

3、加密的密钥统一用security 。

附加题（1分）

Kasisiki测试法和重合指数法的密码分析方法，待破译内容存放在文件V_ciphertext.txt中。将最终结果截图放在待提交的结果文件中。

考核要求

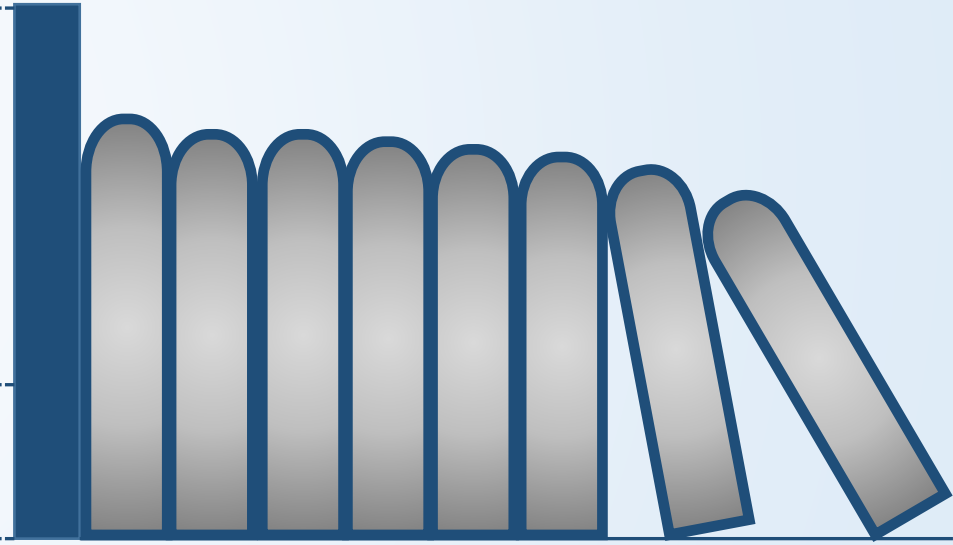
- 将仿射密码算法和维吉尼亚密码算法分别写在不同的程序文件中；
- 课堂检查仿射密码算法和维吉尼亚密码算法的加密解密部分；
- 仿密码破解列出前10个相似度最高的结果，并通过分析标注正确的解密结果；
- 请把测试结果截图及源代码打包成一个压缩包上传到系统中，命名格式如下：

  测试结果：“学号_姓名_实验1_仿射”

  压缩包：“学号_姓名_实验1_维吉尼亚”

- 提交要求：2020年11月10日24点 http://10.251.129.2/index/check.html ；

谢谢



2020年11月3日星期二

实验原理

➢ 维吉尼亚密码分析

- **密钥空间(假设密钥长度为n)**

  因为密钥是可以在26个字母中随意选取的，那么密钥空间为Z_{26}^{n}，也就是26^n。

  密钥空间较大，穷举密钥搜索的方法不适用。

- **频率统计**

  维吉尼亚密码相同的明文字母可能对应不同的密文字母，不同的明文字母也可能对应相同的密文字母，但是仍然可以采用英文字母统计频率来进行破解。

实验原理--维吉尼亚密码分析三步走

- 确定密钥的长度
- 确定密钥字的相对位移
- 穷举搜索密钥字

实验原理--维吉尼亚密码分析三步走

- **第一步确定密钥长度**
  - **Kasiski测试法**

原理：在明文中，如果两个相同的明文片段之间的距离*d*是密钥字长度*m*的倍数，那么这两个明文片段所对应的密文片段一定是相同的。反过来，如果密文中出现两个相同的密文片段（长度至少为3），那么它们对应的明文片段相同的可能性很大。

密钥： deceptivedeceptivedeceptive
明文： wearediscoveredsaveyourself
密文： ZICVTWQNGRZGVTWAVZHCQYGLMGJ
9

密钥 deceptive
密钥长度 9
密文间隔长度 9

实验原理--维吉尼亚密码分析三步走

- **第一步确定密钥长度**
  - **Kasiski测试法**

1、在密文中标出重复的3个或多个字符结构；

2、对每一个字符结构，记下结构的起始位置；

3、计算相邻的起始点的距离d_1,d_2,d_3……；

4、对每个距离求出所有因数m_1,m_2……，那么其中的一个因数就可能是密钥的长度；

举个例子3

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

- ✓ 其中CHR在密文中出现了5次，第1次出现到其他各次出现的距为：165，235，275， 285。
- ✓ gcd(165,235,275,285)=5
- ✓ 可猜测密钥长度为5

实验原理--维吉尼亚密码分析三步走

- 第一步确定密钥长度
  - 重合指数法

原理：自然语言（英语）的重合指数约为0.065，并且单表代换不会改变该值。

概念：设x=x1x2...xn是一个长度为n的英文字母串。则在x中的两个随机元素相同的概率，记为$I_c(x)$。f_i为26个字母中第i个字母在x中出现的次数，p_i为第i个英文字母出现的概率，那么可以得出

$$I_c(x)=\sum_{i=0}^{25}\frac{f_i(f_i-1)}{n(n-1)}\approx\sum_{i=0}^{25}p_i^2=0.065$$

实验原理--维吉尼亚密码分析三步走

举例说明：x=AADDEZDZ　n=8，

A出现的次数f_0 =2，D出现的次数 f_3 =3，E出现的次数 f_4=1，Z出现的次数 f_{25} =2。

那么一个长度为n的字符串中第一次出现A的概率就是$p_0 = \frac{f_0}{n}$，第二次就是$\frac{f_0-1}{n-1}$（因为找到第一次出现的A之后，再从剩下的字符串中找，长度为n-1，剩下A的次数只有f_0 -1），那么在这个字符串中随机选择两个元素是A的概率为$\frac{f_0}{n} * \frac{f_0-1}{n-1}$，当n足够大时，则 $\frac{f_0}{n} \approx \frac{f_0-1}{n-1}$。

由此得出下面的公式，其中p_i是字母在自然语言中统计出的频率

$$I_c(x)=\sum_{i=0}^{25}\frac{f_i(f_i-1)}{n(n-1)} \approx \sum_{i=0}^{25} p_i^2 = 0.065$$

实验原理--维吉尼亚密码分析三步走

- 第一步确定密钥长度
  - 重合指数法

如果猜测正确，则单表代换后的密文字符串重合指数值接近0.065；

否则，重合指数值在0.038~0.065之间。

以例3中的第一行的密文，猜测其密钥长度为5，则可分为5个单表代换的密文字符串

密文	C	H	R	E	E	V	O	A	H	M	A	E	R	A	T	B	I	A	X	X	W	T	N	X	B	E	E	O	P	H	B	S	B	Q	M	Q	E	Q	E	R	B	W
密钥	K_1	K_2	K_3	K_4	K_5	K_1	K_2	K_3	K_4	K_5	K_1	K_2	K_3	K_4	K_5	K_1	K_2	K_3	K_4	K_5	K_1	K_2	K_3	K_4	K_5	K_1	K_2	K_3	K_4	K_5	K_1	K_2	K_3	K_4	K_5	K_1	K_2	K_3	K_4	K_5	K_1	K_2
C1	C					V					A					B					W					E					B					Q					B	
C2		H					O					E					I					T					E					S					E					W
C3			R					A					R					A					N					O					B					Q				
C4				E					H					A					X					X					P					Q					E				
C5					E					M					T					X					B					H					M					R				

实验原理--维吉尼亚密码分析三步走

- 第一步确定密钥长度
  - 重合指数法

根据Kasiski测试法，我们猜测出例3密文的密钥长度为5，那么我们根据重合指数法来进行验证下。为了便于对比，我们分别计算长度为1,2,3,4,5这5组数据的重合指数值。

长度l	重合指数				
	C1	C2	C3	C4	C5
1	0.045				
2	0.046	0.041			
3	0.043	0.050	0.047		
4	0.042	0.039	0.046	0.040	
5	0.063	0.068	0.069	0.061	0.072

根据对比，可以确定密钥的长度为5。

实验原理--维吉尼亚密码分析三步走

- **第二步确定密钥字相对位移**
  - **交互重合指数**

定义：设$x=x_1x_2\ldots x_n$和$y=y_1y_2\ \ldots\ y_n$是两个长度分别为n和n'的字母串。*其*交互重合指数定义为x中的一个随机元素与y中的一个随机元素相同的概率,记为$MI_c(x,y)$，f_i和f_i'为字母A,B,C…在x和y中出现的次数。

$$MI_C(x,y)=\frac{\sum_{i=0}^{25}f_if'_i}{nn'}$$

实验原理--维吉尼亚密码分析三步走

举例说明：x=AADDEZDZ　y=XADFD，

A在x中出现的次数f_0 =2，在y中出现的次数$f_0^{'}$ =1 。

那么A在一个长度为n的字符串中的概率为$\frac{f_0}{n}$， 在一个长度为n′ 的字符串中的概率为$\frac{f_0^{'}}{n'}$，那么从这两个字符串中随机选择两个元素是A的概率为$\frac{f_0}{n} * \frac{f_0^{'}}{n'}$

由此得出下面的公式，

$$MI_C(x,y) = \frac{\sum_{i=0}^{25} f_i f'i}{nn'}$$

实验原理--维吉尼亚密码分析三步走

- 确定密钥字相对位移
  - 考虑不同密钥字加密后密文串的交互重合指数

$c = m + k_i \bmod 26$
$m = c - k_i \bmod 26$

设密钥为$k=k_1,k_2,\ldots,k_m$

c_i和c_j分别为密钥字k_i和k_j加密后的密文字串。从c_i和c_j来中随机抽取一个随机元素同为第h个字母的概率为$p_{(h-ki)}\,p_{(h-kj)}$，$0\le h\le 25$，$p_{(h-ki)}$也就是第h个字母在明文中出现的概率。

那么$MI_c(y_i,y_j)$的值可由下式计算，可以看出$MI_c(y_i,y_j)$的值取决于相对位移k_i-k_j：

$$MI_c(y_i,y_j)\approx\sum_{h=0}^{25}p_{h-ki}p_{h-kj}=\sum_{h=0}^{25}p_hp_{h+ki-kj}$$

实验原理--维吉尼亚密码分析三步走

- **第二步确定密钥字相对位移**

当相对位移等于0时，其实对应的就是自然语言的重合指数，也就是交互重合指数的估计值为0.065；

其他情况时的交互重合MIc指数的估计值为0.032～0.045；

表 2.15　交互重合指数 MI_c 的估计值

相对位移	MI_c 的估计值	相对位移	MI_c 的估计值
0	0.065	7（19）	0.039
1（25）	0.039	8（18）	0.034
2（24）	0.032	9（17）	0.034
3（23）	0.034	10（16）	0.038
4（22）	0.044	11（15）	0.045
5（21）	0.033	12（14）	0.039
6（20）	0.036	13	0.043

实验原理--维吉尼亚密码分析三步走

- **第二步确定密钥字相对位移**

猜测不同密钥间的相对位移l=k_i-k_j，如果$MI_C(y_i, y_j)$的值约为0.065，那么l的值猜测正确，意味着找到了不同密钥字加密的相同明文字母。

$$MI_C(y_i, y_j) = \frac{\sum_{i=0}^{25} f_{i,t} f'_{j,t-l}}{nn'}$$

说明：维吉尼亚这种加密算法密钥字之间的关系体现在了密文字之间。

同一个明文用不同密钥字加密的密文

$c_i = m + k_i \bmod 26$

$c_j = m + k_j \bmod 26$

↓

$c_j - c_i = k_j - k_i \bmod 26$

实验原理--维吉尼亚密码分析三步走

- **第二步确定密钥字相对位移**
  - **交互重合指数,以例3中的C1到C5为例,这里的y_i和y_j就是指Ci和Cj。**

C1

C2

C3

C4

C5

C_i C_j 表 2.16 密文子串的交互重合指数 MI_c 的估计值

i	j	$MI_c(y_i, y_j^g)$ 的值，$0 \leqslant g \leqslant 25$								
1	2	0.028	0.027	0.028	0.034	0.039	0.037	0.026	0.025	0.052
		0.068	0.044	0.026	0.037	0.043	0.037	0.043	0.037	0.028
		0.041	0.041	0.034	0.037	0.051	0.045	0.042	0.036	
1	3	0.039	0.033	0.040	0.034	0.028	0.053	0.048	0.033	0.029
		0.056	0.050	0.045	0.039	0.040	0.036	0.037	0.032	0.027
		0.037	0.036	0.031	0.037	0.055	0.029	0.024	0.037	
1	4	0.034	0.043	0.025	0.027	0.038	0.049	0.040	0.032	0.029
		0.034	0.039	0.044	0.044	0.034	0.039	0.045	0.044	0.037
		0.055	0.047	0.032	0.027	0.039	0.037	0.039	0.035	
1	5	0.043	0.033	0.028	0.046	0.043	0.044	0.039	0.031	0.026
		0.030	0.036	0.040	0.041	0.024	0.019	0.048	0.070	0.044
		0.028	0.038	0.044	0.043	0.047	0.033	0.026	0.046	
2	3	0.046	0.048	0.041	0.032	0.036	0.035	0.036	0.030	0.024
		0.039	0.034	0.029	0.040	0.067	0.041	0.033	0.037	0.045
		0.033	0.033	0.027	0.033	0.045	0.052	0.042	0.030	

$k_1 - k_2 = 9$

$k_1 - k_5 = 16$

$k_2 - k_3 = 13$

实验原理--维吉尼亚密码分析三步走

- **第二步确定密钥字相对位移**
  - **交互重合指数**

续表

i	j	$MI_c(y_i, y_j^g)$ 的值，$0 \leqslant g \leqslant 25$								
2	4	0.046	0.034	0.043	0.044	0.034	0.031	0.040	0.045	0.040
		0.048	0.044	0.033	0.024	0.028	0.042	0.039	0.026	0.034
		0.050	0.035	0.032	0.040	0.056	0.043	0.028	0.028	
2	5	0.033	0.033	0.036	0.046	0.026	0.018	0.043	0.080	0.050
		0.029	0.031	0.045	0.039	0.037	0.027	0.026	0.031	0.039
		0.040	0.037	0.041	0.046	0.045	0.043	0.035	0.030	
3	4	0.038	0.036	0.040	0.033	0.036	0.060	0.035	0.041	0.029
		0.058	0.035	0.035	0.034	0.053	0.030	0.032	0.035	0.036
		0.036	0.028	0.046	0.032	0.051	0.032	0.034	0.030	
3	5	0.035	0.034	0.034	0.036	0.030	0.043	0.043	0.050	0.025
		0.041	0.051	0.050	0.035	0.032	0.033	0.033	0.052	0.031
		0.027	0.030	0.072	0.035	0.034	0.032	0.043	0.027	
4	5	0.052	0.038	0.033	0.038	0.041	0.043	0.037	0.048	0.028
		0.028	0.036	0.061	0.033	0.033	0.032	0.052	0.034	0.027
		0.039	0.043	0.033	0.027	0.030	0.039	0.048	0.035	

$k_2 - k_5 = 7$

$k_3 - k_5 = 20$

$k_4 - k_5 = 11$

实验原理--维吉尼亚密码分析三步走

- **第三步穷举搜索密钥字**

根据上面第二步求出的密钥字之间的关系，穷举搜索26中可能的取值即可，也就是只用猜测k_1的值，其他密钥字可以根据与 k_1的关系推算出来。

通过尝试k1的每个可能的取值，可以确定密钥字为 JANET，k=(9,0,13,4,19)。

The almond tree was in tentative blossom. The days were longer, often ending with magnificent evenings of corrugated pink skies. The hunting season was over, with hounds and guns put away for six months. The vineyards were busy again as the well-organized farmers treated their vines and the more lackadaisical neighbors hurried to do the pruning they should have done in November.

$$k_2 = k_1 + 17$$

$$k_3 = k_1 + 4$$

$$k_4 = k_1 + 21$$

$$k_5 = k_1 + 10$$