云南大学数学与与统计学院

上机实践报告

课程名称：信息论基础实验	**年级：**2013	**上机实践成绩：**
指导教师：陆正福	**姓名：**金洋	
上机实践名称：统计分析攻击实验	**学号：**20131910023	**上机实践日期：** **2016/6/10**
上机实践编号：No. 9	**组号：**	**上机实践时间：** **19:13**

一、实验目的

理解古典密码学、统计分析攻击、信息熵与密码体制的关系

二、实验内容

1. 古典密码体制 Caesar 密码的编程（参阅数据结构与算法主讲课教材的 3.1.4 节）
2. 任取一段较长的且有意义的英文片段，计算其明文熵。
3. 用 Caesar 密码加密上述英文片段，计算其密文熵。
4. 查阅英文字母的频度分布规律，对上述密文进行统计分析攻击。

三、实验环境

1. 个人计算机，任意可以完成实验的平台，如 Java 平台、Python 语言、R 语言、Matlab 平台、Magma 平台等。
2. 对于信息与计算科学专业的学生，建议选择 Java、Python、R 等平台。
3. 对于非信息与计算科学专业的学生，建议选择 Matlab、Magma 等平台。

四、实验记录与实验结果分析

（注意记录实验中遇到的问题。实验报告的评分依据之一是实验记录的细致程度、实验过程的真实性、实验结果的解释和分析。**如果涉及实验结果截屏，应选择白底黑字。**）

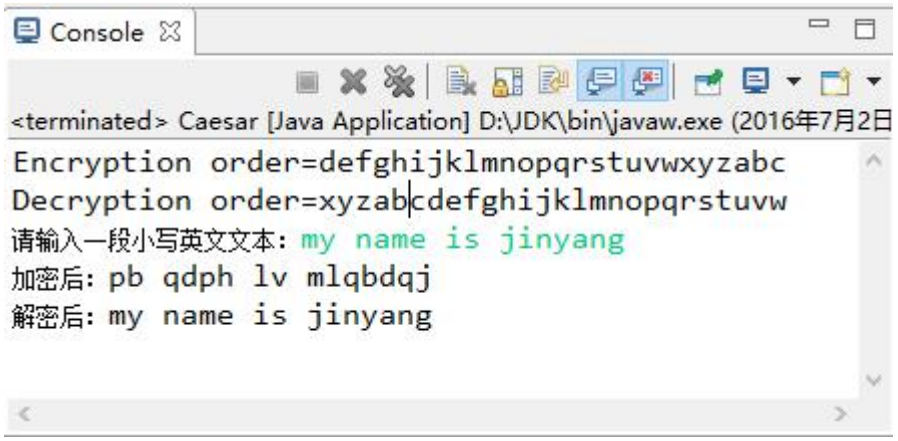
1. 古典密码体制 Caesar 密码的编程（参阅数据结构与算法主讲课教材的 3.1.4 节）

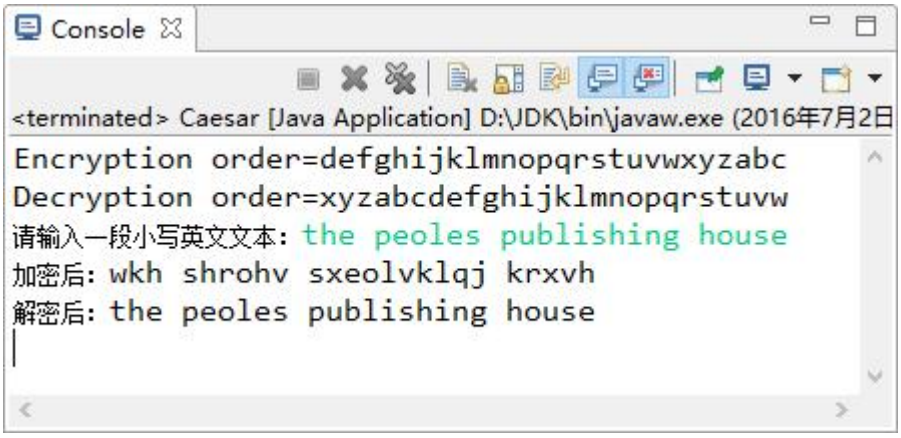
Caesar.java

```
package IT9;

import java.io.*;
public class Caesar {
      public static final int ALPHASIZE=26;
      public static final char[]
alpha={'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'};
      protected char[] encrypt=new char[ALPHASIZE];//加密数组;
      protected char[] decrypt=new char[ALPHASIZE];//解密数组

      public Caesar(){
            for (int i=0;i<ALPHASIZE;i++)
encrypt[i]=alpha[(i+3)%ALPHASIZE];
            for (int i=0;i<ALPHASIZE;i++) decrypt[encrypt[i]-'a']=alpha[i];
      }

      public String encrypt(String secret){
            char[] mess=secret.toCharArray();
            for (int i=0;i<mess.length;i++)
                  if ((mess[i]>='a')&&(mess[i]<='z'))
                        mess[i]=encrypt[mess[i]-'a'];
            return new String(mess);
      }

      public String decrypt(String secret){
            char[] mess=secret.toCharArray();
            for (int i=0;i<mess.length;i++)
                  if ((mess[i]>='a')&&(mess[i]<='z'))
                        mess[i]=decrypt[mess[i]-'a'];
            return new String(mess);
      }
      public static void main(String[] args) throws IOException{
            Caesar cipher=new Caesar();
            System.out.println("Encryption order="+new
String(cipher.encrypt));
            System.out.println("Decryption order="+new
String(cipher.decrypt));

            String secret=new String();

            BufferedReader input=new BufferedReader(new
InputStreamReader(System.in));

            System.out.print("请输入一段小写英文文本：");
            secret=input.readLine();

            secret=cipher.encrypt(secret);
            System.out.println("加密后："+secret);
            secret=cipher.decrypt(secret);
            System.out.println("解密后："+secret);
      }
}
```

测试结果：





2.任取一段较长的且有意义的英文片段，计算其明文熵。

涉及到熵的计算我们导入 IT2 熵的计算实验中的对文本计算熵的程序，并加上本实验 1 中的程序，由于只需使程序实现基本的方法，我们设计的程序只对小写英文字符进行处理。

Entropy.java

```
package IT9;

public class Entropy {
	protected int m,n;//m*n 分布
	protected float[][] p;
	protected float[] px;
	protected float[] py;

	public Entropy() {

	}

	public Entropy(int m,int n, float[][] p) {
		this.p=new float[m][n];
		this.py=new float[m];
```

```
        this.px=new float[n];
        System.arraycopy(p, 0, this.p, 0, p.length);
        this.m=m;
        this.n=n;
    }


    /**
     *
     * @param value 对数的真数
     * @param base 对数的底数
     * @return 对数的值
     */
    public double log(double value, double base) {
        return Math.log(value) / Math.log(base);
    }



    /**
     *
     * @param q:一维分布 q
     * @return：熵
     */
    public float H(float[] q) {
        float h=0;
        int i;
        for (i=0;i<q.length;i++)
            if (q[i]>1e-7)//q[i]=0 时不必相加，约定 0log0=0
                h+=-q[i]*log(q[i],2);
        return h;
    }

    /**
     *
     * @return H(X)
     */
    public float HX() {
        this.px=new float[n];
        int i,j;
        for (i=0;i<n;i++)
            for (j=0;j<m;j++)
                px[i]+=p[j][i];
        return H(px);
    }

    /**
     *
     * @return H(Y)
```

```
 */
public float HY() {
	this.py=new float[m];
	int i,j;
	for (i=0;i<m;i++)
		for (j=0;j<n;j++)
			py[i]+=p[i][j];
	return H(py);
}

/**
 *
 * @return H(X|Y)
 */
public float HX_Y() {
	float[] px_y=new float[n];
	float hx_y=0;
	int i,j;
	for (i=0;i<m;i++) {
		for (j=0;j<n;j++) px_y[j]=p[i][j]/py[i];
		hx_y+=py[i]*H(px_y);
	}

	return hx_y;
}

/**
 *
 * @return H(Y|X)
 */
public float HY_X() {
	float[] py_x=new float[m];
	float hy_x=0;
	int i,j;
	for (i=0;i<n;i++) {
		for (j=0;j<m;j++) py_x[j]=p[j][i]/px[i];
		hy_x+=px[i]*H(py_x);
	}
	return hy_x;
}


/**
 *
 * @return 联合熵 H(X,Y)
 */
public float HXY() {
	return HX()+HY_X();
```

```
	}

	/**
	 *
	 * @return 互信息 I(X,Y)
	 */
	public float IXY() {
		return HX()-HX_Y();
	}
}
```

CalcEntropyofData.java

```
package IT9;
import java.io.*;
public class CalcEntropyofData {

	public static void main(String[] args) throws IOException {
		// TODO Auto-generated method stub
		int i;
		float[] p=new float[26];
		String text=new String();

		BufferedReader input=new BufferedReader(new
InputStreamReader(System.in));
		System.out.print("请输入一段由英文大写字母组成的文本：");


		text=input.readLine();

		for (i=0;i<text.length();i++)
			p[text.charAt(i)-97]++;
		System.out.println("文本中的符号频率分布为：");
		for (i=0;i<26;i++)
			if (p[i]>1e-7) {
				p[i]=p[i]/text.length();
				System.out.printf("%c：%f\n",i+97,p[i]);
			}



		Entropy En=new Entropy();
		System.out.println("文本中的符号频率分布的随机变量的熵
H="+En.H(p));
```

```
        }
}
```

选取以下来自新华社的一则新闻：

The People's Publishing House announced Thursday a reprint of the first edition of the complete works of Marx and Engels, to coincide with the CPC anniversary.

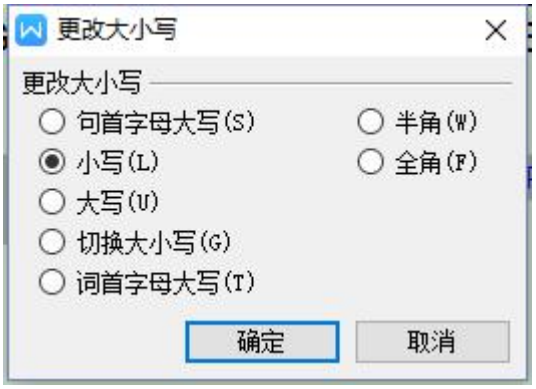
A Marxism with Chinese characteristics exhibition opened at Beijing's Cultural Palace of Nationalities on Thursday.

The exhibition includes over 1,100 artifacts, literature and manuscripts, some 290 pictures, and more than 20 Marxist-themed paintings, sculptures.

Also on Thursday, the State Administration of Cultural Heritage promoted ten exhibitions nationwide to commemorate the anniversary.

The exhibitions include one at the China National Museum of Women and Children, which is on female Communist soldiers, and one featuring revolutionary slogans at Chongqing China Three Gorges Museum.

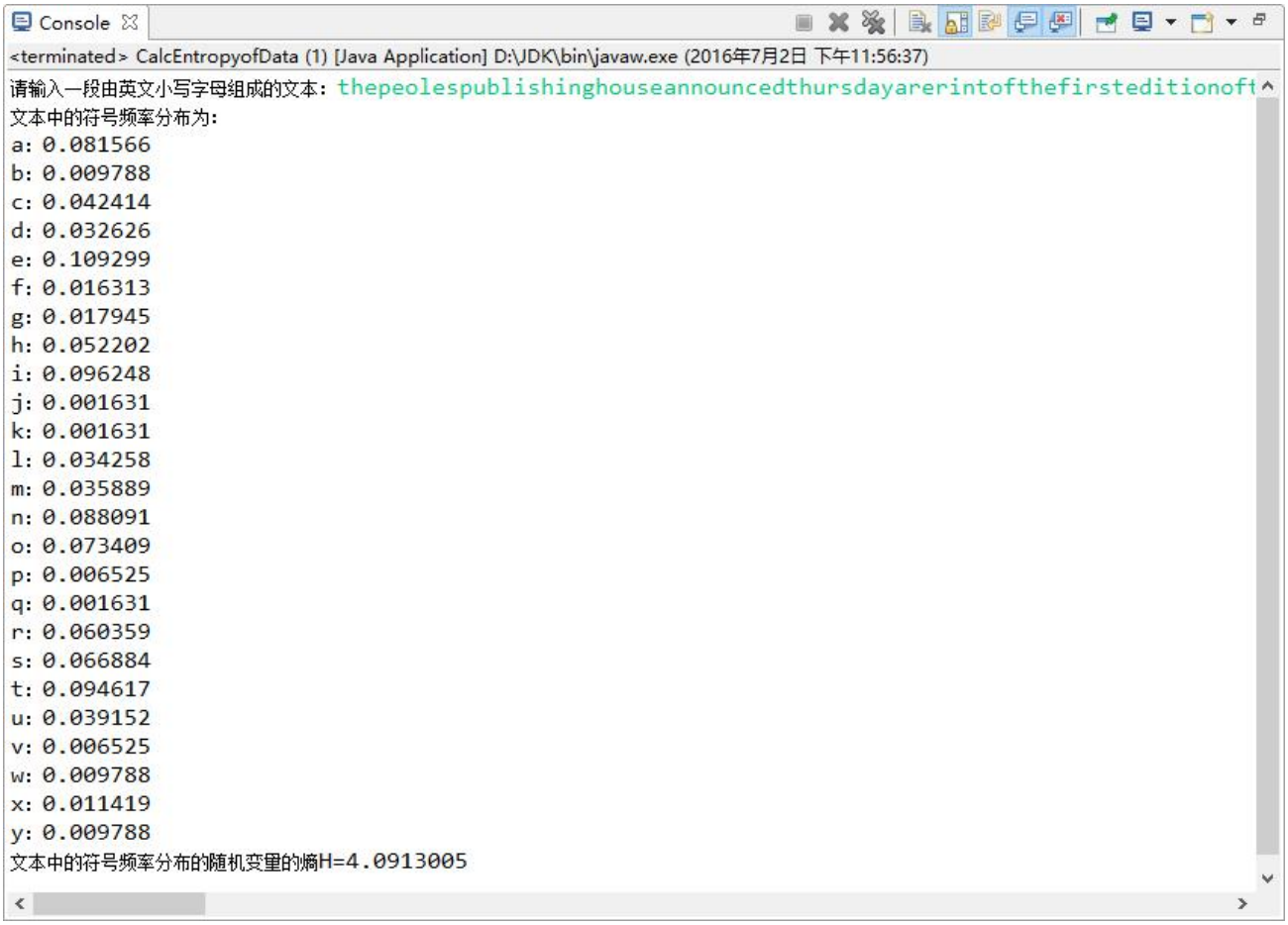
对了对应已经完成的程序，（为简便，我们的程序只对大写字母进行处理），所以我们通过 word 对以上文本删去标点、数字，并改写为小写，具体操作如下：



最后得到如下明文（虽然读者有些费力），但还是属于一段有意义的明文，

thepeolespublishinghouseannouncedthursdayarerintofthefirsteditionofthecomlete worksofmarxandengelstocoincidewiththecpcanniversaryamarxismwithchinesecha racteristicsexhibitionoenedatbeijingsculturalpalaceofnationalitiesonthursdaytheex hibitionincludesoverartifactsliteratureandmanuscritssomeicturesandmorethanmar xistthemedaintingssculturesalsoonthursdaythestateadministrationofculturalherita geromotedtenexhibitionsnationwidetocommemoratetheanniversarytheexhibitionsi ncludeoneatthechinanationalmuseumofwomenandchildrenwhichisonfemalecom munistsoldiersandonefeaturingrevolutionaryslogansatchongqingchinathreegorge smuseum

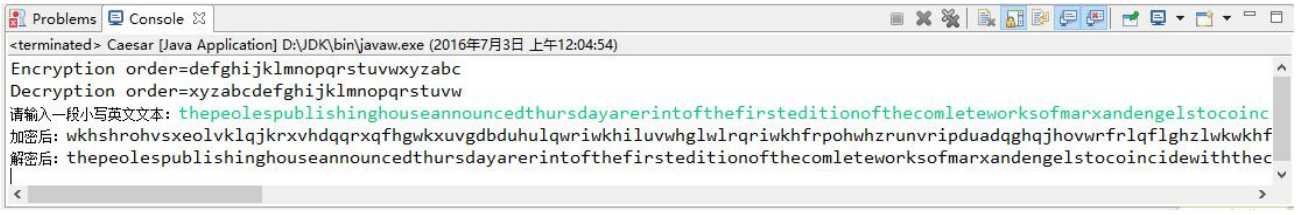
将以上明文作为程序输入



所以明文熵 H=4.09；

3.用 Caesar 密码加密上述英文片段，计算其密文熵。

Caesar 密码加密：



具体地：

```
Encryption order=defghijklmnopqrstuvwxyzabc
Decryption order=xyzabcdefghijklmnopqrstuvw
请输入一段小写英文文本：
thepeolespublishinghouseannouncedthursdayarerintofthefirsteditionofthecom
leteworksofmarxandengelstocoincidewiththecpcanniversaryamarxismwithchines
echaracteristicsexhibitionoenedatbeijingsculturalpalaceofnationalitiesont
hursdaytheexhibitionincludesoverartifactsliteratureandmanuscritssomeictur
```

esandmorethanmarxistthemedaintingsculturesalsoonthursdaythestateadministrationofculturalheritageromotedtenexhibitionsnationwidetocommemoratethean niversarytheexhibitionsincludeoneatthechinanationalmuseumofwomenandchildrenwhichisonfemalecommunistsoldiersandonefeaturingrevolutionaryslogansatchongqingchinathreegorgesmuseum

加密后：

wkhshrohvsxeolvklqjkrxvhdqqrxqfhgwkxuvgdbduhulqwriwkhiluvwhglwlrqriwkhfrpohwhzrunvripduadqghqjhovwrfrlqflghzlwkwkhfsfdqqlyhuvdubdpdualvpzlwkfklqhvhfkdudfwhulvwlfvhaklelwlrqrhqhgdwehlmlqjvfxowxudosdodfhriqdwlrqdolwlhvrqwkxuvgdbwkhhaklelwlrqlqfoxghvryhuduwlidfwvolwhudwxuhdqgpdqxvfulwvvrphlfwxuhvdqgpruhwkdqpdualvwwkhphgdlqwlqjvvfxowxuhvdovrrqwkxuvgdbwkhvwdwhdgplqlvwudwlrqrifxowxudokhulwdjhurprwhgwhqhaklelwlrqvqdwlrqzlghwrfrpphprudwhwkhdqqlyhuvdubwkhhaklelwlrqvlqfoxghrqhdwwkhfklqdqdwlrqdopxvhxprizrphqdqgfkloguhqzklfklvrqihpdohfrppxqlvwvroglhuvdqgrqhihdwxulqjuhyroxwlrqdubvorjdqvdwfkrqjtlqjfklqdwkuhhjrujhvpxvhxp

解密后：

thepeolespublishinghouseannouncedthursdayarerintofthefirsteditionofthecompleteworksofmarxandengelstocoincidewiththecpcanniversaryamarxismwithchinesecharacteristicsexhibitionoenedatbeijingsculturalpalaceofnationalitiesonthursdaytheexhibitionincludesoverartifactsliteratureandmanuscritssomeicturesandmorethanmarxistthemedaintingsculturesalsoonthursdaythestateadministrationofculturalheritageromotedtenexhibitionsnationwidetocommemoratethean niversarytheexhibitionsincludeoneatthechinanationalmuseumofwomenandchildrenwhichisonfemalecommunistsoldiersandonefeaturingrevolutionaryslogansatchongqingchinathreegorgesmuseum

计算密文熵：

具体地：

请输入一段由英文小写字母组成的文本：
wkhshrohvsxeolvklqjkrxvhdqqrxqfhgwkxuvgdbduhulqwriwkhiluvwhglwlrqriwkhfrp
ohwhzrunvripduadqghqjhovwrfrlqflghzlwkwkhfsfdqqlyhuvdubdpdualvpzlwkfklqhv
hfkdudfwhulvwlfvhaklelwlrqrhqhgdwehlmlqjvfxowxudosdodfhriqdwlrqdolwlhvrqw
kxuvgdbwkhhaklelwlrqlqfoxghvryhuduwlidfwvolwhudwxuhdqgpdqxvfulwvvrphlfwxu
hvdqgpruhwkdqpdualvwwkhphgdlqwlqjvvfxowxuhvdovrrqwkxuvgdbwkhvwdwhdgplqlvw
udwlrqrifxowxudokhulwdjhurprwhgwhqhaklelwlrqvqdwlrqzlghwrfrpphprudwhwkhdq
qlyhuvdubwkhhaklelwlrqvlqfoxghrqhdwwkhfklqdqdwlrqdopxvhxprizrphqdqgfklogu
hqzklfklvrqihpdohfrppxqlvwvroglhuvdqgrqhihdwxulqjuhyroxwlrqdubvorjdqvdwfk
rqjtlqjfklqdwkuhhjrujhvpxvhxp
文本中的符号频率分布为：
a：0.011419
b：0.009788
d：0.081566
e：0.009788
f：0.042414
g：0.032626
h：0.109299
i：0.016313
j：0.017945
k：0.052202
l：0.096248
m：0.001631
n：0.001631
o：0.034258
p：0.035889
q：0.088091
r：0.073409
s：0.006525
t：0.001631
u：0.060359
v：0.066884
w：0.094617
x：0.039152
y：0.006525
z：0.009788
文本中的符号频率分布的随机变量的熵 H=4.0913005

所以密文熵 H=4.09，与明文熵一致。

4.查阅英文字母的频度分布规律，对上述密文进行统计分析攻击。

查阅英文字母的频度分布规律为：

字母	英语中出现的频率
a	8.167%
b	1.492%
c	2.782%
d	4.253%
e	12.702%
f	2.228%
g	2.015%
h	6.094%
i	6.966%
j	0.153%
k	0.772%
l	4.025%
m	2.406%
n	6.749%
o	7.507%
p	1.929%
q	0.095%
r	5.987%
s	6.327%
t	9.056%
u	2.758%
v	0.978%
w	2.360%
x	0.150%
y	1.974%
z	0.074%

英文中频率最高的是 e：12.702%；密文中频率最高的是 h：0.109299；故可能密文的 h 译码为 e；

英文中频率第二高的是 t：9.056%；密文中频率近似第二高的是 w：0.094617 和 l：0.096248 故可能密文的 w 译码为 t 或 l；

英文中频率第三高的是 a：8.167%；

……

其实，由于是凯撒密码，只需知道移动位数即可。这里根据英文中频率最高的是 e：12.702%；密文中频率最高的是 h：0.109299；可以初步判断出移动位数为 3；

而根据英文中频率第二高的是 t：9.056%；密文中频率近似第二高的是 w：0.094617 和 l：0.096248 故可能密文的 w 译码为 t 或 l；结合上一个频率最高的字母的分析，在此时基本可以确定移动位数为 3。试着将其解密密文，便可得到明文

thepeolespublishinghouseannouncedthursdayarerintofthefirsteditionofthecomlete worksofmarxandengelstocoincidewiththecpcanniversaryamarxismwithchinesecha racteristicsexhibitionoenedatbeijingsculturalpalaceofnationalitiesonthursdaytheex hibitionincludesoverartifactsliteratureandmanuscritssomeicturesandmorethanmar xistthemedaintingssculturesalsoonthursdaythestateadministrationofculturalherita geromotedtenexhibitionsnationwidetocommemoratetheanniversarytheexhibitionsi ncludeoneatthechinanationalmuseumofwomenandchildrenwhichisonfemalecom munistsoldiersandonefeaturingrevolutionaryslogansatchongqingchinathreegorge smuseum

经过一些标点的处理便可得到

The People's Publishing House announced Thursday a reprint of the first edition of the complete works of Marx and Engels, to coincide with the CPC anniversary.

A Marxism with Chinese characteristics exhibition opened at Beijing's Cultural Palace of Nationalities on Thursday.

The exhibition includes over 1,100 artifacts, literature and manuscripts, some 290 pictures, and more than 20 Marxist-themed paintings, sculptures.

Also on Thursday, the State Administration of Cultural Heritage promoted ten exhibitions nationwide to commemorate the anniversary.

The exhibitions include one at the China National Museum of Women and Children, which is on female Communist soldiers, and one featuring revolutionary slogans at Chongqing China Three Gorges Museum.

五、实验体会

（请认真填写自己的真实体会）

1、使用 Caesar 密码加密解密后，明文熵和密文熵相等，这是因为 Caesar 密码加解密只对字母进行相同个数的位置平移，虽然加密前后，字母在变化，但是他们对应的概率还是同一批数，所以明文熵和密文熵将相等。
2、和所有的利用字母表进行替换的加密技术一样，恺撒密码非常容易被破解，而且在实际应用中也无法保证通信安全。

六、参考文献

1. Thomas M. Cover, Joy A. Thomas. Elements of Information Theory (2^{nd} Edition) [M]. John Wiley & Sons, Inc. Chapter 4

2. Toggle navigation.Letter frequency (English) [EB/OL]. http://en.algoritmy.net/article/40379/Letter-frequency-English, 2015;