一种新式 Vigenere 密码的破译和研究

何晓琴[1] 陆一南[2]
(重庆高等电力专科学校计算机科学系 重庆 400053)[1]
(马来西亚林国荣创意科技大学信息工程系 吉隆坡 63000)[2]

摘 要 随着信息技术的飞速发展,如何保护信息的安全,使之不被窃取及不至于被篡改或破坏,已成为当今普遍关注的热点问题。为了解决这个热门问题,主要研究了一种新式 Vigenere 密码的解译算法,其通过确定 Vigenere 密钥的长度和密钥字以及明文,最终得出该种新式 Vigenere 密码的破译算法。

关键词 Vigenere 密码,Kasiski 测试法,重合指数,重合互指数,相对移位
中图法分类号 TP393.11 **文献标识码** A

Modeling Research in Complex Network of Traditional Economic System

HE Xiao-qin[1] LU Yi-nan[2]
(Department of Computer Science, Chongqing Electric Power Specialist University, Chongqing 400053, China)[1]
(Department of Information Engineering, Limkokwing University of Creative Technology, Kuala Lumpur 63000, Malaysia)[2]

Abstract With the rapid development of information technology, how to protect the security of the information not stolen and not subject to tampering or destruction, has become a hot issue in present. In order to solve this question, this paper mainly researched the interpreter of a new vigenere password algorithm. We used vigenere to determine the length of the key and the key word and plaintext. Eventually we got the new vigenere password decipher algorithm.

Keywords Vigenere password, Kasiski test method, Coincident index, Coincidence interaction index, Rrelative shift

1 引言

当今社会由于信息技术的飞速发展,使得人们的生活面貌发生了很大的变化,同时信息技术的发展也极大地促进了社会生产力的发展。在计算机网络日益普及的信息时代,信息本身就是时间,就是财富。如何保护信息的安全使之不被窃取及不至于被篡改或破坏,已成为当今计算机界普遍关注的重大问题。而 Vigenere 密码则在密码学的发展史中起到了里程牌式的作用,如何有效解译 Vigenere 密码已经成为目前计算机密码学科的一个热门问题。

2 问题的提出

对于如何破译 Vigenere 密码,我们以如下密文为例:

Ahmnywiwoeeraltllbisqxvrgusewhsbahbskoermbmvhubknfeibhgvzhvrosrfmoelchsudubkfgyrlvozsiohgoydfohzifloairghxmeerrvfoaivbtfvnrjgsjvmhfanqoenisayyaaghusatuvgywatrefrsjhrfmbmkoabhjjptrgczwahvfkbmsdpbacjvmvfsndsoshgnfkaobjrsslqbbblyeghfamudreoaqjiepizkaaapsfaaifxbboutunhvldaflchooourzcwkufgctwaajnmsjvmurfrsudnyzjgblqosygztvsmbmzhbhzqkoojlchjzeyskrsyethoevldqnmnfknvtvgakoabhpsyelbinjlnbgvrjldbacgtltebiodldbaalsjcbhbgfvtuvbtoplyuochlngbarloeanhvguayngfwtbyrkvdssubkywuiraqlxhrrjsydahrzcflsoirrbxteaosdmpegvtlgbcnahnclcnesbxfohegrdmfbealkllsvqnfuogjfvllaalabjlbhgbbloiatwaloejbfyvjohyrflvpzrhbskoermbmbpgbhuwkengv

3 解决思路和具体方法

3.1 确定 Vigenere 密钥字长度

Kasiski 测试法:由普鲁士军官 Kasiski 在 1863 年提出的一种重码分析法。这种方法的基本原理是:若用给定的 d 个密钥表周期地对明文字母加密,则当明文中有两个相同字母组在明文序列中间隔的字母数为 d 的倍数时,这两个明文字母组对应的密文字母组必相同。但反过来,若密文中出现两个相同的字母组,它们所对应的明文字母组未必相同,但相同的可能性很大。如果我们将密文中的相同字母组找出来,并对其相间字母数综合研究,找出它们的相间字母数的最大公因子,就有可能提取出有关密钥子的长度 d 的信息。

密文串 mbm 在密文中共出现了 4 次,起始位置分别是 41,167,321,594,从第 1 个到其它 3 个的距离分别为 126,280,553,这 3 个数的最大公因子是 7,所以 7 很可能是密钥字的长度,下面我们将使用重合指数法来进一步确定密钥的长度。

定义 1(重合指数法) 设 $x=x_1,x_2,\cdots,x_n$ 是 n 个字母的一个串,x 的重合指数定义为 x 中的两个随机元素相同的概率,记为 $I_c(x)$。

假定 $f_0,f_1,\cdots,f_{25}$ 分别表示 x 中字母 $A,B,C,\cdots Z$ 出现的次数。我们能以 $\binom{n}{2}$ 种方法选择 x 中的两个元素,对每一

到稿日期:2013-03-15 返修日期:2013-06-20
何晓琴 女,硕士,副教授,主要研究方向为计算机软件与理论;陆一南 男,主要研究方向为计算机网络,E-mail:450925968@qq.com(通信作者)。

个 i，$0\leqslant i\leqslant 25$，x 中的两个元素都被选择为 i 的方法有 $\binom{f_i}{2}$ 种，因此：

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

现在假定 x 是一个英文明文串。记字母 $A,B,\cdots,Z$ 出现的期望概率分别为 $p_0,p_1,\cdots,p_{25}$。通过对大量的小说、杂志、报纸等的汇编统计，人们已经获得了英文的 26 个字母的概率分布的一个估计，并可计算得：

$$I_c(x)\approx\sum_{i=0}^{25}p_i^2=0.065$$

且经多表代换字符串的重合指数不变。

假定 $y=y_1y_2\cdots y_{n'}$ 是通过 Vigenere 密码所获得的一个密文串。把 y 分成 d 个长为 n/d 的子串，记为 $Y_1,Y_2,\cdots,Y_d$。如果 d 的确是密钥字长度，那么每个 $I_c(Y_i)(1\leqslant i\leqslant d)$ 都将大概等于 0.065。如果 d 不是密钥字的长度，那么字串将看起来更随机些，因为它们将是采用不同的密钥移位加密获得的。

而对一个完全随机的串 x，$I_c(x)\approx 26\left(\frac{1}{26}\right)^2=\frac{1}{26}=0.038$，因为两个值 0.065 和 0.038 间隔得充分远，所以我们通常能确定正确的密钥字长度。

现我们代入计算可得：

当 $d=1$ 时，重合指数为 0.045；

当 $d=2$ 时，两个重合指数分别为 0.044，0.045；

当 $d=3$ 时，3 个重合指数分别为 0.044，0.043，0.044；

当 $d=4$ 时，4 个重合指数分别为 0.044，0.044，0.043，0.045；

当 $d=5$ 时，5 个重合指数分别为 0.043，0.041，0.046，0.044，0.050；

当 $d=6$ 时，6 个重合指数分别为 0.044，0.043，0.048，0.043，0.039，0.043；

当 $d=7$ 时，7 个重合指数分别为 0.067，0.063，0.076，0.059，0.062，0.060，0.060。

这也为密钥字是 7 提供了有力的证据。

3.2 确定 Vigenere 密码字

从以上结论可知我们已经假定了 Vigenere 密钥字的长度为 7。下面我们将采用重合互指数法来确定相对移位，重合互指数的精确定义如下。

定义 2 假定 $x=x_1x_2\cdots x_n$ 和 $y=y_1y_2\cdots y_{n'}$ 分别是长为 n 和 n' 的字母串。x 和 y 的重合互指数定义为 x 的一个随机元素等于 y 的一个随机元素的概率，记为 $MI_c(x,y)$。

现在我们假定 Y_i 与 Y_j 是通过明文的移位加密获得的 7 个子串中的任意 2 个，并假定 $K=(k_1,k_2,\cdots,k_7)$ 是密钥字，考虑 Y_i 中的一个随机字母和 Y_j 中的一个随机字母，两个字母都是 A 的概率是 $p_{-k_i}p_{-k_j}$，两个字母都是 B 的概率是 $p_{1-k_i}p_{1-k_j}$，等等(所有的下标都是模 26 运算)。因此，我们可估算 $MI_c(Y_i,Y_j)\approx\sum_{h=0}^{25}p_{h-k_i}p_{h-k_j}=\sum_{h=0}^{25}p_hp_{h+k_i-k_j}$。易知，$MI_c(Y_i,Y_j)$ 的估计值只依赖于差 $(k_i-k_j)\bmod 26$，也就是 Y_i 和 Y_j 的相对移位。而当相对移位不是 0 时，这些估计值均在 0.031～0.045 之间，而相对移位是 0 时，估计值是 0.065。因而通过计算并查找那些接近于 0.065 的数，找出相应的 (i,j) 就能确定相对移位。

表 1 重合互指数

i j	重合互指数数值表								
	0.032	0.051	0.044	0.037	0.039	0.037	0.032	0.043	0.047
2	0.029	0.033	0.034	0.033	0.043	**0.061**	0.031	0.033	0.046
	0.040	0.023	0.033	0.041	0.031	0.042	0.040	0.044	
	0.039	0.040	0.044	0.036	0.040	0.027	0.041	**0.064**	0.036
3	0.028	0.040	0.041	0.015	0.034	0.040	0.038	0.040	0.047
	0.042	0.033	0.046	0.047	0.040	0.035	0.034	0.031	
	0.036	0.044	0.033	0.041	0.035	0.036	0.028	0.056	0.043
4	0.034	0.038	0.041	0.029	0.040	0.050	0.036	0.038	0.033
1	0.029	0.036	**0.065**	0.035	0.035	0.045	0.042	0.023	
	0.038	0.037	0.039	0.037	0.042	0.036	0.026	0.049	0.049
5	0.034	0.033	0.041	0.029	0.040	0.048	0.040	0.033	0.039
	0.030	0.032	**0.064**	0.044	0.031	0.037	0.044	0.027	
	0.034	0.034	0.040	0.038	0.039	0.034	0.042	0.032	0.048
6	0.040	0.032	0.031	0.036	0.040	0.047	0.043	0.034	0.035
	0.027	0.039	0.045	**0.062**	0.035	0.034	0.035	0.045	
	0.035	0.045	0.034	0.035	0.042	0.042	0.030	0.047	0.046
7	0.038	0.029	0.034	0.038	0.034	0.049	0.044	0.037	0.037
	0.036	0.032	**0.059**	0.042	0.028	0.036	0.043	0.027	
	0.032	0.034	0.049	0.046	0.043	0.034	0.034	0.041	0.042
3	0.032	0.024	0.038	0.030	0.037	0.043	0.049	0.029	0.038
	0.072	0.049	0.028	0.033	0.036	0.027	0.034	0.047	
	0.030	0.037	0.038	0.025	0.034	**0.066**	0.042	0.032	0.042
4	0.040	0.030	0.042	0.043	0.028	0.038	0.047	0.038	0.034
	0.045	0.039	0.040	0.043	0.040	0.028	0.040	0.039	
	0.036	0.032	0.045	0.027	0.029	**0.060**	0.050	0.031	0.035
2 5	0.036	0.032	0.042	0.046	0.031	0.033	0.049	0.044	0.035
	0.043	0.038	0.039	0.043	0.043	0.027	0.036	0.040	
	0.036	0.033	0.034	0.032	0.031	0.039	**0.061**	0.042	0.028
6	0.029	0.036	0.043	0.038	0.040	0.035	0.038	0.046	0.047
	0.036	0.037	0.036	0.046	0.043	0.040	0.031	0.042	
	0.038	0.038	0.045	0.031	0.031	0.057	0.050	0.031	0.033
7	0.041	0.031	0.035	0.051	0.032	0.033	0.047	0.050	0.038
	0.040	0.035	0.037	0.035	0.036	0.031	0.036	0.039	
	0.042	0.037	0.045	0.037	0.036	0.045	0.038	0.027	0.037
4	0.030	0.023	0.041	**0.067**	0.036	0.029	0.049	0.042	0.037
	0.039	0.044	0.028	0.034	0.042	0.037	0.035	0.043	
	0.043	0.035	0.047	0.036	0.033	0.039	0.041	0.029	0.033
5	0.036	0.029	0.031	**0.069**	0.041	0.030	0.037	0.042	0.039
3	0.042	0.039	0.033	0.030	0.048	0.041	0.035	0.042	
	0.033	0.040	0.042	0.046	0.034	0.039	0.044	0.034	0.026
6	0.035	0.027	0.036	0.039	**0.063**	0.035	0.024	0.038	0.045
	0.045	0.032	0.037	0.039	0.035	0.048	0.044	0.040	
	0.042	0.037	0.038	0.028	0.038	0.034	0.040	0.031	0.038
7	0.041	0.030	0.034	**0.069**	0.042	0.029	0.037	0.046	0.032
	0.035	0.044	0.033	0.027	0.048	0.048	0.043	0.035	
	0.046	0.028	0.033	0.040	0.032	0.045	0.044	0.032	0.033
5	0.047	0.036	0.033	0.048	0.040	0.031	0.038	0.041	0.030
	0.040	0.045	0.033	0.033	0.047	0.031	0.031	**0.065**	
	0.061	0.037	0.028	0.030	0.043	0.045	0.040	0.038	0.036
4 6	0.037	0.041	0.042	0.037	0.041	0.033	0.036	0.039	0.039
	0.035	0.042	0.037	0.030	0.036	0.036	0.039	0.041	
	0.046	0.027	0.030	0.043	0.032	0.038	0.049	0.033	0.032
7	0.045	0.041	0.036	0.044	0.036	0.032	0.035	0.037	0.036
	0.039	0.047	0.035	0.037	0.045	0.035	0.030	0.059	
	0.063	0.029	0.030	0.036	0.044	0.043	0.041	0.033	0.040
6	0.034	0.045	0.037	0.038	0.037	0.037	0.032	0.046	0.031
5	0.040	0.038	0.043	0.031	0.037	0.028	0.043	0.045	
	0.035	0.032	0.032	0.043	0.029	0.044	0.041	0.033	0.029
7	0.052	0.032	0.043	0.040	0.038	0.030	0.045	0.030	0.041
	0.033	0.052	0.033	0.042	0.037	0.035	0.033	**0.066**	
	0.043	0.026	0.040	0.027	0.038	0.042	0.038	0.028	0.045
6 7	0.038	0.041	0.036	0.040	0.036	0.041	0.031	0.044	0.034
	0.046	0.040	0.050	0.034	0.030	0.031	**0.059**	0.040	

通过程序 2 我们计算出了 $26\times 21=546$ 个值 $MI_c(Y_i,Y_j^g)$，$1\leqslant i<j\leqslant 7$，$0<g\leqslant 25$，详细请参见表 1。

对每对(i,j)，筛选出接近 0.065 的 $MI_c(Y_i,Y_j^g)$的值，在表格中用加粗字体标出了这 18 个值。这些值表明：Y_1 和 Y_2 的相对移位可能是 15，Y_1 和 Y_3 的相对移位可能是 8，Y_1 和 Y_4 的相对移位可能是 21，Y_1 和 Y_5 的相对移位可能是 21，Y_1 和 Y_6 的相对移位可能是 22，Y_1 和 Y_7 的相对移位可能是 21。这些关系给出了关于未知量 k_1,k_2,k_3,k_4,k_5 的如下关系式：

$k_2=k_1+15, k_3=k_1+8, k_4=k_1+21$

$k_5=k_1+21, k_6=k_1+22, k_7=k_1+21$

所以 Vigenere 密钥字的可能形式为：

$(k_1, k_1+15, k_1+8, k_1+21, k_1+21, k_1+22, k_1+21)$，$k_1 \in Z_{26}$

k_1 遍历 26 个英文字母后，可以得出 26 种可能的 Vigenere 密钥，详细数据请见表 2。

表 2　26 种可能的密钥

apivvwv	ixqdded	qfyllml	yngttut
bqjwwxw	jyreefe	rgzmmnm	zohuuvu
crkxxyx	kzsffgf	**shannon**	
dslyyzy	latgghg	tiboopo	
etmzzaz	mbuhhih	ujcppqp	
funaaba	ncviiji	vkdqqrq	
gvobbcb	odwjjkj	wlerrsr	
hwpccdc	pexkklk	xmfssts	

为了在穷搜索方式下节省尽可能多的时间，我们仅尝试在 mbm 下的解密结果。我们发现密钥为 shannon 时 mbm 解密结果为 you，故 shannon 极有可能为所求密钥字。而后我们将其代入程序 4 后发现密文可解密为通顺的英文段落，故最终确定 Vigenere 密钥密码字为 shannon。

4　程序模拟的结果

我们运用编写好的程序 3 来验算模拟我们推导出的结果（程序 1、2、4 不再累述）：

```
//本程序旨在根据论文得出的数据，求出将密文分为 7 段以及 7 段之
  间的相互移位后，穷搜 26 种可能，得到 26 种可能的密钥段。
public class Step3 {
public static void main(String[]args) {
char[]k=new char[7];//建立长度为 7 的字符数组
  int i=0;
for (i=0;i<26;i++)//穷搜 26 种可能，求出密钥段
{
k[0]=(char)(97+i);
k[1]=(char)(k[0]+15);
k[2]=(char)(k[0]+8);
k[3]=(char)(k[0]+21);
k[4]=(char)(k[0]+21);
k[5]=(char)(k[0]+22);
k[6]=(char)(k[0]+21);
for(int j=0;j<7;j++){
  if(k[j]>122)
k[j]=(char)((int)k[j]-26);
}
System. out. println(k[0]+" "+k[1]+" "+k[2]+" "+k[3]+" "
+k[4]+" "+k[5]+" "+k[6]);//将密钥段打印出来
}}}
```

我们根据程序 3 所求出的 Vigenere 密钥，通过运行程序 4 后解密得到的 Vigenere 译文如下：

iamaliveheremybelovedforthereasontoadoreyouohhowanxiousihavebeenforyouandhowsorryiamaboutallyoumusthavesufferedinhavingnonewsfromusmayheavengrantthatthisletterreachesyoudonotwritetomethiswouldcompromiseallofusandaboveallldonotreturnunderanycircumstancesitisknownthatitwasyouwhohelpedustogetawayfromhereandallwouldbelostifyoushouldshowyourselfweareguardeddayandnightidonotcareyouarenothereddonotbetroubledonmyaccountnothingwillhappentomethenationalassemblewillshowleniencyfarewellthemostlovedofmenbequietifyoucantakecareofyourselfformyselficannotwriteanymorebutnothingintheworldcouldstopmetoadoreyouuptothedeath

我们通过精准的英语断句和语法整理后得到了该条 Vigenere 密文的明文结果如下：

I am alive here, my beloved, for the reason to adore you. Oh! How anxious I have been for you and how sorry I am about all you must have suffered in having no news from us. May heaven grant that this letter reaches you. Do not write to me, this would compromise all of us and above all, do not return under any circumstances. It is known that it was you who helped us to get away from here and all would be lost if you should show yourself. We are guarded day and night. I do not care you are not here. Do not be troubled on my account. Nothing will happen to me. The national assemble will show leniency. Farewell the most loved of men. Be quiet if you can take care of yourself. For myself I cannot write any more, but nothing in the world could stop me to adore you up to the death.

结束语　我们检验程序 4 解译出的 Vigenere 密文的明文结果后得知，我们使用的新式 Vigenere 密码的破译方法是正确的，我们通过确定 Vigenere 密钥的长度和密钥字以及明文，最终得出该种新式 Vigenere 密码的破译算法。

参 考 文 献

[1] Latora V, Marchiori M. The Architecture of complex systems [M]//Gell-Mann M, Tsallis C. Nonextensive entropy: interdisciplinary applications. New York: Oxford University Press, 2002: 377-386

[2] Newman M E J. The structure and function of complex network [J]. SIAM Review, 2003, 45(2): 167-256

[3] Newman M E J. Scientific collaboration networks. I. Network construction and fundamental result [J]. Physical Review E, 2001, 64(1)

[4] 张建辉，王典洪. 一种基于 Java Applet 的 Vigen～re 密文破译方法[J]. 计算机与现代化，2003，5：17-20

[5] 徐小华. Vigenere 密码加密和解密的程序实现[J]. 信息技术，2010，2：121-123

[6] 王飞. 基于 vigenere 密码的文件加解密器的研究与实现[J]. 信息系统工程，2011，7：50-52

[7] 刘桂海，张璟. 基于双密钥的三维 Vigenere 密码的研究与应用[J]. 计算机工程与应用，2008，12：131-133

[8] 韩春艳，包旭雷，王光义. 一种新的数字混沌密码序列及其性能分析[J]. 重庆邮电大学学报：自然科学版，2010，22(3)：334-338