密码和密码模型

Abstract: 介绍初等的密码知识、加密解密过程及相关的数学原理

Keywords: 密钥、加密、解密

1 Introduction

密码作为军事和政治斗争中的一门技术，已经有了上千年的历史，但是，密码学作为一门科学，还不过是近几十年的事情。特别是在计算机科学蓬勃发展促进下，数据安全作为一个新的分支活跃在计算机领域。

据史料记载,密码最早产生于希腊.古希腊北路军司令莱山得在征服雅典之后,信使赶到,献上了一条皮带,上面有文字,通报敌军欲断其归路的企图.4 世纪希腊出现了隐蔽书信内容的初级密码. 1200年罗马政府和意大利政府开始有系统地使用密码.19世纪随著资本主义的发展和资产阶级相互斗争的需要,出现了无线电密码通信.

对于密码学的研究需要加密和解密两个过程，密码的传输方不希望密码被破译，而解密方则向竭力去破解，这就好比是矛和盾。特别是在二次世界大战中，密码的传输和破译起了非常重要的作用。1942年，美国从破译日本海军密报中，获悉日军对中途岛地区的作战意图和兵力部署，从而能以劣势兵力击破日本海军的主力，扭转了太平洋地区的战局。在保卫英伦三岛和其他许多著名军事事件中，密码破译的成功都起到了极其重要的作用，这些事例也从反面说明了密码保密的重要地位和意义。

我们生活在一个资讯发达的社会里,利用互联网互通消息已经不是一件罕见的事.大家有没有想过,在互联网上传送的信息有机会被人截取呢？其实这是经常发生的事,那么我们还有什么秘密可言,于是在信息传输过程中的信息密码学便应运而生，保护我们的信息不被窃取.

在信息的传递过程中，如果A方要把信息通过公共信道向另外一方B传送信息m，但是由于公共通道缺乏安全保护，信息很容易被第三者窃取。甚至被篡改，所以在信息的传递过程前，需要把这样的信息变成秘密的形式。那么这条可以直接识别或使用的代码(如文字等)我们称之为明码;而密码就是将明码经过了一定处理,转换成一种外人无法直接识别或使用的信息.把明文变成密文的过程称为加密。如果有人知道了密码而把密文变为明文的过程称为解密。而在密码中的关键信息称之为密钥。因此密钥在保密通讯中占有极其重要的地位，通常

情况下密钥由通讯双方秘密商定。

置换密码是一个最容易实现且最为人们熟悉的密码，也称凯撒密码，它只需要把每个字母由其它的字母来替换而形成密文。替换的规则是随机的或者是系统的。将讯息(明码)中的字母,用不同的字母代替.其一做法是系统地将字母向后推几个位置，不妨设三个位置,因此加密时a用D取代,b是E,依此类推.最后的x,y,z 三字则转过头来,分别是A,B,C.我们可用下表,将明码字母与密码字母有效地变换.对照表(1)

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

例子:

1. this message is top secret →WKLVPHVVDJHLVWRSVHFUHW

2. we are the champion →ZHDUHWKHFKDPSLRQ

3. hello how are you →KHOORKRZDUBRX

凯撒密码是一种可以单独应用的密码系统,它可以用对照表或轮盘来进行,也可以用电脑来进行而不必用到任何数字,但是密码系统的基本结构若能使用数字是最理想的. 原因如下：

1. 电脑十分适合以很快的速度来处理数字,即使很大的数字都没有问题. 2. 有很多数学函数可用来把一个数,以很复杂的方法变成另外一个数. 3. 这样可以设计出很有效率,但十分安全的密码系统.

用00-25这26个数字来取代26个英文字母,如下表:

a→00 h→07 o→14 v→21 b→01 i→08 p→15 w→22 c→02 j→09 q→16 x→23 d→03 k→10 r→17 y→24 e→04 l→11 s→18 z→25 f→05 m→12 t→19 g→06 n→13 u→20

加密公式

$$C \equiv P + s(\mod 26) \tag{1.1}$$

这里P是明码，C是加密后的密文，s是密钥.对应上例是$s = 3$.

而实际上用数字来取代英文字母,并不一定要顺着次序,对照表也可无规率和自定的.但是只要知道明码和密码的对照表,也即明码和密码的转换方法,那么能破解密码所包含的讯息了.那么如何根据所得到的密文进行解密呢？如果我们知道了s（移位因子），那么逆推就可以了。这里的s就是凯撒密码的密钥。逆推公式为：

$$C - s \equiv P(\mod 26) \tag{1.2}$$

但是对于解密方而言，前提是在知道采用何种加密形式加密的情况下，如果不知道密钥，那么如何进行处理？这时一般情况有两种办法。一种是穷举法。也就是把s的值逐个带入：$1, 2, ..., 25$，最后根据所得到的字义结果来进行判断。这种办法对于直接平移的有效，但是对于没有特定次序的，而需要根据明码和密码对照表来确定的密码，就无能为力了。在这

种情况下，如何处理？这时可以借助于频率解析法来进行处理。

频率解析法(frequency analysis)是根据英文字母在明码资料中出现的个别频率而来的.通常各英文字母出现的频率百分比形成下表:

Table 1: 频率公式

a	0.0856	b	0.0139	c	0.0279	d	0.0378
e	0.1304	f	0.0289	g	0.0199	h	0.0528
i	0.0627	j	0.0013	k	0.0042	l	0.0339
m	0.0249	n	0.0707	o	0.0797	p	0.0199
q	0.0012	r	0.0677	s	0.0007	t	0.1045
u	0.0249	v	0.0092	w	0.0149	x	0.0017
y	0.0199	z	0.0008				

计算密文资料中每个字母出现的频率，在截获信息足够多的情况下，因为密钥只有一个，则可以根据出现频繁的多少来确定其对应的明文字母.

仿射变换密码

$$c \equiv ap + b(\mathsf{mod}26) \tag{1.3}$$

这里要求自然数a与26必须互素，也就是与a的最大公约数必须为1，那么a的取值共有12种可能的取法，而b共有26 种取法，这样由明文到密文的变换共有312种。对于这种变换，可以通过

$$c - b \equiv ap(\mathsf{mod}26) \Rightarrow a^{-1}(c-b) \equiv p(\mathsf{mod}26) \tag{1.4}$$

进行解密。只需要两边同时乘以a关于26的逆。这里其逆可以定义为如下形式：

$$a^{-1}a \equiv 1(\mathsf{mod}26) \tag{1.5}$$

实际上，是通过如下公式：$a*x-k*26=1$,这里x就定义为a^{-1}.对于该值，可以通过辗转相除法得到。在这种变换下，相继的明文字母对应着间隔为a的密文字母。

对应这种密文的解密，一般情况下可以根据密文中的字母的频率，假设密文中出现的频率最高的字母对应于英文中最常见的字母。例如在一则消息中，Z 出现14次，B出现12次，V11次，U10次，T10次，Y9次。我们假设Z-E,B-T,则有下面两个同余公式，

$$26 \equiv 5a + b(\mathsf{mod}26), 2 \equiv 20a + b(\mathsf{mod}26) \tag{1.6}$$

两式相减，可得$24 \equiv -15a(\mathsf{mod}26)$ ，它相当于$24 \equiv 11a(\mathsf{mod}26)$ ，因为11关于26的逆是19，则$a \equiv 19*24(\mathsf{mod}26) \equiv 14(\mathsf{mod}26)$，但是14不满足与26互素的条件。所以在这种情况下求得的结果不是我们希望的结果。

再做另外的尝试，V-T,B-E,则有下面两个同余公式

$$22 \equiv 20a + b(\mathsf{mod}26), 2 \equiv 5a + b(\mathsf{mod}26) \tag{1.7}$$

两式相减，可得$23 \equiv 15a(\mathsf{mod}26)$ ，因为15关于26的逆是7，则$a \equiv 7 * 23 \equiv 5(\mathsf{mod}26)$，这里5满足与26互素的条件，近一步可得$b \equiv 2 - 5 * 5 \equiv 3(\mathsf{mod}26)$.在这种情况下求得的结果是可能的一种结果，把这个结果带入到其他的密文中去，即可以求得其他的对应的明文。无论如何，应用频率方法是一种有效的针对上述密码的求解方法。其关键点在于对于同一个明文对应的密文保持不变。

三重图密码

对于我们刚才所讲过的两种加密系统，无论是什么情况，都可以根据密文中字母的频率，重复模式以及字与字之间的组合的方式来进行计算解密。关键在于明文的一个字母在密文中总用相同的密文字母来表示。

防止使用频率解密的加密方式之一就是每次加密一组字母而不是加密单个字母。这样使得原来的同样的字母在密文中就会以不同的方式来表达。为了避免出现像上面讨论的素数的限制，我们模的数就取一个素数29（加上逗号、空格、问号）。下面我们通过一个具体的实例对该加密方式进行表达：我们考虑三重图系统：例如对明文ADD加密，这几个数等价于1，4，4，则矩阵左乘以矩阵

$$M = \begin{pmatrix} 0 & 2 & 3 \\ 1 & 4 & 7 \\ 2 & 3 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 20 \\ 45 \\ 38 \end{pmatrix} \equiv \begin{pmatrix} 20 \\ 16 \\ 9 \end{pmatrix} (\mathsf{mod}29) \tag{1.8}$$

这样明文ADD对应的是密文TPI.这样就说明了在多重图加密系统中，相同的明文也许对应着不同的密文。加密的过程很简单，只需要找到一合适的加密矩阵。这一矩阵要每个元素求必须是整数。这样如果需要对该系统进行解密，也就是如何根据密文计算出明文。关键在于需要计算出该矩阵的逆。

欲求的矩阵若是3阶矩阵，则未知的量有9个，这样就需要9个已知的信息。下面以一个二阶矩阵的逆的计算来说明一下如何计算：

假设已给的信息，即对应的明文和密文的信息为：$(go\ qb) \leftrightarrow (de\ ar)$, 对应的数字应该是

$$\begin{pmatrix} 4 & 5 \\ 1 & 18 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 7 & 15 \\ 17 & 2 \end{pmatrix} \tag{1.9}$$

$$\begin{pmatrix} 4 & 5 & 7 & 15 \\ 1 & 18 & 17 & 2 \end{pmatrix} \Leftrightarrow (7^{-1} \equiv 15(\mathsf{mod}26)) \begin{pmatrix} 105 & 225 & 60 & 75 \\ 60 & 75 & 1 & 18 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 17 & 8 & 23 \\ 17 & 2 & 1 & 18 \end{pmatrix} \tag{1.10}$$

$$\Leftrightarrow \begin{pmatrix} 1 & 17 & 8 & 23 \\ 0 & -287 & -135 & -373 \end{pmatrix} \Leftrightarrow \begin{pmatrix} 1 & 17 & 8 & 23 \\ 0 & 25 & 21 & 17 \end{pmatrix} \Leftrightarrow (25^{-1} \equiv 25(\mathsf{mod}26)) \tag{1.11}$$

$$\begin{pmatrix} 1 & 17 & 8 & 23 \\ 0 & 625 & 525 & 425 \end{pmatrix} (\mathsf{mod}\ 26) \Leftrightarrow \begin{pmatrix} 1 & 0 & -77 & -130 \\ 0 & 1 & 5 & 9 \end{pmatrix} (\mathsf{mod}26) \Leftrightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 5 & 9 \end{pmatrix} (\mathsf{mod}26) \tag{1.12}$$

这样就可以计算出方程的逆矩阵，上述矩阵的右侧。这样就可以对原来的密文直接破译。对于以上的密码设计，只要公开密钥，就可以对以上的密码进行解密。

公开密钥机制

在1976年，W.Diffie,M.Hellman提出了一个双密钥系统的设想。在信息传递过程中设置两个密钥，一个是加密的，一个是解密的。虽然这两个密钥是互逆的。但是很难从加密的密钥计算出解密密钥的方法，从而即使将加密的密钥公开也不会对解密的密钥产生威胁。RSA加密系统是世界上第一套符合公开钥匙加密条件的密码系统,于1977年由瑞维斯特(R. Rivest),薛米尔(A. Shamir) 及艾多曼(L. Adleman)三人所研发出来,故称为RSA。它的安全性基于下面这样一个假设：一个包含有两个足够大的素数因子所组成的合数的因子分解在当前的计算机技术下，其计算机机时的消耗是相当可观的。

加密一个明文，公开钥匙n是由两个质数p与q相乘而得.而p和q只有持有人才得知.若想强行破解,则须先将n进行质因数分解才行,但因为当p 和q 的值越大时,则质因数分解所耗用的时间会成指数增长；和另外一个数e,它与数$r = (p-1)(q-1)$ 互素且满足$2^e > n$.

加密时,须先将文字讯息转换为十进制的值。将M的编码数字e 次幂并且求出它的关于n的模,其加密公式如下$M^e \equiv C(\mathsf{mod}n)$.

如果要想对该密码进行解密，关键首先要计算一个d 值，使得$ed \equiv 1(\mathsf{mod}r)$，接着可以利用下面的公式$C^d \equiv M(\mathsf{mod}n)$,通过这个公式就可以达到对原来的密码进行解密的目的。

这样对于RSA加密体系而言，公开密钥是n, e, 解密密钥是r, d.

例如：利用RSA系统为“phone”进行加密. 首先,利用对照表把它转化为数字

$$1608151405 \tag{1.13}$$

然后选取p和q的值,为简化计算,才用较小的数值.

设公开钥匙为$(n, e) = (247, 31)$,则$n = 13 * 19 = 247, r = 12 * 18 = 216$.我们便能把假

设$M < n, M_1 = 16, M_2 = 08, M_3 = 15$, 通过$ed \equiv 1(\mathsf{mod}r)$,可以计算$d = 7$

$$c_1 \equiv M_1^e \equiv 16^{31} \equiv 81(\mathsf{mod}247), M_1 \equiv 81^7 \equiv 16(\mathsf{mod}247) \tag{1.14}$$

$$c_2 \equiv M_2^e \equiv 8^{31} \equiv 122(\mathsf{mod}247), M_2 \equiv 122^7 \equiv 8(\mathsf{mod}247) \tag{1.15}$$

$$c_3 \equiv M_3^e \equiv 15^{31} \equiv 219(\mathsf{mod}247), M_3 \equiv 219^7 \equiv 15(\mathsf{mod}247) \tag{1.16}$$

1977年，这一密码体制的发明者以两个素数相乘，得到一个129位的数n，他们认为破译这一密码需要4亿亿年，尽管这一工作非常不易，在贝尔通讯公司的一位科研人员的协调下，利用internet,五大洲600余人使用1600多台计算机，历时八个月在1994年得以破译：这些魔文是容易受惊的鱼鹰.

1999年,RSA-155(155位)被成功分解,花了五个月时间在一台有3.2G内存的Cray C916电脑上完成。现在的超级电脑即使有1000GHz,要破解1024位的RSA加密系统,都要很多年,甚至比地球还要长寿.除非有其他更有效率的质数分解法,否则在地球毁灭之时,也未能计算出来.密码学专家对RSA也束手无策,可见,一般人要破解RSA,几乎是没有可能.所以,这套技术是可靠的,而网上的电子交易,都进行了加密,因此应该也是可靠的.