第 3 章　对称密码技术

本章导读

- 本章主要介绍对称密码技术及其相关的内容，包括一些加密算法、密钥的产生和密钥的分配等。
- 对称密码是一种加密密钥和解密密钥相同的密码体制。
- 对称密码分为分组密码和流密码。分组密码每次操作(如加密和解密)是针对一个分组而言。流密码则每次加密(或者解密)一位或者一个字节。
- 对密码的攻击方法有基于密码算法性质的密码分析和穷举搜索攻击。
- 对称密码主要分两个阶段：20 世纪 70 年代以前的对称密码(主要指计算机出现以前)和 20 世纪 70 年代以后的对称密码。
- 20 世纪 70 年代以前的对称密码我们称为古典加密技术，主要使用代换或者置换技巧。20 世纪 70 年代以后的对称密码则同时使用代换和置换技巧。
- 古典加密技术分为两类：一类是单字母代换密码，它将明文的一个字符用相应的一个密文字符代替。另一类是多字母代换密码，它是对多于一个字母进行代换。单字母代换密码又分为单表代换密码和多表代换密码。
- DES 是第一个加密标准，它与古典加密技术不一样，DES 同时使用了代换和置换两种技巧。用 56 位密钥加密 64 位明文。
- AES 是用来取代 DES 的高级加密标准，其结构与 DES 不同，它是用 128、192 或者 256 位密钥加密 128 位的分组。
- RC6 是 RSA 公司提交给 NIST 的一个候选高级加密标准算法，其效率非常高。
- RC4 是广泛使用的一种同步流密码。
- 在密码学中很多场合都要使用随机数，安全的随机数应该满足随机性和不可预测性。
- 密钥分配为通信的双方发送会话密钥。

密码技术是信息系统最重要的安全机制。密码技术主要分为对称密码技术(也称单钥或者传统密码技术)和非对称密码技术(也称双钥或者公钥密码技术)。在对称密码技术中，加密密钥和解密密钥相同，或者一个密钥可以从另一个导出。而非对称密码技术则使用两个密钥，加密密钥和解密密钥不相同。对称密码技术主要使用两种技巧：代换和置换。代换是将明文中的每个元素映射成另一个元素。置换是将明文中的元素重新排列。在 20 世纪 70 年代以前的加密技术都是对称加密技术，并且在这些加密技术只使用了代换或者置换技巧。这个时期的加密技术也称为古典加密技术。在 20 世纪 70 年代以后出现的对称加密技术则同时使用了代换和置换两种技巧。这两个阶段的加密技术还有一个典型区别是，古典加密技术一般将加密算法保密，而现代的对称加密技术则公开加密算法，加密算法的安全性只取决于密钥，不依赖于算法。非对称密码技术则产生于 20 世纪 70 年代。

3.1 基本概念

密码学(Cryptology)是以研究秘密通信为目的,即对所要传送的信息采取一种秘密保护,以防止第三者对信息的窃取的一门学科。密码学作为数学的一个分支,包括密码编码学(Cryptography)和密码分析学(Cryptanalysis)两部分。密码编码学是研究加密原理与方法,使消息保密的技术和科学,它的目的是掩盖消息内容。密码分析学则是研究破解密文的原理与方法。密码分析者(Cryptanalyst)是从事密码分析的专业人员。

采用加密的方法伪装消息,使得未授权者不可理解。被伪装的原始的消息(Message)称为明文(Plaintext)。将明文转换为密文过程称为加密(Encryption),加了密的消息称为密文(Ciphertext),而把密文转变为明文的过程称为解密(Decryption)。加密解密过程如图 3.1 所示。从明文到密文转换的算法称为密码(Cipher)。把一个加密系统采用的基本工作方式叫做密码体制(Cryptosystem)。实际上在密码学中见到"系统或体制"(System)、"方案"(Scheme)和"算法"(Algorithm)等术语本质上是一回事,在本书中我们也使用这些术语。加密和解密算法通常是在一组密钥(Key)控制下进行的,分别称为加密密钥和解密密钥。如果加密密钥和解密密钥相同,则密码系统为对称密码系统。

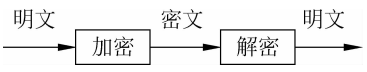

图 3.1 加密解密过程

3.2 对称密码模型

对称密码也称传统密码,它的特点是发送方和接收方共享一个密钥。对称密码分为两类:分组密码(Block Ciphers)和流密码(Stream Ciphers)。分组密码也称为块密码,它是将信息分成一块(组),每次操作(如加密和解密)是针对一组而言。流密码也称序列密码,它是每次加密(或者解密)一位或者一个字节。

一个对称密码系统(也称密码体制)由 5 个组成部分组成。一用数学符号来描述为 $S=\{M,C,K,E,D\}$,如图 3.2 所示。

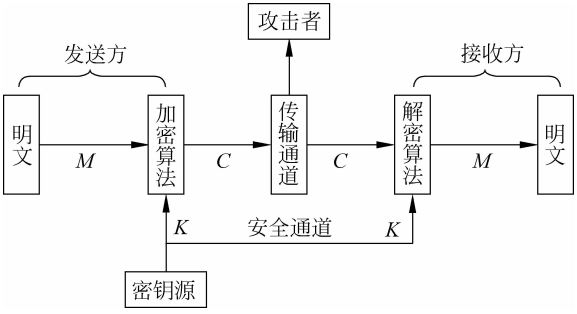

图 3.2 对称密码系统模型

① 明文空间 M,表示全体明文的集合。

② 密文空间 C,表示全体密文的集合。

③ 密钥空间 K,表示全体密钥的集合,包括加密密钥和解密密钥。

④ 加密算法 E,表示由明文到密文的变换。

⑤ 解密算法 D,表示由密文到明文的变换。

在发送方,对于明文空间的每个明文,加密算法在密钥的作用下生成对应的密文。接收方将接收的密文,用解密算法在解密密钥的控制下变换成明文。我们可以看到加密算法有两个输入,一个是明文,另一个是密钥。加密算法的输出是密文。解密算法本质上是加密算法的逆运行,解密算法的输入是密文和密钥,输出是明文。

对明文 M 用密钥 K,使用加密算法 E 进行加密常常表示为 $E_k(M)$,同样用密钥 K 使用解密算法 D 对密文 C 进行解密表示为 $D_k(C)$。在对称加密体制中,解密密钥相同,有

$$C = E_k(M)$$

$$M = D_k(C) = D_k(E_k(M))$$

从对称密码模型可以看到,发送方和接收方主要进行加密和解密运算,我们希望这个运算越容易越好,对于攻击者而言,我们希望他们破译密文的计算越难越好。因此一个好的密码体制至少满足下面的条件。

① 已知明文 M 和加密密钥 K 时,计算 $C=E_k(M)$容易。

② 加密算法必须足够强大,使破译者不能仅根据密文破译消息,即在不知道解密密钥 K 时,由密文 C 计算出明文 M 是不可行的。

③ 由于对称密码系统双方使用相同的密钥,因此还必须保证能够安全地产生密钥,并且能够以安全的形式将密钥分发给双方。

④ 对称密码系统的安全只依赖于密钥的保密,不依赖于加密和解密算法的保密。

3.3　密 码 攻 击

分析一个密码系统是否是安全,一般是在假定攻击者知道所使用的密码系统情况下进行分析的。一般情况下,密码分析者可以得到密文,知道明文的统计特性,加密体制,密钥空间及其统计特性,但不知道加密截获的密文所用的特定密钥。这个假设称为 Kerckhoff 假设。分析一个密码系统的安全性一般是建立在这个假设的基础上。当然,如果攻击者不知道所使用的密码体制,那么破译是更难的。但是,不应当把密码系统的安全性建立在攻击者不知道所使用的密码体制这个前提之下。因此,在设计一个密码系统时,其目的应当是在 Kerckhoff 假设下达到一定的安全程度。

攻击对称密码体制有两种方法:密码分析和穷举攻击(Brute Force Search)。密码分析是依赖加密算法的性质和明文的一般特征等试图破译密文得到明文或试图获得密钥的过程。穷举攻击则是试遍所有可能的密钥对所获密文进行解密,直至得到正确的明文;或者用一个确定的密钥对所有可能的明文进行加密,直至得到所获得的密文。

3.3.1　穷举攻击

穷举攻击是最基本也是比较有效的一种攻击方法。从理论上讲,可以尝试所有的密钥。

因此只要有足够的资源，任何密码体制都可以用穷举攻击将其攻破。幸运的是，攻击者不可能有无穷的可用的资源。

穷举攻击的代价与密钥大小成正比。穷举攻击所花费的时间等于尝试次数乘以一次解密(加密)所需的时间。显然可以通过增大密钥位数或加大解密(加密)算法的复杂性来对抗穷举攻击。当密钥位数增大时，尝试的次数必然增大。当解密(加密)算法的复杂性增大时，完成一次解密(加密)所需的时间增大。从而使穷举攻击在实际上不能实现。表 3.1 是穷尽密钥空间所需的时间。从表中我们可以发现，当密钥长度达到 128 位以上时，以目前的资源来说，穷举攻击将不成功。

表 3.1 穷尽密钥空间所需的时间

密钥长度(位)	密 钥 数 目	尝试 1 次/微秒 所需时间	尝试 10^6 次/微秒 所需时间
32	$2^{32}=4.3\times10^9$	2^{31} 微秒=35.8 分	2.15 毫秒
56	$2^{56}=7.2\times10^{16}$	2^{55} 微秒=1142 年	10.01 小时
128	$2^{128}=3.4\times10^{38}$	2^{127} 微秒=5.4×10^{24} 年	5.4×10^{18} 年
168	$2^{168}=3.7\times10^{50}$	2^{167} 微秒=5.9×10^{36} 年	5.9×10^{30} 年
26 个字母排列	$26!=4\times10^{26}$	2×10^{26} 微秒=6.4×10^{12} 年	6.4×10^{6} 年

3.3.2 密码攻击类型

密码分析是基于 Kerckhoff 假设。密码分析者所使用的策略取决于加密方案的性质以及可供密码分析者使用的信息，正是基于密码分析者所知的信息量，可把对密码的攻击分为以下几种类型。

- 唯密文攻击(Ciphertext-Only Attack)。密码分析者有一些消息的密文，这些消息都用同一算法加密。密码分析者的任务是恢复尽可能多的明文，或者是最好能推算出加密消息的密钥来，以便采用相同的密钥解出其他被加密的消息。这种情况下，密码分析者知道的东西只有两样：加密算法和待破译的密文。
- 已知明文攻击(Known-Plaintext Attack)。密码分析者除知道加密算法和待破译的密文外，而且也知道有一些明文和同一个密钥加密的这些明文所对应的密文，即知道一定数量的明文和对应的密文。
- 选择明文攻击(Chosen-Plaintext Attack)。密码分析者知道加密算法和待破译的密文，并且可以得到所需要的任何明文所对应的密文，这些明文和待破译的密文是用同一密钥加密得来的，即知道选择的明文和对应的密文。如在公钥密码体制中，攻击者可以利用公钥加密他任意选定的明文。
- 选择密文攻击(Chosen-Ciphertext Attack)。密码分析者知道加密算法和待破译的密文，密码分析者能选择不同的被加密的密文，并可得到对应的解密的明文，即知道选择的密文和对应的明文。解密这些密文所使用的密钥与解密待破解的密文的密钥是一样的。这种攻击主要用于公钥密码算法。
- 选择文本攻击(Chosen Text Attack)。选择文本攻击是选择明文攻击和选择密文攻击的结合。密码分析者知道加密算法和待破译的密文，并且知道任意选择的明文和它对应的密文，这些明文和待破译的密文是用同一密钥加密得来的，以及有目的选

择的密文和它对应的明文，解密这些密文所使用的密钥与解密待破解的密文的密钥是一样的。

在以上任何一种情况下，攻击者的目标是为了确定正在使用的密钥。显然，上述 5 种攻击类型的强度按序递增，如果一个密码系统能够抵抗选择明文攻击，那么它也能抵抗唯密文攻击和已知明文攻击，一般来说，一个密码体制是安全的，通常是指前三种攻击下的安全性，即攻击者一般容易具备前三种攻击条件。在这几种攻击类型中，唯密文攻击难度最大，因为攻击者可利用的信息最少。在此情况下一种可能的攻击方法是对所有可能的密钥尝试的强行攻击法，即穷举攻击。如果密钥量非常大，则该方法是不现实的。因此，攻击者通常运用各种统计方法对密文本身的分析。如果攻击者知道的信息越多，就越容易破解密文。在多数情况下，密码分析者能够获得除密文外的更多信息，如能够获得一段或者多段明文以及对应的密文，或者可能知道某种明文模式将出现在某个消息中，此时可以进行已知明文攻击，攻击者可以从转换明文的方法来推导密钥。

对密码设计者而言，被设计的加密算法一般要能经受得住已知明文的攻击。如果无论攻击者有多少密文，由一个加密算法产生的这些密文中包含的信息不足以唯一决定对应的明文，也无论用什么技术方法进行攻击都不能被攻破，这种加密算法是绝对安全(Unconditional Security)。绝对安全指不论攻击者具有多少计算能力都无法破解密文。除一次一密(One-Time Pad)外，没有绝对安全的加密算法。因此，加密算法的使用者应该挑选满足下列标准中的一个或两个的算法。

(1) 破译该密码的成本超过被加密信息的价值。

(2) 破译该密码的时间超过该信息有用的生命周期。

如果满足上述的两个准则，一个加密算法就可认为是在计算上安全(Computational Security)的。计算上安全是指在计算能力有限的情况下(如计算所需时间比宇宙生存时间还长)，无法破解此密文。目前的加密算法一般在计算上是安全的。

3.3.3　密码分析方法

当密钥长度增加到一定的大小时，穷举攻击变得不实际。因此用密码分析的方法攻击密码越来越引起人们的重视，目前比较流行的密码分析方法是线性密码分析和差分密码分析。这两种方法主要是针对现代密码的攻击。

线性分析是一种已知明文攻击，最早由 Matsui 在 1993 年提出。线性分析是一种统计攻击，它以求线性近似为基础。通过寻找现代密码算法变换的线性近似来攻击。如用这种方法只需要知道 2^{43} 个已知明文的情况下就可以找到 DES 的密钥。

差分密码分析在许多方面与线性密码分析相似，它与线性密码分析的主要区别在于差分密码分析包含了将两个输入的异或与其相对应的两个输出的异或相比较。差分密码分析也是一个选择明文攻击。差分密码分析被公认为尽年来密码分析最大的成就。差分密码分析出现于 20 世纪 70 年代，但在 1990 年才公开发表。它的基本思想是：通过分析明文对的差值与密文对的差值的影响来恢复某些密钥位。差分分析可用来攻击任何一个固定迭代轮函数结构的密码算法。

3.4 古典加密技术

古典加密技术主要使用代换或者置换技术。代换(Substitution)是将明文字母替换成其他字母、数字或者符号。置换(Permutation)则保持明文的所有字母不变,只是打乱明文字母的位置和次序。这些古典代换加密技术分为两类,一类是单字母代换密码(Monogram Substitution Cipher),它将明文的一个字符用相应的一个密文字符代替。另一类是多字母代换密码(Polygram Substitution Cipher),它是对多于一个字母进行代换。在单字母代换密码中又分为单表代换密码(Monoalphabetic Substitution Cipher)和多表代换密码(Polyalphabetic Substitution Cipher)。单表代换密码只使用一个密文字母表,并且用密文字母表中的一个字母来代替一个明文字母表中的一个字母。多表代换密码是将明文消息中出现的同一个字母,在加密时不是完全被同一个固定的字母代换,而是根据其出现的位置次序,用不同的字母代换。

3.4.1 单表代换密码

单表代换密码只使用一个密文字母表,并且用密文字母表中的一个字母来代替一个明文字母表中的一个字母。设 M 和 C 分别表示为含 n 个字母的明文字母表和密文字母表。

$$M = \{m_0, m_1, \cdots, m_{n-1}\}$$

$$C = \{c_0, c_1, \cdots, c_{n-1}\}$$

如果 f 为一种代换方法,那么密文为 $C=E_k(m)=c_0c_1\cdots c_{n-1}=f(m_0)f(m_1)\cdots f(m_{n-1})$。

单表代换密码常见的方法有加法密码、乘法密码和仿射密码。在本章的例子中,我们将用小写字母表示明文,用大写字母表示密文。明文和密文空间都假设为 26 个字母,即属于 Z_{26},当然很容易推广到 n 个字母的情况。

1. 加法密码

对每个 $c, m \in Z_n$,加法密码的加密和解密算法是

$$C = E_k(m) = (m + k) \bmod n$$

$$M = D_k(c) = (c - k) \bmod n$$

k 是满足 $0<k<n$ 的正整数。若 n 是 26 个字母,加密方法是用明文字母后面第 k 个字母代替明文字母,因此,代换密码中的加密和解密可以看做是字母表上的一个字母的置换。Caesar 密码是典型的加法密码。

Caesar 密码是已知最早的单表代换密码,采用加法加密的方法,由 Julius Caesar 发明,最早用在军方。将字母表中的每个字母,用它后面的第 3 个字母代替,如下:

明文:meet me after the toga party

密文:PHHW PH DIWHU WKH WRJD SDUWB

可将代换方式定义如下:

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

让每个字母等价一个数字如下：

a	b	c	d	e	f	g	h	i	j	k	l	m	n
0	1	2	3	4	5	6	7	8	9	10	11	12	13

o	p	q	r	s	t	u	v	w	x	y	z
14	15	16	17	18	19	20	21	22	23	24	25

对每个明文字母 m，用密文字母 c 代换，那么 Caesar 密码算法如下。

加密：$C=E(m)=(m+3) \bmod 26$

解密：$M=D(c)=(c-3) \bmod 26$

移位可以是任意的，如果用 $k(1\leqslant k\leqslant 25)$ 表示移位数，则通用的 Caesar 密码算法表示如下。

加密：$C=E_k(m)=(m+k) \bmod 26$

解密：$M=D_k(c)=(c-k) \bmod 26$

Caesar 密码安全性分析如下：

前面我们已经介绍过，对密码的分析是基于 Kerckhoff 假设。因此假设攻击者知道使用 Caesar 密码加密。如果攻击者只知道密文，即唯密文攻击，只要穷举测试所有可能字母移位的距离，最多尝试 25 次。实际上攻击者为了加快穷举速度，只要对密文中一个单词进行猜想解密，就可以加快判断密钥的正确性。如果攻击者知道一个字符以及它对应的密文，即已知明文攻击，那么攻击者很快就通过明文字符和对应的密文字符之间的距离推出密钥。这个例子说明一个密码体制安全至少要能够抵抗穷举密钥搜索攻击，普通的做法是将密钥空间变得足够大。但是，很大的密钥空间并不是保证密码体制安全的充分条件，下面的例子可以说明这一点。

我们对 Caesar 密码进行改进，假设密文是 26 个字母的任意代换，密钥是明文字母到密文字母的一个字母表，密钥长度是 26 字长。

例如字母代换表如下：

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

若加密的明文为 ifwewishtoreplaceletters，那么对应的密文为 WIRFRWAJUHYFTSDVFSFUUFYA。

上面的字母代换表由通信双方事先设计好，一个更实际的构造字母代换表的方法是使用一个密码句子。如密钥句子为 the message was transmitted an hour ago，按照密钥句子中的字母依次填入字母表(重复的字母只用一次)，未用的字母按自然顺序排列。这样可以构造如下的字母代换表。

原字母表如下：

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

代换字母表如下：

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

若明文为 please confirm receipt，使用上面的代换字母表，则密文为 CDSTKSEBUARJOJSESRCL。

使用上面的方法代换，总共有 $26!=4\times10^{26}$ 种密钥，从表 3.1 可以看到穷举搜索这么多的密钥很困难。但这并不表示该密码不容易破解。破解这类密码的突破点是由于语言本身的特点是充满冗余的，每个字母使用的频率不相等。由于上面的加密后的密文实际上是明文字母的一个排列，因此单表代换密码没有改变字母相对出现的频率，明文字母的统计特性在密文中能够反映出来，即保持明文的统计特性不变。当通过统计密文字母的出现频率，可以确定明文字母和密文字母之间的对应关系。英文字母中单字母出现的频率如图 3.3 所示。

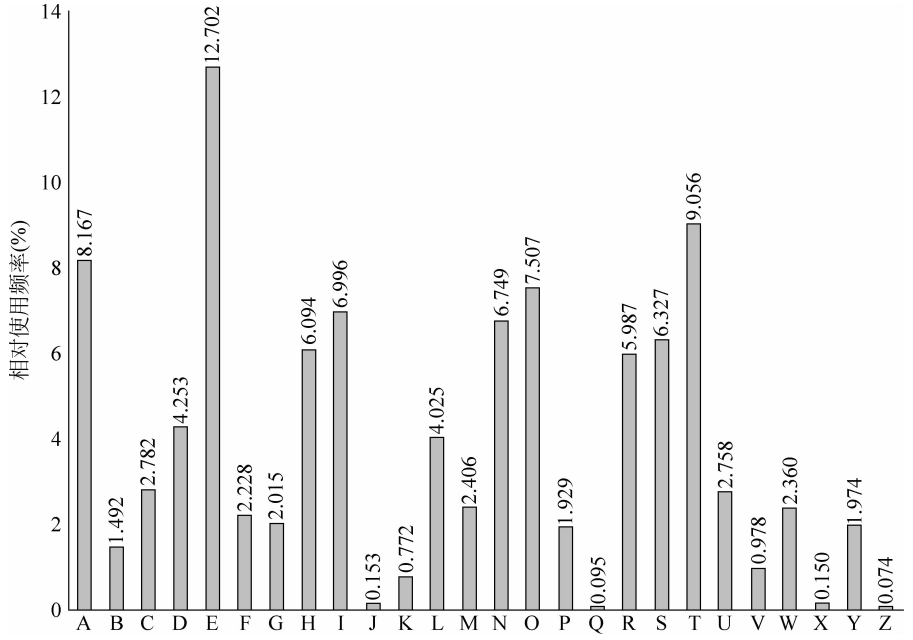

图 3.3 英文字母中单字母出现的频率

图 3.3 中的 26 个字母按照出现频率的大小可以分为下面 5 类。

(1) e：出现的频率大约为 0.127。

(2) t,a,o,i,n,s,h,r：出现的频率大约在 0.06～0.09 之间。

(3) d,l：出现的频率约为 0.04。

(4) c,u,m,w,f,g,y,p,b：出现的频率大约在 0.015～0.028 之间。

(5) v,k,j,x,q,z：出现的频率小于 0.01。

双字母和三字母组合都有现成的统计数据，常见的双字母组合和三字母组合统计表能够帮助破解密文。

出现频率最高的 30 个双字母(按照频率从大到小排列)如下：

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 20 个三字母(按照频率从大到小排列)如下：

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

例 3.1　已知下面的密文是由单表代换产生的：

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZVUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSXEPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ

试破译该密文。

首先统计密文中字母出现的频率，然后与英文字母出现频率比较。密文中字母的相对频率统计如下。

字母	次数	频率(%)	字母	次数	频率(%)	字母	次数	频率(%)	字母	次数	频率(%)
A	2	1.67	H	7	5.83	O	9	7.50	V	5	4.17
B	2	1.67	I	1	0.83	P	16	13.33	W	4	3.33
C	0	0.00	J	1	0.83	Q	3	2.50	X	5	4.17
D	6	5.00	K	0	0.00	R	0	0.00	Y	2	1.67
E	6	5.00	L	0	0.00	S	10	8.33	Z	14	11.67
F	4	3.33	M	8	6.67	T	3	2.55	V	5	4.17
G	2	1.67	N	0	0.00	U	10	8.33	W	4	3.33

将统计结果与图 3.3 进行比较，可以猜测密文中 P 与 Z 可能是 e 和 t，密文中的 S，U，O，M 出现频率比较高，可能与明文字母中出现频率相对较高的 a，o，i，n，s，h，r 这些字母对应。密文中出现频率很低的几个字母 C，K，L，N，R，I，J 可能与明文字母中出现频率较低的字母 v，k，j，x，q，z 对应。就这样边试边改，最后得到明文如下：

it was disclosed yesterday that several informal but direct contacts have been made with political representatives of the viet cong in moscow

在尝试过程中，如果同时使用双字母和三字母的统计规律，那么更容易破译密文。如上面的密文中出现最多的双字母是 ZW，它可能对应明文双字母出现频率较大的 th，那么 ZWP 就可能是 the，这样就更容易试出明文。

2. 乘法密码

对每个 $c, m \in Z_n$，乘法密码的加密和解密算法是

$$C = E_k(m) = (mk) \bmod n$$

$$M = D_k(c) = (ck^{-1}) \bmod n$$

其中 k 和 n 互素，即 $\gcd(k,n)=1$，否则不存在模逆元，不能正确解密。显然乘法密码的密码空间大小是 $\varphi(n)$，$\varphi(n)$是欧拉函数。可以看到乘法密码的密钥空间很小，当 n 为 26 字母，则与 26 互素的数是 1、3、5、7、9、11、15、17、19、21、23、25，即 $\varphi(n)=12$ 因此乘法密码的密钥空间为 12。

乘法密码也称采样密码，因为密文字母表是将明文字母按照下标每隔 k 位取出一个字母排列而成。

例 3.2　英文字母，选取密码为 9，使用乘法密码的加密算法，那么明文字母和密文字母

的代换表构造如下。

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

若明文为 a man liberal in his views，那么密文为 AENVUJKXUNLUGHUKQG。

3. 仿射密码

加法密码和乘法密码结合就构成仿射密码，仿射密码的加密和解密算法是

$$C = E_k(m) = (k_1 m + k_2) \bmod n$$

$$M = D_k(c) = k_1^{-1}(c - k_2) \bmod n$$

仿射密码具有可逆性的条件是 $\gcd(k,n)=1$。当 $k_1=0$ 时，仿射密码变为加法密码，当 $k_2=0$ 时，仿射密码变为乘法密码。

仿射密码中的密钥空间的大小为 $n\varphi(n)$，当 n 为 26 字母，$\varphi(n)=12$，因此仿射密码的密钥空间为 $12\times26=312$。

例 3.3 设密钥 $K=(7,3)$，用仿射密码加密明文 hot。

三个字母对应的数值是 7、14 和 19。分别加密如下：

$$(7\times7+3) \bmod 26 = 52 \bmod 26 = 0$$

$$(7\times14+3) \bmod 26 = 101 \bmod 26 = 23$$

$$(7\times19+3) \bmod 26 = 136 \bmod 26 = 6$$

三个密文数值为 0、23 和 6，对应的密文是 AXG。

例 3.4 假设获得仿射密码加密的密文是：

FMXVEDKAPHRERBNDKRXRSREFMORUD5DXYV5HVUPEDKAPRKDLYEVLRHHHRH

试破译该密码。

同样可以统计密文中各字母出现的频率，然后与英文字母出现频率比较，在尝试过程中同时要考虑仿射密码的条件。

各个字母出现的频率统计如下。

字母	频率(次数)	字母	频率(次数)	字母	频率(次数)	字母	频率(次数)
A	2	H	5	O	1	V	4
B	1	I	0	P	2	W	0
C	0	J	0	Q	0	X	2
D	7	K	5	R	8	Y	1
E	5	L	2	S	3	Z	0
F	4	M	2	T	0		
G	0	N	1	U	2		

这里虽然只有 57 个字母，但它足以分析仿射密码，最大频率的密文字母是 R(8 次)，D(7 次)，E、H、K(每个 5 次)和 S、F、V(各 4 次)。首先，我们可以猜想 R 是 e 的加密，而 D 是 t 的加密，因为 e 和 t 是两个出现频率最高的字母。e 和 t 对应的数值是 4 和 19，R 和 D 对应的数值是 17 和 3。对于仿射密码，有 $c=(k_1m+k_2)$。

所以我们有如下的关于两个未知数线性方程组：

$$17=4k_1+k_2$$

$$13=19k_1+k_2$$

这个方程组有唯一解 $k_1=6, k_2=19$，但这不是一个合法的密钥，因为 gcd(6,26)=2，不等于 1。

我们再猜测 R 是 e 的加密，而 E 是 t 的加密，继续使用上述的方法，得到 $k_1=13$，这也是一个不合法的密钥。再试一种可能性：R 是 e 的加密，H 是 t 的加密，则有 $k_1=8$，这也是不合法的。继续进行，我们猜测 R 是 e 的加密，K 是 t 的加密，这样可得 $k_1=3, k_2=5$，首先它至少是一个合法的密钥，下一步工作就是检验密钥 $K=(3,5)$ 的正确性。如果我们能得到有意义的英文字母串，则可证实是有效的。对密文进行解密有

algorithms are quite general definitions of arithmetic processes

3.4.2　多表代换密码

单表代换密码是将明文的一个字母唯一地代换为一个字母。加密后的密文具有明文的特征，通过统计密文中字母出现的频率能够比较方便地破解密文。要提高密码的强度，应该让明文结构在密文中尽量少出现。多表代换密码和多字母代换密码能够减少这种密文字母和明文字母之间的对应关系。这一节介绍多表代换密码，下一节将介绍多字母代换密码。

多表代换密码是对每个明文字母信息采用不同的单表代换，也就是用一系列(两个以上)代换表依次对明文消息的字母进行代换的加密方法。

如果明文字母序列为 $m=m_1m_2\cdots$，令 $f=f_1, f_2, \cdots$ 为代换序列，则对应的密文字母序列为

$$C=E_k(m)=f_1(m_1)f_2(m_2)\cdots$$

若代换系列为非周期无限序列则相应的密码为非周期多表代换密码。这类密码对每个明文字母都采用不同的代换表或密钥进行加密，称作是一次一密密码(One-Time Pad Cipher)。这是一种在理论上唯一不可破的密码，一次一密对于明文的特征可实现完全隐蔽，但由于需要的密钥量和明文消息长度相同而难以广泛使用。

实际中经常采用周期多表代换密码，它通常只使用有限的代换表，代换表被重复使用以完成对消息的加密。此时代换表系列为

$$f=f_1, f_2, \cdots, f_d, f_1, f_2, \cdots, f_d, \cdots$$

在对明文字母序列为 $m=m_1m_2\cdots$ 进行加密时，相应的密文字母系列为

$$C=E_k(m)=f_1(m_1)f_2(m_2)\cdots f_d(m_d)f_1(m_{d+1})f_2(m_{d+2})\cdots f_d(m_{2d})\cdots$$

当 $d=1$ 时，多表代换密码变为单表代换密码。

下面介绍一种比较有名的多表代换密码——维吉尼亚密码。