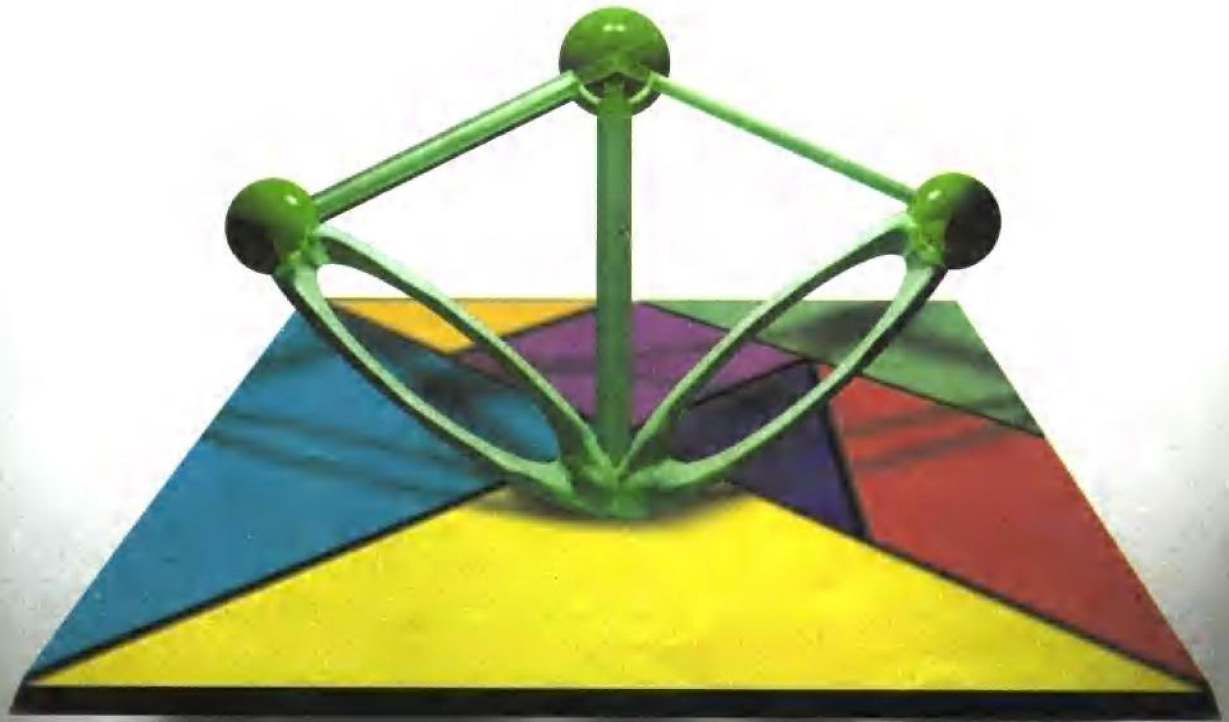
计算机科学组合学丛书

计算机密码学

——计算机网络中的数据保密与安全

（第2版）

卢开澄　编著



清华大学出版社　http//www.tup.tsinghua.edu.cn

计算机科学组合学系列丛书之四

计 算 机 密 码 学

——计算机网络中的数据保密与安全

（第 2 版）

卢开澄　编著

清 华 大 学 出 版 社

(京)新登字 158 号

内 容 简 介

本书是在 1990 年第一版的基础上改写而成的，共 8 章：传统密码与密码学基本概念；分组密码；公钥密码；与公钥密码有关的若干算法；密码学的信息论基础；线性反馈移位寄存器和序列密码；数字签名、Hash 函数、安全协议及其它；网络的安全保密。为了便于广大读者阅读，本书备有若干附录。其中包含 DES 和 IDEA 两个分组密码的源程序可供使用。

本书可作为计算机系或其它专业关于"网络通信保密安全"的教材或参考书。

图书在版编目(CIP)数据

计算机密码学：计算机网络中的数据保密与安全/卢开澄编著. —2 版. —北京：清华大学出版社，1998
ISBN 7-302-02783-8

Ⅰ. 计… Ⅱ. 卢… Ⅲ. 电子计算机-密码-理论 Ⅳ. TP309

中国版本图书馆 CIP 数据核字(97)第 29000 号

出版者：清华大学出版社(北京清华大学校内，邮编 100084)
http://www.tup.tsinghua.edu.cn
印刷者：北京市清华园胶印厂
发行者：新华书店总店北京发行所
开 本：787×1092 1/16 印张：19.75 字数：464 千字
版 次：1998 年 7 月第 2 版 1999 年 8 月第 3 次印刷
书 号：ISBN 7-302-02783-8/TP・1450
印 数：28001～33000
定 价：19.80 元

计算机科学组合学丛书

前　言

电子计算机的出现是20世纪的大事，它改变了我们这个世界的面貌。可以毫不夸张地说，它的影响遍及所有的角落，几乎无处不感觉到它的存在。数学更不例外。严格地说，电子计算机本身就是近代数学的辉煌成就。将计算机与数学割裂开来，既不合理也不可能。组合学也就是在计算机科学蓬勃发展的刺激下面崛起的，从而成为近若干年来最活跃的数学分支。它研究的问题有的可追溯到欧拉和哈密尔顿等18世纪的数学家，但它成为一新的分支还是近若干年的事。它从与计算机科学相结合中获得了广阔的发展空间，从而也为计算机科学奠定了理论基础。

什么是计算机科学呢？有的学者将它定义为研究算法的一门学科。研究算法无疑是计算机科学的重要领域，也是本丛书的核心内容，贯穿始终。组合学家在20世纪70年代初建立的算法复杂性“NP理论”，至今仍然令无数计算机科学工作者与数学工作者为之折腰。

计算机科学里的组合学内容十分广泛。本丛书涉及组合分析、图论、组合算法、近代密码学、组合优化、编码理论及算法复杂性等七部分。

组合分析是算法的理论基础。组合分析之与组合算法犹如数学分析之与计算数学，众所周知，前者是后者的理论根基。

图论原本是组合数学这个“家族”的主要成员，只因它已成长壮在，故自立门户独立出去。

算法复杂性的NP理论是近30年的一大成就。研究表明对于一类叫做NPC类的困难问题，至今都不存在有效算法，但它们难度相当，只要其中任何一个找到多项式解法，则全体都获得解决；或证明它们根本不存在有效办法。不论是前者还是后者都还看不见露到海平面上的桅杆塔，它吸引了众多的有志之士。密码学是其中十分引人入胜的分支。如若设计好的密码，对它的破译等价于某一NPC类困难问题，无疑这样的密码将是牢不可破的。

在计算机网络深入普及的信息时代，信息本身就是时间，就是财富。信息的传输通过是脆弱的公共信道，信息储存于“不设防”的计算机系统中，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，已成为当今被普遍关注的重大问题。密码是有效而且可行的办法。在计算机网络的刺激下，近代密码学便在算法复杂性理论的基础上建立起来了。密码作为一种技术，自从人类有了战争，不久便有了它。但作为一门学科则是近20多年的事。甚至于它已成为其它学科的基础。密码也从此走出“军营”，进入百姓家。

实际中的“优化”问题是大量的，半个多世纪以来它曾经几度辉煌。近来在计算机科学的影响下，又出现了若干闪光点，十分耀眼，引人注目。

实际上密码也是一种编码。如果说密码学研究的编码是保证通信的保密与安全，则编码理论研究的是通信中如何纠错与检错。计算机纠错码是既实用、理论上又饶有趣味的分支。

本丛书是作者在清华大学计算机科学与技术系长期工作的总结。它不是一部“长篇”记述，而是互相关联又彼此相对独立，因此难免有少量交叉。它们涉及的面如此之广，囿于作者的水平，缺点和错误在所难免，敬请读者不吝指正。谢谢。

前　言

自从人类有了战争，就有了密码，所以密码作为一种技术源远流长，可以追溯到远古时代，而且还有过自己的辉煌经历。但成为一门学科则是近20余年的事，这是受计算机科学蓬勃发展的刺激的结果。今天在计算机被广泛应用的信息时代，信息本身就是时间，就是财富。大量信息用数据形式存放在计算机系统里。信息的传输则通过公共信道。这些计算机系统和公共信道是不设防的，是很脆弱的，容易受到攻击和破坏，信息的丢失不容易被发现，而后果是极其严重的。如何保护信息的安全已不仅仅是军事和政府部门感兴趣的问题，其他企事业单位也愈感迫切。因为在网络化的今天，计算机犯罪每年使他们遭受的损失极其巨大，而且还在发展中。密码是有效而且可行的保护信息安全的办法，有效是指密码能做到使信息不被非法窃取，不被篡改或破坏，可行是说它需要付出的代价是可以接受的。

密码形成一门新的学科是在70年代。它的理论基础之一应该首推1949年Shannon的一篇文章“保密通信的信息理论”，这篇文章过了30年后才显示出它的价值。现在，密码学有了突飞猛进的发展，而且成为有些学科的基础。

在近代密码学上值得一书的大事有两件：一是1977年美国国家标准局正式公布实施了美国的数据加密标准(DES)，公开它的加密算法，并批准用于非机密单位及商业上的保密通信。密码学的神秘面纱从此被揭开。二是Diffie和Hellman联合写的一篇文章“密码学的新方向”，提出了适应网络上保密通信的公钥密码思想，掀起了公钥密码研究的序幕。受他们的思想启迪，各种公钥密码体制被提出，特别是RSA公钥密码的提出在密码学史上是一个里程碑。可以这么说：“没有公钥密码的研究就没有近代密码学。”

在密码学的发展过程中，计算机科学和数学工作者作出了卓越的贡献。数学中许多分支如数论、概率统计、近世代数、信息论、椭圆曲线理论、算法复杂性理论、自动机理论、编码理论等都可以在其中找到各自的位置。它的踪影遍及数学许多分支，而且还推动了并行算法的研究，从而成为近若干年来非常引人入胜的领域。但还应该强调指出的是密码学毕竟不等于就是数学，它还有自己的空间。

随着计算机网络不断渗透到各个领域，密码学的应用也随之扩大。数字签名、身份鉴别等都是由密码学派生出来的新技术和应用。

中国不能没有自己的密码系统，中国也必须有自己的数据加密标准。近年来，我国引进了很多设备，唯有密码设备不能依靠引进，开展这方面的研究是当务之急。

本书是作者在清华大学计算机系从事数据安全的教学科研基础上写成的。

第一章　传统密码与密码学基本概念

1.1　引　　论

密码学以研究秘密通信为目的。即研究对传输信息采取何种秘密的变换以防止第三者对信息的窃取。

在今天的信息社会里，通信安全保密问题的研究已不仅仅出于军事、政治和外交上的需要。科学技术的研究和发展及商业等方面，无一不与信息息息相关。所以信息就是生命，信息就是时间，信息就是财富。由于信息是共享的，信息的扩散会产生社会影响，所以保护信息的安全是信息时代的迫切需要。

保护信息的安全无疑是十分重要的，然而信息的丢失不容易被发现。同时它又是具有时间性的。同一信息在不同的时间里的价值也是不一样的，有时候获得信息的时间比信息本身还重要。

信息的存储和传输是通过载体进行的，例如，信使便是以人作为载体的。近代通信的载体有电波或电信号、磁盘等。

保密有载体保密和通信保密两种。密码学主要研究通信保密，而且仅限于数据通信保密。此外，还有语音保密和图象传输保密等。语音保密是研究电话通信的保密技术，不在本书讨论范围内。

近年来，密码学研究之所以十分活跃，主要原因是它与计算机科学的蓬勃发展息息相关。此外还由于电信事业以及防止日益严重的计算机犯罪的需要。由于公共和私人部门的一些机构愈来愈多地应用电子数据处理，将数据存储在数据库中，因此防止非法泄露、删除、修改等是必须正视的问题。特别是，电子资金传输系统是一个由通信网络互相联结的金融机构，并通过这种网络传输大量资金，这是密码通信通向民用的典型例子。因此，信息的安全性也成为全社会关心的问题，密码学从此也成为一门新的学科，引起了数学家和计算机科学工作者的日益浓厚的兴趣。

一般说来，由数据库收集或存储的大量数据，或在传输过程中的数据，由于传输中的公共信道和存储的计算机系统非常脆弱，容易受到两种形式的攻击：一种是从传输信道上截取信息，或从存储的载体上偷窃或拷贝信息，我们称之为被动攻击，其结果是导致数据的暴露和对私有权的侵犯；另一种是对在传输过程中或对存储的数据进行非法删除、更改或插入等操作，这称之为主动攻击，其结果可能引起数据或文件的混乱，严重时可能导致信息处理系统完全失控。对于这两种可能遭受到的攻击除了制定法律外，还需要有合适的保护措施。密码技术就是一种有效的方法。事实证明，这也是最经济可行的方法，它使得在一种潜在不安全的环境中保证通信的安全。正是因为密码对于通信安全的极端重要性，所以应该强调说，不安全的密码技术比没有还要坏。因为它给人们以安全的假象。

密码技术还有效地被用于信息鉴别、数字签名等，用以防止电子欺骗，这对信息处理

系统的安全起到极其重要的作用。

近代密码学研究并非是传统密码技术的旧话重提，它有其自己的特点。快速电子计算机和现代数学方法一方面为加密技术提供了新的概念和工具，另一方面也给破译者以有力武器。总之，较之传统的密码系统有更丰富多彩的内容。

密码加密算法的对立面就是密码分析，也就是密码的破译技术研究。加密与破译是一对矛盾，了解破译对研究加密是非常必要的。

1.2 基本概念

什么是密码？简单地说它就是一组含有参数 k 的变换 E。设已知信息 m，通过变换 E_k 得密文 c，即

$$c=E_k(m)$$

这个过程称之为加密，参数 k 称之为密钥。加密算法 E 确定之后，由于密钥 k 不同，密文 c 也不同。

当然不是所有含参数 k 的变换都可以作为密码，它要求计算 $E_k(m)$ 不困难，而且若第三者不掌握密钥 k，即使截获了密文 c，他也无法从 c 恢复信息 m，也就是反过来从 c 求 m 极为困难。以后称 m 为明文。

通信双方一为发信方，或简称为发方，另一方为收信方或简称收方。传统的保密通信机理可用图 1.1 表示。

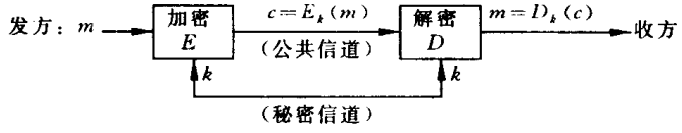


图 1.1

从密文 c 恢复明文 m 的过程称之为解密。解密算法 D 是加密算法 E 的逆运算，解密算法也是含参数 k 的变换。传统密码加密用的密钥 k 与解密用的密钥 k 是相同的，所以有时也叫对称密码。通信双方用的密钥 k 是通过秘密方式由双方私下约定产生的，只能由通信双方秘密掌握。如果丢失了密钥，则密码系统不攻自破。密钥的重要性可想而知。举一个最简单的例子。设已知明文 m 为

during the last twenty years there has been an explosion of public academic research in cryptography

明文的意思是“近 20 年对密码学的公开研究已急剧增加。”

先将明文分成 5 个字符一组得

durin gthel asttw entyy earst hereh asbee nanex plosi onofp ublic acade micre searc hincr yptog raphy

再将每组按相反顺序写下，例如 durin，倒过来写成 nirud，于是得密文 c 如下：

nirudlehtgwttsayytnetsraehereheebsaxenanisolppfonocilbuedacaercimcraes

rcnihgotpyyhpar

这里加密算法便是将明文先分组再逆序书写，密钥是每组的字符长。本例 $k=5$。若不知道加密算法，这密文相对于明文面目全非，从而达到加密的目的。当然这个加密算法不是很安全的，破译不难。

1.3 若干传统密码与它的破译

在这一节将介绍若干经典、传统的密码，附带讨论其中若干破译技巧。本书的重点是讨论加密算法。加密与破译是一对矛和盾，整个密码学也分成两大分支：加密方法与密码分析。密码分析则是研究破译的一门技术。但了解破译技术对研究加密算法是必要的。加密是一门科学，密码分析也是一门学问。有的加密算法对不掌握密码分析方法的人乍一看十分神秘，似乎“牢不可破”，其实不堪一击。前面已强调过：不可靠的密码比没有还坏。

1.3.1 密码举例

最早的一种密码是在公元前两世纪，由一位希腊人提出来的。他将 26 个字母排列在一个 5×5 的方格里，其中 i 和 j 填在同一格，见表 1.1：

表 1.1

	1	2	3	4	5
1	a	b	c	d	e
2	f	g	h	ij	k
3	l	m	n	o	p
4	q	r	s	t	u
5	v	w	x	y	z

于是每个字母对应一数 $\xi\eta$，其中 ξ 是该字母所在行的标号，η 是列标号。如 c 对应 13，r 对应 42 等等。使用这种密码可以将明文

secure message transmission is of extreme importance in information based society

转换为密文

43 15 13 45 42 15 32 15 43 43
11 22 15 44 42 11 33 43 32 24
43 43 24 34 33 24 43 34 21 15
53 44 42 15 32 15 24 32 35 34
42 44 11 33 13 15 24 33 24 33
21 34 42 11 44 24 34 33 12 11
43 15 13 43 34 13 24 15 44 54

在古代这种棋盘密码曾被广泛应用。

1.3.2 凯撒密码

凯撒密码是每一字母向前推移 k 位。例如 $k=5$ 便有明文和密文对应关系如下：

明文：a b c d e f g h i j k l m n o

密文：F G H I J K L M N O P Q R S T

明文：p q r s t u v w x y z

密文：U V W X Y Z A B C D E

于是对于明文：

data security has evolved rapidly

可加密得密文：

I F Y F X J H Z W N Y D M F X J A T Q A J I W F U N I Q D

不同的 k 可得不同的密文。

若令 26 个字母分别对应于整数 0～25，如表 1.2 所示：

表 1.2

a	b	c	d	e	f	g	h	i	j	k	l	m
1	2	3	4	5	6	7	8	9	10	11	12	13
n	o	p	q	r	s	t	u	v	w	x	y	z
14	15	16	17	18	19	20	21	22	23	24	25	0

则凯撒加密变换实际上是

$$c \equiv m + k \bmod 26$$

其中 m 是明文对应的数据，c 是与明文对应的密文数据，k 是加密用的参数，也称做为密钥。例如

data security

对应于数据序列

4 1 20 1 19 5 3 21 18 9 20 25

$k=5$ 时得密文序列

9 6 25 6 24 10 8 0 23 14 25 4

对应的密文为

I F Y F X J H Z W N Y D

若选取 k_1，k_2 两个参数，其中 $(k_1, 26)=1$，即 k_1 和 26 互素，令

$$c \equiv k_1 m + k_2 \bmod 26$$

这种变换称之为仿射变换。$k_1=1$ 时便是凯撒变换。

例如：$k_1=7$，$k_2=10$，则明文

please send moneys

的对应数据为

16 12 5 1 19 5 19 5 14 4

13 15 14 5 25 19

通过变换 $c \equiv 7m + 10 \bmod 26$ 可得

18 16 19 17 13 19 13 19 4 12

23 11 4 19 3 13

对应的密文为

R P S Q M S M S D L W K D S C M

1.3.3 单表置换

下面引进更一般的变换形式，例如

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

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

其中 TSINGHUAVERY 是 Tsinghua University 略去前面已出现的字符依次写下的。后面 BCD…WXZ 则是按英文字母表顺序续上前面未出现的字母而构成的，故 Tsinghua University便称之为该密码的密钥词组。密钥词组可作为密码的标志，记住这个密钥词组就能掌握字母加密置换的全体。

前面已提到加密算法和破译技术是一对矛和盾。为了更好地研究加密方法，了解必要的破译技巧是有好处的。

下面举一例子说明密码分析是如何进行的，若已知密文

G J X X N G G O T Z N U C O T W M O H Y

J T K T A M T X O B Y N F G O G I N U G

J F N Z V Q H Y N G N E A J F H Y O T W

G O T H Y N A F Z N F T U I N Z B N E G

N L N F U T X N X U F N E J C I N H Y A

Z G A E U T U C Q G O G O T H J O H O A

T C J X K H Y N U V O C O H O U H C N U

G H H A F N U Z H Y N C U T W J U W N A

E H Y N A F O W O T U C H N P H O G L N

F Q Z N G O F U V C N V J H T A H N G G

N T H O U C G J X Y O G H Y N A B N T O

T W G N T H N T X N A E B U F K N F Y O

H H G I U T J U C E A F H Y N G A C J H

O A T A E I O C O H U F Q X O B Y N F G

原来的明文经过加密变换显然掩盖了真面目，然而不是没有留下可以利用的蛛丝马迹。比如字母出现的频率和前后连缀的关系不会由于简单的置换而不留下痕迹。下面将从此入手，逐步剥去伪装恢复其原来面目。

从大量非技术性的英文书籍、报刊、文章中摘取适当长度的章节进行统计，发现英文字母出现的频率有惊人的相似之处。比如 e 出现的次数最多，其它如 t，a，o 等出现较多，几乎处处如此。不仅是单个字母如此，相邻的连缀字母也是如此。

出现频率较高的双字母有 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 等。三字母出现频率高的有 the，ing，and，her，ent 等。

经过大量的统计，可得出英文字母出现的频率如表 1.3 所示。

表 1.3

a	b	c	d	e	f
0.0856	0.0139	0.0279	0.0378	0.1304	0.0289
g	h	i	j	k	l
0.0199	0.0528	0.0627	0.0013	0.0042	0.0339
m	n	o	p	q	r
0.0249	0.0707	0.0797	0.0199	0.0012	0.0677
s	t	u	v	w	x
0.0607	0.1045	0.0249	0.0092	0.0149	0.0017
y	z				
0.0199	0.0008				

对上面密文中 280 个字母进行统计，并将出现的频数列表如下：

A	B	C	D	E	F	G	H	I	J	K	L	M
16	5	13	0	7	17	23	26	5	12	3	2	2
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
36	25	1	5	0	0	22	20	4	6	9	14	7

若按频数从大到小依次排列得

N	H	O	G	T	U	F	A	Y	C	J	X
36	26	25	23	22	20	17	16	14	13	12	9
E	Z	W	B	I	Q	V	K	L	M	P	
7	7	6	5	5	5	4	3	2	2	1	

机械地依照频率的大小顺序确定明文和密文的对应关系是不恰当的。一般来说，e，t，a，o，n，i，r，s，h 在高频率字母群中出现；d，l，u，c 和 m 在中频率字母群中出现；p，f，y，w，g，b，v 属于低频部分。j，k，q，x，z 比较罕见。从频率表中可以看出高频字母群和中频字母群之间有一明显的间断。也就是说，高频中频率最低为 h，它的频率为 0.0528，中频中频率最高为 d，它的频率为 0.0378，频率之差为

$$0.0528-0.0378=0.015$$

而 e 在高频字母群中特别突出。

这在密文的频数分析中也很明显。可以有理由假定高频字母群 e，t，…等 9 个字符正好是密文中前 9 位高频字符的对应明文。而且其中密文字符 N 非常可能就是 e 的对应密文。

现将密文中频率高的 9 个字母的前后连缀关系组成表 1.4。表中每个格子中有上下两个数，上面一个数表示前缀关系，下面一个数则表示后缀关系。例如 N 行 Y 列只有上面一个数 9，这表示 YN 出现 9 次，而 NY 一次也没有。又如 G 行 N 列上面一个数为 5，下面

一个数为4，这说明NG出现5次，而GN出现4次。其它依此类推。

表 1.4

↓ \ →	N	H	O	G	T	U	F	A	Y	
N		3 1		4 5	4	5	3 7	5	9	e
H	1 3	2 2	5 4	2 1	1 1	1 1	2	1 1	10	t
O		5 5		6 10	1 7	1	1 1	2	3	i
G	5 4	1 2	4 6	2 2		2	2	2		?
T	4	1 4	7 1			4 3	1	2 3		n
U	5	1 1	1	2	3 4	2 3				a
F	7 3	2	1 1	3	1	3 2		3	1	?
A	5	1 1	2	2	3 2		4		1	o
Y	9	10	3				1	1		h

现在依次判别如下：

① 密文N是明文e，用N↔e表示。

② 高频字母群中的a，i，o三个元音互联的机会极小。由此可推断O，U，A可能是a，i，o的密文。H显然很少同U和A相接触，然而经常和O相联，所以它不可能是元音。

③ a，i，o三个元音互相联结的机会显然很少，但根据统计i o的机会相对稍高。从上面的统计表中可知OA出现两次，OU出现一次。其它如UO，UA，AO，AU等均不出现。所以若OA分别是i o的密文，则应有

$$O\leftrightarrow i,\quad A\leftrightarrow o。$$

④ 从上面的推断结果来看，U可能是a的密文，OU出现一次，表明i a也是常见的。特别考虑到NU出现5次，而UN不出现，说明e a出现次数较多，而a e很少，这些事实都证明U↔a的可能性很大。

⑤ 高频辅音中最容易识别的是字母n，n前面是元音的机率高达0.80。从元音表中可见T前面和N，O，U，A连接的总次数为17。

⑥ Y的特点也很突出，YN出现9次，然而NY不出现。根据N↔e和h e出现机会多，而e h出现机会极少，所以推断Y↔h是非常可能的。

⑦ 根据英文中t h经常出现，而h t很罕见的特点，断定可能有H↔t。

⑧ 在高频率的字母群中，还有r，s两个未能判定。但r和a，e，i，o的连接多于s，而s多于同辅音相连。在余下的高频密文中，G和F中不容易得出有说服力的结论。虽有差

别，但并不明显。现在已经在280个字母中识别出159个，先将它代入部分密文得：

G	J	X	X	N	G	G	O	T	Z	N	U	C	O	T	W
				e			i	n		e	a		i	n	
M	O	H	Y	J	T	K	T	A	M	T	X	O	B	Y	N
w	i	t	h		n		n	o	w	n		i		h	e
F	G	O	G	I	N	U	G	J	F	N	Z	V	Q	H	Y
		i			e	a				e				t	h
N	G	N	E	A	J	F	H	Y	O	T	W	G	O	T	H
e		e		o			t	h	i	n			i	n	t
Y	N	A	F	Z	N	F	T	U	I	N	Z	B	N	F	G
h	e	o			e		n	a		e			e		
N	L	N	F	U	T	X	N	X	U	F	N	E	J	C	I
e		e		a	n		e		a		e				
N	H	Y	A	Z	G	A	E	U	T	U	C	Q	G	O	G
e	t	h	o			o		a	n	a				i	
O	T	H	J	O	H	O	A	T	C	J	X	K	H	Y	N
i	n	t		i	t	i	o	n					t	h	e
U	V	O	C	O	H	Q	U	H	C	N	U	G	H	H	A
a		i		i	t		a	t		e	a		t	t	o
F	N	U	Z	H	Y										
	e	a		t	h										

⑨ ___ith可能是with，故令M↔w。

⑩ 若将M代以w，代入上面密文，观察是否可以导出其它信息，请见上面有═══等号的部分

with—n—nown

和

int—ition

可猜测分别是with unknown和intuition，故令J↔u，K↔k，现将已得到的置换排列如下：

a	b	c	d	e	f	g	h	i	j	k	l	m
U				N			Y	O		K		
n	o	p	q	r	s	t	u	v	w	x	y	z
T	A					H	J		M			

⑪ 从对应关系

t	u	v	w
H	J		M

可以推测

l↔v

⑫ 由⑧知 F 和 G 究竟如何对应于 r 和 s，无确切的根据。现在由于

$$H\leftrightarrow t,\quad J\leftrightarrow u$$

所以可以令

$$F\leftrightarrow r,\quad G\leftrightarrow s$$

⑬ 由于 U↔a，可令 V↔b。

⑭ 将以上结果代入密文得

G	J	X	X	N	G	G	O	T	Z	N	U	C	O	T	W
s	u			e	s	s	i	n		e	a		i	n	
M	O	H	Y	J	T	K	T	A	M	T	X	O	B	Y	N
w	i	t	h	u	n	k	n	o	w	n		i		h	e
F	G	O	G	I	N	U	G	J	F	N	Z	V	Q	H	Y
r	s	i	s		e	a	s	u	r	e		b		t	h
N	G	N	E	A	J	F	H	Y	O	T	W				
e	s	e		o	u	r	t	h	i	n					

⑮ 由 s u (XX)e s s 可令 X↔c，得 success。括号( )里的 XX 是尚未破译的密文。

⑯ 由 ci (B) hers 可令 B↔p，得 ciphers。

⑰ 由(E)our 可令 E↔f 得 four。

⑱ thin(W)可令 W↔g 得 thing。

由此得置换的密钥词组为 NEW YORK CITY，得置换

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

全部明文为：

Success in dealing with unknown ciphers is measured by these four things in the order named, perseverance, careful methods of analysis, intuition, luck. The ability at least to read the language of the original text is very desirable but not essential. Such is the opening sentence of Parker Hitt's Manual for the solution of Military Ciphers.

标点符号是加上的。这段原文很有意思，翻译为："破译一未知密码是否成功，可由以下四个因素来衡量，按其顺序为：毅力、审慎的分析方法、直观和运气。阅读原文的文字的起码能力是需要的，然而不是必不可少的。这是 Parker Hitt 的'军事密码破译指南'一书的开场白。"

这个例子说明英文字符的频率差异是英文本身给密码带来多余的东西，为破译提供了可以利用的弱点。

前面介绍的密码体制都是属于单表置换。即一个明文字母对应的密文字母是确定的。正因为如此，频率分析对该密码体制进行了有效的攻击。下面将介绍一种多表密码，即一个明文字母可以表示成多个密文字母。

1.3.4 维吉利亚(Vigenere)密码

Vigenere 是法国的密码专家,以他的名字命名的加密算法是多表密码的典型代表。方法如下:

设密钥 $k=k_1k_2\cdots k_n$,明文 $M=m_1m_2\cdots m_n$,加密变换

$$E_k(M)=c_1c_2\cdots c_n$$

其中 $c_i\equiv(m_i+k_i) \bmod 26$, $i=1,2,\cdots,n$。

例如, M=data security, k=best。首先将 M 分解成长为 4 的序列

data secu rity

每一节利用密钥 k=best 加密得密文

$$c=E_k(M)=\text{EELT TIUN SMLR}$$

表 1.5 是维吉利亚方阵,利用它可以进行加密和脱密。

表 1.5 维吉利亚方阵

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

至于密钥 k 可以通过周期性地延长,周而复始,反复进行以至无穷。即令

$$\mathrm{k}=\mathrm{k}_1\mathrm{k}_2\mathrm{k}_3\cdots$$

其中

$$\mathrm{k}_{i+\ln}\equiv \mathrm{k}_i \ mod \ \mathrm{n}$$

例如,利用密钥 k=best 对明文 data 加密得密文 EELT,第一个 E 是在 b 行 d 列上;第二个 E 是在 e 行 a 列上;同样的道理 L 在 s 行 t 列上;第四个 T 是在 t 行 a 列上。其余

依此类推。脱密时第一个明文d是b行含密文E的列，同理第二个明文a是e行含密文E的列，同理推出其它。

维吉利亚方阵给我们以启发，可用类似想法得到很多其它的多表密码，其中最著名的有比欧福特(Beaufort)方法(见表1.6)。

表1.6 比欧福特方阵

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

它是由维吉利亚方阵作如下修改而得到的：比欧福特方阵的每一行恰好是维吉利亚方阵对应行的逆序排列。利用比欧福特方阵进行加密，相当于已知密钥

$$k=k_1k_2\cdots k_n$$

明文

$$M=m_1m_2\cdots m_n$$

可得密文

$$c=E_k(M)=c_1c_2\cdots c_n$$

其中

$$c_i\equiv(k_i+m_i)\bmod 26,\quad i=1,2,\cdots,n$$

维吉利亚方阵也可用来做比欧福特密码的加密、脱密。具体方法留给读者自己去思考。

多表密码加密算法结果将使得对单表置换用的简单频率分析方法失效。

1.3.5 对维吉利亚密码的分析

(1) 若随机地选取英文字母构成序列，任一字母被选上的概率为$\frac{1}{26}$。

若两个随机产生的英文字母序列并列，在特定的位置上出现相同字母的概率为

$$26\times\left(\frac{1}{26}\right)^2=1/26=0.0385$$

因特定的一字母重复的概率为$(1/26)^2$，但共有26个字母，所以特定位置上出现相同字母的概率为$\frac{1}{26}$。

如果任取两段英文并列，在特定位置上同时出现a的概率应为

$$(0.0856)^2=0.0073274$$

同时出现b的概率为

$$(0.0139)^2=0.0001932$$

其余可依次类推。

若p_1表示a出现的概率，p_2表示b出现的概率，…，p_{26}表示z出现的概率，那么

$$x=\sum_{i=1}^{26}p_i^2=0.0687$$

x是0.0385的1.784倍。也就是说，若任取两段英文并列，则在特定位置上出现相同字母的概率是随机产生的两段字母序列的1.784倍。

若对某一密文出现的字母频率进行统计得

$$f_A,f_B,f_C,\cdots,f_Z$$

而且满足

$$f_A+f_B+f_C+\cdots+f_Z=1$$

所以频率的平均值仍然是$\frac{1}{26}$。

我们引进

$$\begin{aligned}k^2&=\sum_{\xi=A}^{Z}\left(f_\xi-\frac{1}{26}\right)^2\\&=\sum_{\xi=A}^{Z}f_\xi^2-\frac{2}{26}\sum_{\xi=A}^{Z}f_\xi+\left(\frac{1}{26}\right)^2\cdot 26\end{aligned}$$

由于 $$\sum_{\xi=A}^{Z}f_\xi=1,$$

$\therefore$ $$\begin{aligned}k^2&=\sum_{\xi=A}^{Z}f_\xi^2-\frac{2}{26}+\frac{1}{26}\\&=\sum_{\xi=A}^{Z}f_\xi^2-\frac{1}{26}\end{aligned}$$

若当$f_A=f_B=\cdots=f_Z=\frac{1}{26}$时，$k=0$，$k$值愈大表示$f_A,f_B,\cdots,f_Z$起伏比较大，或$f_A,f_B,\cdots,f_Z$对于平均值$\frac{1}{26}$的偏离程度愈强烈，在计算$k^2$时需要计算$\sum_{\xi=A}^{Z}f_\xi^2$，这个值实际上