

高等院校信息科学系列教材

# 现代密码学

陈鲁生 沈世镒 编著

科学出版社

2002

## 内 容 简 介

本书系统地介绍现代密码学的基本内容,取材具有典型性.全书共分9章,第1章介绍密码学中的一些基本概念,第2章介绍古典密码的加密方法和一些典型的古典密码体制,第3章介绍 Shannon 的密码学理论,第4章和第5章分别讨论分组密码和公钥密码,第6章介绍序列密码和线性移位寄存器序列,第7章和第8章分别讨论数字签名和 Hash 函数,第9章介绍一些重要的密码协议.本书每章末均附有习题,其中有些习题是对正文内容的补充,以供学生复习巩固书中所学内容.

本书可作为高等院校信息科学专业或其他相关专业本科生的教材,也可作为相关领域中的教学、科研人员以及工程技术人员的参考书.

### 图书在版编目(CIP)数据

现代密码学/陈鲁生,沈世镒编著. - 北京:科学出版社,2002  
ISBN 7-03-010607-5

I. 现… II. ①陈… ②沈… III. 密码-理论 IV. TN918.1

中国版本图书馆 CIP 数据核字(2002)第 046822 号

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

印刷

科学出版社发行 各地新华书店经销

\*

2002 年 7 月第 一 版 开本:B5(720×1000)

2002 年 7 月第一次印刷 印张:10 1/2

印数:1-5 000

字数:197 000

定价:14.00 元

(如有印装质量问题,我社负责调换〈科印〉)

## 序 言

1998 年教育部进行高校专业调整时,设立了《信息与计算科学》专业.该专业的设立,受到很多高等院校的热烈响应,据不完全统计,几年来已有约 280 所院校招收了该专业的本科生,其中大部分院校计划开设信息科学方面的系列课程.

为了配合高等院校在学科专业设置上的改革与深化,来自几十所高等院校的有关专业的部分领导和教师,于 1999 年、2000 年召开了第一、二届《信息专业发展与学术研讨会》,与会者热烈讨论并探讨了许多与信息学科的学科发展和建设的基本问题.会议一致认为教材建设是目前最为紧迫的任务,因此成立了教材编审协调组来组织该系列教材的编写.

2001 年教材编写协调组召集了有多位经验丰富的教师和出版社参加的教材建设会议.会议明确了教材建设是一项长期的工作,并决定首先编写和出版这套教材来满足近期急需.为了保证教材的质量,会议对每本教材的要求、内容和大纲进行了具体研讨,并请具有多年教学经验的重点院校教授担任各教材的负责人.

为了贴近教学的实际,每部教材都配有习题或思考题,同时对内容也做了结构化安排,以便教师能根据实际情况部分选讲.本套教学用书不仅适用于教学,也可供相关读者参考.

在教材编写和出版过程中,作者对内容的取舍、章节的安排、结构的设计以及表达方式等方面多方听取意见,并进行了反复修改.在感谢作者们辛勤劳作的同时,编委会还特别感谢科学出版社的鞠丽娜编辑,她不辞辛劳,在统筹印刷出版、督促进度、征求意见、组织审校等方面做了大量工作.这套教材能在保证质量的前提下,及时与读者见面,和她的努力是分不开的.

从长远的教学角度考虑,为了适应不同类型院校、不同要求的课程需要,教材编审协调组将不断组织教材的修订、编写(译),从而使信息科学教学用书做到逐步充实、完善、提高和多样化.在此衷心希望采用该系列用书的教师、学生和读者对书中存在的问题及时提出修改意见和建议.

高等院校信息科学系列教材编委会

2002 年 3 月

# 前 言

随着计算机和通信网络的广泛应用,信息的安全性已受到人们的普遍重视.信息安全已不仅仅局限于政治、军事以及外交等领域,而且现在也与人们的日常生活息息相关.现在,密码学理论和技术已得到了迅速的发展,它是信息科学和技术中的一个重要研究领域.

多年来,我们一直在南开大学为信息科学专业的本科生和研究生讲授现代密码学课程,本书就是在此基础上编写而成的,旨在为高等院校信息科学专业或相关专业的本科生提供一本关于现代密码学的教材.

本书系统地介绍了现代密码学的基本内容.全书共 9 章.第 1 章介绍密码学中的一些基本概念.第 2 章介绍古典密码,讨论了古典密码的基本加密方法和分析方法,并介绍了一些典型的古典密码体制.第 3 章介绍 Shannon 的密码学理论.第 4 章讨论分组密码,主要介绍数据加密标准 DES 和高级加密标准 AES,这是两种不同类型的分组密码.第 5 章讨论公钥密码,介绍三种常见的公钥密码体制,并对公钥密码中用到的素数的生成方法进行了讨论.第 6 章介绍序列密码和线性移位寄存器序列.第 7 章和第 8 章分别讨论数字签名和 Hash 函数.第 9 章介绍一些重要的密码协议.

在本书的编写过程中,我们力求简明扼要,容易理解.对书中介绍的密码体制的数学基础,我们都做了简明扼要的介绍.书中所用到的数学结论基本上都做了证明,只有少数的数学结论由于证明过于复杂或者牵扯到更多的数学知识,我们只给出了结论,而没有给出证明,有兴趣的读者可以参阅相应的文献.众所周知,如果没有相应的数学基础,要理解一个密码体制是困难的.我们假定本书的读者具备了简单的概率论、高等代数、有限域以及数论等基本知识.另外,了解一点有关计算复杂性的知识对于理解各种密码体制和密码协议是有用的.由于严格地定义计算复杂性需要用到理论计算机科学中的一些知识,所以本书对此只做了一点简单的直观描述,感兴趣的读者可以参阅有关的文献.

本书适合高等院校信息科学、计算机科学以及通信等专业的高年级本科生使用,也可供相关领域的科研人员以及工程技术人员参考.

由于时间仓促,书中难免有疏漏和不当之处,敬请读者批评指正.

作 者

2002 年 5 月

# 目 录

第 1 章 引言	1
1.1 密码学的发展概况	1
1.2 密码学的基本概念	1
第 2 章 古典密码	4
2.1 古典密码中的基本加密运算	4
2.1.1 单表古典密码中的基本加密运算	4
2.1.2 多表古典密码中的基本加密运算	5
2.2 几种典型的古典密码体制	6
2.2.1 几种典型的单表古典密码体制	7
2.2.2 几种典型的多表古典密码体制	7
2.3 古典密码的统计分析	12
2.3.1 单表古典密码的统计分析	12
2.3.2 多表古典密码的统计分析	16
习题	21
第 3 章 Shannon 理论	23
3.1 密码体制的数学模型	23
3.2 熵及其性质	25
3.3 伪密钥和惟一解距离	31
3.4 密码体制的完善保密性	34
3.5 乘积密码体制	37
习题	38
第 4 章 分组密码	40
4.1 分组密码的基本原理	40
4.2 数据加密标准 DES	41
4.2.1 DES 加密算法	42
4.2.2 DES 的解密过程	48
4.2.3 DES 的安全性	48
4.3 多重 DES	49
4.3.1 双重 DES	49
4.3.2 三重 DES	50
4.4 DES 的工作模式	50

4.5	高级加密标准 AES .....	53
4.5.1	AES 的数学基础 .....	54
4.5.2	AES 的输入输出和中间状态 .....	57
4.5.3	AES 的加密过程 .....	59
4.5.4	密钥扩展 .....	62
4.5.5	AES 的解密过程 .....	63
	习题 .....	66
第 5 章	公钥密码 .....	69
5.1	公钥密码的理论基础 .....	69
5.2	RSA 公钥密码 .....	70
5.2.1	基本的数论知识 .....	70
5.2.2	RSA 公钥密码体制 .....	73
5.2.3	RSA 的安全性讨论 .....	74
5.2.4	模 $n$ 求逆的算法 .....	75
5.2.5	模 $n$ 的大数幂乘的快速算法 .....	77
5.2.6	因子分解 .....	77
5.3	大素数的生成 .....	78
5.3.1	素数的分布 .....	79
5.3.2	Legendre 符号和 Jacobi 符号 .....	80
5.3.3	Solovay-Strassen 素性测试法 .....	81
5.3.4	Miller-Rabin 素性测试法 .....	84
5.4	EIGamal 公钥密码 .....	86
5.4.1	EIGamal 公钥密码体制 .....	86
5.4.2	EIGamal 公钥密码体制的安全性 .....	87
5.4.3	有限域上离散对数的计算方法 .....	88
5.5	椭圆曲线上的 Menezes-Vanstone 公钥密码 .....	93
5.5.1	有限域上的椭圆曲线 .....	93
5.5.2	Menezes-Vanstone 公钥密码体制 .....	96
	习题 .....	98
第 6 章	序列密码与移位寄存器 .....	100
6.1	序列密码的基本原理 .....	100
6.2	移位寄存器与移位寄存器序列 .....	101
6.3	线性移位寄存器的表示 .....	103
6.4	线性移位寄存器序列的周期性 .....	105
6.5	线性移位寄存器的序列空间 .....	106
6.6	线性移位寄存器序列的极小多项式 .....	108

6.7 m 序列的伪随机性 ..... 112

6.8 B-M 算法与序列的线性复杂度 ..... 116

6.9 线性移位寄存器的非线性组合 ..... 119

习题..... 121

**第 7 章 数字签名**..... 122

7.1 基于公钥密码的数字签名 ..... 122

7.2 ElGamal 签名方案 ..... 123

7.3 数字签名标准 DSS ..... 125

7.4 基于离散对数问题的一般数字签名方案 ..... 126

习题..... 128

**第 8 章 Hash 函数** ..... 129

8.1 Hash 函数的性质..... 129

8.2 基于分组密码的 Hash 函数 ..... 130

8.3 Hash 函数 MD4 ..... 132

8.4 安全 Hash 算法 SHA ..... 135

习题..... 137

**第 9 章 密码协议**..... 139

9.1 密钥分配与密钥协商 ..... 139

9.1.1 密钥分配 ..... 140

9.1.2 密钥协商 ..... 143

9.2 秘密分享 ..... 146

9.3 身份识别 ..... 148

9.4 零知识证明 ..... 151

习题..... 153

**主要参考文献**..... 155

# 第 1 章 引 言

## 1.1 密码学的发展概况

密码学是一门既古老又年轻的学科,其历史可以追溯到几千年以前.

古代的行帮暗语和一些文字猜迷游戏等,实际上就是对信息的加密.这种加密方法通过原始的约定,把需要表达的信息限定在一定的范围内流通.

古典密码主要应用于政治、军事以及外交等领域,可以说,自从有了战争,就有了保密通信.交战双方都为了保护自己的通信安全、窃取对方的情报而研究各种信息加密技术和密码分析技术.

在 1949 年之前,密码技术基本上可以说是一门技巧性很强的艺术,而不是一门科学.在这一时期,密码专家常常是凭借直觉和信念来进行密码设计和分析,而不是推理证明.

在 1949 年,C. E. Shannon 发表了“保密系统的通信理论 (Communication Theory of Secrecy Systems)”一文,为密码学奠定了坚实的理论基础,使密码学成为一门真正的科学.但从 1949 年至 1975 年,密码学的理论研究工作进展不大.

1976 年,W. Diffie 和 M. E. Hellman 发表了“密码学中的新方向 (New Directions in Cryptography)”一文,提出了一种崭新的密码设计思想,导致了密码学的一场革命.他们首次证明了从发送端到接收端无密钥传输的保密通信是可能的,从而开创了公钥密码学的新纪元.1977 年,美国国家标准局 (National Bureau of Standards) 正式公布了数据加密标准 DES (Data Encryption Standard),将 DES 算法公开,从而揭开了密码学的神秘面纱.从此,密码学的研究进入了一个崭新的时代.

随着计算机科学的蓬勃发展,社会已进入信息时代.电子计算机和通信网络的广泛应用,一方面为人们的生活和工作提供了很大的方便,另一方面也提出了许多亟待解决的问题,其中信息的安全性就是一个突出的问题.因此,密码学理论和技术已成为信息科学与技术中的一个重要研究领域.随着计算机网络的迅速发展,特别是近年来电子商务的兴起,现代密码学的应用已不仅仅局限于政治、军事以及外交等领域,其商用价值和社会价值也已得到了充分的肯定.

## 1.2 密码学的基本概念

没有加密的信息称为明文 (plaintext). 加密后的信息称为密文 (ciphertext). 从明



文到密文的变换称为加密(encryption). 从密文到明文的变换称为解密(decryption).

加密和解密都是在密钥(key)的控制下进行的. 给定一个密钥,就可确定一对具体的加密变换和解密变换.

一个密码体制(cryptosystem)通常由五部分组成:

- (1) 明文空间  $M$ :全体明文的集合.
- (2) 密文空间  $C$ :全体密文的集合.
- (3) 密钥空间  $K$ :全体密钥的集合. 通常每个密钥  $k$  都由加密密钥  $k_e$  和解密密钥  $k_d$  组成,  $k = \langle k_e, k_d \rangle$ ,  $k_e$  与  $k_d$  可能相同也可能不同.
- (4) 加密算法  $E$ :由加密密钥控制的加密变换的集合.
- (5) 解密算法  $D$ :由解密密钥控制的解密变换的集合.

设  $m \in M$  是一个明文,  $k = \langle k_e, k_d \rangle \in K$  是一个密钥,则

$$c = E_{k_e}(m) \in C$$

$$m = D_{k_d}(c) \in M,$$

其中  $E_{k_e}$  是由加密密钥  $k_e$  确定的加密变换,  $D_{k_d}$  是由解密密钥  $k_d$  确定的解密变换.

在一个密码体制中,要求解密变换是加密变换的逆变换. 因此,对任意的  $m \in M$  都有

$$D_{k_d}(E_{k_e}(m)) = m$$

成立.

密钥空间中不同密钥的个数称为密码体制的密钥量. 它是衡量密码体制安全性的一个重要指标.

如果一个密码体制的加密密钥与解密密钥相同,则称其为单密钥密码体制或对称密码体制;否则,称其为双密钥密码体制或非对称密码体制.

在一个双密钥密码体制中,由加密密钥  $k_e$  计算解密密钥  $k_d$  是困难的,公开  $k_e$  不会损害  $k_d$  的安全性,则可以将加密密钥  $k_e$  公开. 这样的密码体制称为公钥密码体制(public-key cryptosystem).

一个好的密码体制至少应该满足下述两个条件:

- (1) 在已知明文  $m$  和加密密钥  $k_e$  时,计算  $c = E_{k_e}(m)$  容易. 在已知密文  $c$  和解密密钥  $k_d$  时,计算  $m = D_{k_d}(c)$  容易.
- (2) 在不知解密密钥  $k_d$  时,不可能由密文  $c$  推知明文  $m$ .

对于一个密码体制,如果能够根据密文确定明文或密钥,或者能够根据明文和相应的密文确定密钥,则我们说这个密码体制是可破译的;否则,称其为不可破译的.

密码分析者攻击密码体制的方法主要有以下三种:

- (1) 穷举攻击:密码分析者通过试遍所有的密钥来进行破译. 显然,可以通过增大密钥量来对抗穷举攻击.
- (2) 统计分析攻击:密码分析者通过分析密文和明文的统计规律来破译密码.

对抗统计分析攻击的方法是设法使明文的统计特性与密文的统计特性不一样.

(3) 解密变换攻击:密码分析者针对加密变换的数学基础,通过数学求解的方法来设法找到相应的解密变换.为对抗这种攻击,应该选用具有坚实的数学基础和足够复杂的加密算法.

密码分析者通常可以在下述四种情况下对密码体制进行攻击:

- (1) 唯密文攻击(ciphertext-only attack):密码分析者仅知道一些密文.
- (2) 已知明文攻击(known-plaintext attack):密码分析者知道一些明文和相应的密文.
- (3) 选择明文攻击(chosen-plaintext attack):密码分析者可以选择一些明文,并得到相应的密文.
- (4) 选择密文攻击(chosen-ciphertext attack):密码分析者可以选择一些密文,并得到相应的明文.

其中唯密文攻击的强度最弱,其他情况下的攻击强度依次增加.

对于一个密码体制,如果密码分析者无论截获了多少密文以及无论用什么方法进行攻击都不能破译,则称其为绝对不可破译的密码体制.绝对不可破译的密码在理论上是存在的.但是,如果能够利用足够的资源,那么任何实际的密码都是可以破译的.因此,更有实际意义的是在计算上不可破译(computationally unbreakable)的密码.所谓计算上不可破译是指密码分析者根据可利用的资源来进行破译所用的时间非常长,或者破译的时间长到使原来的明文失去保密的价值.

应当指出,对任何一种攻击方法,我们都假定密码分析者事先知道所使用的密码体制,这一点称为 Kerckhoff 假设,是由 Auguste Kerckhoff(1835- 1903)提出的.在设计密码体制时,应当记住的一点是永远不要低估密码分析者的能力.

图 1.1 给出的是一个密码系统模型.

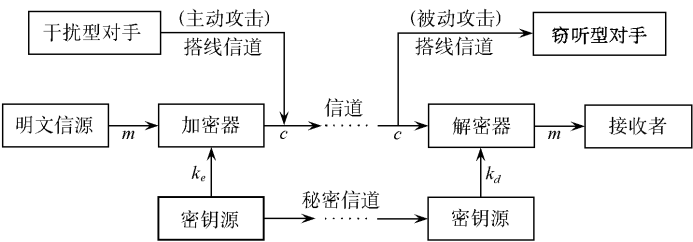


图 1.1 密码系统模型

密码系统的使用者通常称为用户.密码系统的破坏者有时称为对手.对手分为“窃听型”和“干扰型”两种.“窃听型”对手只是截取信道上传送的信息,而“干扰型”对手则会篡改信道上传送的信息.

## 第 2 章 古典密码

本章介绍古典密码体制中的基本加密运算、几种典型的古典密码体制以及关于古典密码体制的一些破译方法. 虽然古典密码都比较简单而且容易破译, 但研究古典密码的设计原理和分析方法对于理解、设计以及分析现代密码是十分有益的.

### 2.1 古典密码中的基本加密运算

明文字母表  $X$  是指明文空间  $M$  中出现的所有不同的字母的集合. 密文字母表  $Y$  是指密文空间  $C$  中出现的所有不同的字母的集合.

对于一个密码体制, 如果明文字母对应的密文字母在密文中保持不变, 则称其为单表密码体制; 如果明文中不同位置的同一明文字母在密文中对应的密文字母不同, 则称其为多表密码体制.

设  $q$  是一个正整数,  $Z_q = \{0, 1, 2, \dots, q-1\}$ , 则  $Z_q$  在模  $q$  加法和模  $q$  乘法运算下构成一个交换环. 记  $Z_q^* = \{k \in Z_q \mid \gcd(k, q) = 1\}$ , 则  $Z_q^*$  在模  $q$  乘法运算下构成一个乘法群, 其中  $\gcd(k, q)$  表示  $k$  和  $q$  的最大公因子.

#### 2.1.1 单表古典密码中的基本加密运算

##### 1. 加法密码

设  $X = Y = Z_q$ ,  $K = Z_q$ . 对任意  $m \in X$ ,  $k \in K$ , 密文

$$c = E_k(m) = (m + k) \bmod q.$$

显然, 加法密码的密钥量为  $q$ .

##### 2. 乘法密码

设  $X = Y = Z_q$ ,  $K = Z_q^*$ . 对任意  $m \in X$ ,  $k \in K$ , 密文

$$c = E_k(m) = km \bmod q.$$

解密变换为

$$m = k^{-1}c \bmod q.$$

关于  $k$  模  $q$  求逆的算法将在 5.2.4 节中介绍. 显然, 乘法密码的密钥量为  $\varphi(q)$ . 这里  $\varphi(q)$  是小于  $q$  且与  $q$  互素的非负整数的个数, 称为 Euler 函数.

##### 3. 仿射密码

设  $X = Y = Z_q$ ,  $K = \{(k_1, k_2) \mid k_1 \in Z_q, k_2 \in Z_q^*\}$ . 对任意  $m \in X$ ,  $k = (k_1, k_2) \in$

$\mathbb{K}$ , 密文

$$c = E_k(m) = (k_1 + k_2 m) \bmod q.$$

解密变换为

$$m = k_2^{-1}(c - k_1) \bmod q.$$

显然, 加法密码和乘法密码都是仿射密码的特例. 仿射密码的密钥量为  $q\varphi(q)$ .

#### 4. 置换密码

设  $X = Y = Z_q$ ,  $\mathbb{K}$  为  $Z_q$  上全体置换的集合. 对任意  $m \in X, k = \sigma \in \mathbb{K}$ , 密文

$$c = E_k(m) = \sigma(m).$$

显然, 仿射密码是置换密码的特例. 置换密码的密钥量为  $q!$ .

### 2.1.2 多表古典密码中的基本加密运算

#### 1. 简单加法密码

设  $X^n = Y^n = Z_q^n$ ,  $\mathbb{K} = Z_q^n$ . 对任意  $m = (m_1, m_2, \dots, m_n) \in X^n, k = (k_1, k_2, \dots, k_n) \in \mathbb{K}$ , 密文

$$c = E_k(m) = (m_1 + k_1, m_2 + k_2, \dots, m_n + k_n),$$

其中的加法都是模  $q$  加法. 显然, 简单加法密码的密钥量为  $q^n$ .

#### 2. 简单乘法密码

设  $X^n = Y^n = Z_q^n$ ,  $\mathbb{K} = \{(k_1, k_2, \dots, k_n) \mid k_i \in Z_q^*, 1 \leq i \leq n\}$ . 对任意  $m = (m_1, m_2, \dots, m_n) \in X^n, k = (k_1, k_2, \dots, k_n) \in \mathbb{K}$ , 密文

$$c = E_k(m) = (k_1 m_1, k_2 m_2, \dots, k_n m_n),$$

其中的乘法都是模  $q$  乘法. 显然, 简单乘法密码的密钥量为  $\varphi(q)^n$ .

#### 3. 简单仿射密码

设  $X^n = Y^n = Z_q^n$ ,

$$\mathbb{K} = \{(\langle k_{11}, k_{21} \rangle, \langle k_{12}, k_{22} \rangle, \dots, \langle k_{1n}, k_{2n} \rangle) \mid k_{1i} \in Z_q, k_{2i} \in Z_q^*, 1 \leq i \leq n\}.$$

对任意  $m = (m_1, m_2, \dots, m_n) \in X^n, k = (\langle k_{11}, k_{21} \rangle, \langle k_{12}, k_{22} \rangle, \dots, \langle k_{1n}, k_{2n} \rangle) \in \mathbb{K}$ , 密文

$$c = E_k(m) = (k_{11} + k_{21}m_1, k_{12} + k_{22}m_2, \dots, k_{1n} + k_{2n}m_n),$$

其中的加法和乘法都是模  $q$  加法和乘法. 显然, 简单仿射密码的密钥量为  $q^n \varphi(q)^n$ .

#### 4. 简单置换密码

设  $X^n = Y^n = Z_q^n$ ,  $\mathbb{K} = \{(k_1, k_2, \dots, k_n) \mid k_i \text{ 是 } Z_q \text{ 上的置换}, 1 \leq i \leq n\}$ . 对任意  $m = (m_1, m_2, \dots, m_n) \in X^n, k = (k_1, k_2, \dots, k_n) \in \mathbb{K}$ , 密文

$$c = E_k(m) = (k_1(m_1), k_2(m_2), \dots, k_n(m_n)).$$

显然,简单置换密码的密钥量为 $(q!)^n$ .

5. 换位密码

设 $X^n=Y^n=Z_q^n$ , 密钥空间  $\mathbf{K}$  为 $\{1,2,\cdots,n\}$ 上的全体置换的集合. 对任意明文 $m=(m_1,m_2,\cdots,m_n)\in X^n,k=\sigma\in \mathbf{K}$ , 密文

$$c=E_k(m)=(m_{\sigma(1)},m_{\sigma(2)},\cdots,m_{\sigma(n)}).$$

显然,换位密码的密钥量为 $n!$ .

6. 广义置换密码

设 $X^n=Y^n=Z_q^n$ ,  $\mathbf{K}$  为 $Z_q^n$ 上的全体置换的集合. 对任意 $m\in X^n,k=\sigma\in \mathbf{K}$ , 密文

$$c=E_k(m)=\sigma(m).$$

显然,广义置换密码的密钥量为 $(q^n)!$ .

7. 广义仿射密码

设 $X^n=Y^n=Z_q^n$ ,  $\mathbf{K}=\{(\alpha,H)\mid \alpha\in Z_q^n,H\text{ 为 }Z_q\text{ 上的 }n\text{ 阶可逆方阵}\}$ . 对任意 $m=(m_1,m_2,\cdots,m_n)\in X^n,k=(\alpha,H)\in \mathbf{K}$ , 密文

$$c=E_k(m)=\alpha+mH,$$

其中的运算都是模 $q$ 运算. 显然,广义仿射密码的密钥量为 $q^nr_n$ ,其中 $r_n$ 是 $Z_q$ 上不同的 $n$ 阶可逆方阵的个数.

2.2 几种典型的古典密码体制

表 2.1 给出了英文字母表 $\{a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z\}$ 和整数集 $Z_{26}$ 之间的一一对应关系.

表 2.1 字母与数字的对应关系

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

为方便起见,在本节中表示字母的符号同时也将表示字母所对应的整数,根据上下文能够容易地区分一个符号是表示字母还是表示字母所对应的整数.

2.2.1 几种典型的单表古典密码体制

1. Caesar 体制

Caesar 体制是一种典型的加法密码,其密钥  $k=3$ . 表 2.2 给出了 Caesar 体制中明文字母与密文字母的对应关系.

表 2.2 Caesar密表

明文字母	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
密文字母	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

2. 标准字头密码体制

这是一种置换密码. 它利用一个密钥字来构造置换作为密钥. 譬如,如果选择 cipher 作为密钥字,则标准字头密码体制中明文字母与密文字母的对应关系如表 2.3 所示.

表 2.3 标准字头密码体制的密表

明文字母	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
密文字母	c	i	p	h	e	r	a	b	d	f	g	j	k	l	m	n	o	q	s	t	u	v	w	x	y	z

2.2.2 几种典型的多表古典密码体制

1. Playfair 体制

Playfair 体制被英国军队作为一种战地密码体制使用多年,在第一次世界大战期间,美国和英国军队使用过一段时间. Playfair 体制是由英国的著名科学家——“Wheatstone 电桥”的设计者 Charles Wheatstone 发明的,但名字取自于率先发起应用这种密码体制的 Lionel Playfair.

Playfair 体制的密钥是一个  $5 \times 5$  的矩阵  $P = (p_{ij})_{5 \times 5}$ . 构造方法如下:

- (1) 构造字母表  $\{a,b,c,d,e,f,g,h,i,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z\}$  的一个置换. 这里将  $j$  当作  $i$ ,实际上只有 25 个字母.
- (2) 将上述置换按行排列成一个  $5 \times 5$  的矩阵  $P = (p_{ij})_{5 \times 5}$ .

用 Playfair 体制对明文字母串进行加密时,首先在明文字母串的适当位置插入一些特定的字母,譬如字母 q,使得明文字母串的长度为偶数,并且将明文字母串按两个字母一组进行分组,每组中的两个字母不同. 对任意的明文字母对  $m_1 m_2$ ,

设它们对应的密文对为  $c_1 c_2$ , 加密方法如下:

- (1) 如果  $m_1$  和  $m_2$  在  $P$  中同一行, 则密文  $c_1$  和  $c_2$  分别为紧靠  $m_1$  和  $m_2$  右端的字母. 这里将第一列看作是最后一列的右端.
- (2) 如果  $m_1$  和  $m_2$  在  $P$  中同一列, 则密文  $c_1$  和  $c_2$  分别为紧靠  $m_1$  和  $m_2$  下方的字母. 这里将第一行看作是最后一行的下方.
- (3) 如果  $m_1$  和  $m_2$  既不在  $P$  中的同一行, 也不在同一列, 则密文  $c_1$  和  $c_2$  分别为由  $m_1$  和  $m_2$  确定的矩形的其他两个角上的字母.  $c_1$  和  $m_1$  同行,  $c_2$  和  $m_2$  同行.

例 2.1 设密钥矩阵

$$P = \begin{bmatrix} c & i & p & h & e \\ r & a & b & d & f \\ g & k & l & m & n \\ o & q & s & t & u \\ v & w & x & y & z \end{bmatrix},$$

明文为

Playfair cipher was actually invented by wheatstone.

将明文分组为

pl ay fa ir ci ph er wa sa ct ua lq  
ly in ve nt ed by wh ea ts to ne

则密文为

bs dw rb ca ip he cf ik qb ho qf ks  
mx ek ze mu hf dx yi if ut uq uf

□

## 2. Vigenere 体制

Vigenere 体制是 1586 年由法国密码学家 Blaise de Vigenere 发明的. Vigenere 体制就是多表简单加法密码.

设明文  $m = m_1 m_2 \cdots m_n$ , 密钥  $k = k_1 k_2 \cdots k_n$ , 则密文

$$c = E_k(m) = c_1 c_2 \cdots c_n,$$

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

当密钥的长度比明文短时, 密钥可以周期性地重复使用, 直至完成明文中每个字母的加密.

表 2.4 称为 Vigenere 方阵, 利用它可以方便地进行加密和解密. 当用密钥字母  $k_i$  对明文字母  $m_i$  进行加密时, Vigenere 方阵中的第  $k_i$  行第  $m_i$  列的字母就是相应的密文字母.

表 2.4 Vigenere方阵

	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

例 2.2 设明文为

This cryptosystem is not secure,

密钥为 cipher,则密文为

VPXZGI AXIVWP UBTTMJ PWIZIT WZT. □

3. Beaufort 体制

Beaufort 体制与 Vigenere 体制非常相似,也是一种多表简单加法密码.



设明文  $m = m_1 m_2 \cdots m_n$ , 密钥  $k = k_1 k_2 \cdots k_n$ , 则密文

$$c = E_k(m) = c_1 c_2 \cdots c_n,$$

其中  $c_i = (k_i + 25 - m_i) \bmod 26, i = 1, 2, \cdots, n$ .

同 Vigenere 体制一样, 当密钥的长度比明文短时, 密钥可以周期性地重复使用, 直至完成明文中每个字母的加密.

表 2.5 称为 Beaufort 方阵, 它是由英国海军上将 Francis Beaufort 设计的, 利用它可以方便地进行加密和解密. 当用密钥字母  $k_i$  对明文字母  $m_i$  进行加密时, Beaufort 方阵中的第  $k_i$  行第  $m_i$  列的字母就是相应的密文字母.

表 2.5 Beaufort 方阵

	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

### 例 2.3 设明文为

This cryptosystem is also not secure,  
密钥为 cipher, 则密文为

IAJOBZ DSVSLS JOKUVY BWWSQC IPKEJZ X. □

### 4. Vernam 体制

美国电话电报公司的 Gilbert Vernam 在 1917 年为电报通信设计了一种非常方便的密码, 后来称之为 Vernam 密码. Vernam 密码在对明文加密前首先将明文编码为  $(0, 1)$  字符串.

设  $m = m_1 m_2 \cdots m_i \cdots$  为明文,  $k = k_1 k_2 \cdots k_i \cdots$  为密钥, 其中  $m_i, k_i \in \text{GF}(2), i \geq 1$ . 则密文  $c = c_1 c_2 \cdots c_i \cdots$ , 其中

$$c_i = m_i \oplus k_i, i \geq 1.$$

这里  $\oplus$  为模 2 加法.

在用 Vernam 密码对明文加密时, 如果对不同的明文使用不同的密钥, 则这时 Vernam 密码为“一次一密”密码, 在理论上是不可破译的. 如果存在不同的明文使用相同的密钥, 则这时 Vernam 密码就比较容易破译.

当对手获得了一个密文  $c = c_1 c_2 \cdots c_i \cdots$  所对应的明文  $m = m_1 m_2 \cdots m_i \cdots$  时, 很容易得到密钥  $k = k_1 k_2 \cdots k_i \cdots$ , 其中

$$k_i = m_i \oplus c_i, \quad i \geq 1.$$

因此, 如果密钥  $k$  重复使用, 对手可以立即解密得到相应的明文.

### 5. Hill 体制

Hill 体制是 1929 年由 Lester S. Hill 发明的. 它实际上就是广义仿射密码的一个特例. Hill 体制的基本思想是将  $n$  个明文字母通过线性变换转换为  $n$  个密文字母. 解密时只需做一次逆变换即可. 密钥就是变换矩阵.

设明文  $m = (m_1, m_2, \cdots, m_n) \in Z_{26}^n$ , 密文  $c = (c_1, c_2, \cdots, c_n) \in Z_{26}^n$ , 密钥为  $Z_{26}$  上的  $n \times n$  阶可逆方阵  $K = (k_{ij})_{n \times n}$ , 则

$$\begin{aligned} c &= mK \pmod{26}, \\ m &= cK^{-1} \pmod{26}. \end{aligned}$$

例 2.4 设  $n = 2$ , 密钥为

$$K = \begin{bmatrix} 11 & 8 \\ 3 & 7 \end{bmatrix},$$

容易计算

$$K^{-1} = \begin{bmatrix} 7 & 18 \\ 23 & 11 \end{bmatrix}.$$

设明文为 Hill, 则相应的明文向量为  $(7, 8)$  和  $(11, 11)$ . 于是, 相应的密文向量分别

为

$$(7,8)\begin{bmatrix} 11 & 8 \\ 3 & 7 \end{bmatrix} = (77 + 24, 56 + 56) = (23,8),$$
$$(11,11)\begin{bmatrix} 11 & 8 \\ 3 & 7 \end{bmatrix} = (121 + 33, 88 + 77) = (24,9).$$

因此,明文 Hill 的密文为 XIYJ. □

## 2.3 古典密码的统计分析

### 2.3.1 单表古典密码的统计分析

单表古典密码体制的密文字母表实际上是明文字母表的一个排列. 因此,明文字的统计特性在密文中能够反映出来. 当截获的密文足够多时,可以通过统计密文字母的出现频率,来确定明文字母和密文字母之间的对应关系.

表 2.6 给出了 26 个英文字母出现的频率. 当然,由于用于统计的明文的长度的不同,以及明文反映的内容的不同,不同的文献中给出的 26 个英文字母出现的频率可能略有差别.

表 2.6 26 个英文字母的出现频率

字 母	频 率	字 母	频 率
a	0.082	n	0.067
b	0.015	o	0.075
c	0.028	p	0.019
d	0.043	q	0.001
e	0.127	r	0.060
f	0.022	s	0.063
g	0.020	t	0.091
h	0.061	u	0.028
i	0.070	v	0.010
j	0.002	w	0.023
k	0.008	x	0.001
l	0.040	y	0.020
m	0.024	z	0.001

26 个英文字母按出现频率的大小可以分为以下五类:

- (1) e:出现频率约为 0.120.
- (2) t,a,o,i,n,s,h,r:出现频率约在 0.06 到 0.09 之间.
- (3) d,l:出现频率约为 0.04.
- (4) c,u,m,w,f,g,y,p,b:出现频率约在 0.015 到 0.028 之间.
- (5) v,k,j,x,q,z:出现频率小于 0.01.

在密码分析中,除了单字母的统计特性外,知道双字母和三字母的统计特性也是非常有用的.

出现频率最高的 30 个双字母(按频率从大到小排列)为

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(按频率从大到小排列)为

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

下面的例子说明了对单表古典密码的统计分析方法. 该例子取自 D. R. Stinsou 的“Cryptography:Theroy and Practice”一书.

例 2.5 设某一段明文经单表密码加密后的密文为

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

设加密变换为  $E_k$ ,解密变换为  $D_k$ . 密文中共有 168 个字母,各个密文字母的出现次数和出现频率如表 2.7 所示.

表 2.7 例 2.5 中各个密文字母的出现次数和出现频率

密文字母	出现次数	出现频率	密文字母	出现次数	出现频率
A	0	0.000	N	9	0.054
B	1	0.006	O	0	0.000
C	15	0.089	P	1	0.006
D	13	0.077	Q	4	0.024
E	7	0.042	R	10	0.060
F	11	0.065	S	3	0.018
G	1	0.006	T	2	0.012
H	4	0.024	U	5	0.030
I	5	0.030	V	5	0.030
J	11	0.065	W	8	0.048
K	1	0.006	X	6	0.036
L	0	0.000	Y	10	0.060
M	16	0.095	Z	20	0.119

从表 2.7 可以看出,密文字母 Z 的出现次数明显地比其他密文字母的出现次数多,出现频率约为 0.12. 因此,可以猜测

$$D_k(Z) = e.$$

除 Z 外,出现至少 10 次的密文字母为 C,D,F,J,M,R,Y,它们的出现频率约在 0.06 到 0.095 之间.因此,可以猜测

$$\{D_k(C), D_k(D), D_k(F), D_k(J), D_k(M), D_k(R), D_k(Y)\} \subseteq \{t, a, o, i, n, s, h, r\}.$$

现在考虑密文中形如- Z 和 Z- 的双字母的出现情况.在密文中没有出现的形如- Z 和 Z- 的双字母在表 2.8 中没有列出.

表 2.8 例 2.5 的密文中包含字母 Z 的双字母的出现次数

- Z	出现次数	Z-	出现次数
DZ	4	ZW	4
NZ	3	ZU	3
RZ	2	ZR	2
HZ	2	ZV	2
XZ	2	ZC	2
FZ	2	ZD	2
		ZJ	2

因为 ZW 出现 4 次,WZ 没有出现,并且 W 的出现频率为 0.048,所以可以猜测

$$D_k(W) = d.$$

注意到 DZ 出现 4 次,ZD 出现 2 次,可以猜测  $D_k(D) \in \{r, s, t\}$ .

在  $D_k(Z) = e$  和  $D_k(W) = d$  的假设前提下,继续观察密文.可以注意到在密文的开始部分出现了 ZRW 和 RZW,并且在后面还出现了 RW.因为 R 在密文中经常出现,并且 nd 在明文是常见的双字母,所以很有可能

$$D_k(R) = n.$$

下面我们列出明文与密文的部分对应关系:

```

- - - - - e n d - - - - - e - - - - n e d - - - e - - - - -
Y I F Q F M Z R W Q F Y V E C F M D Z P C V M R Z W N M D Z V E J B T X C D D U M J
- - - - - e - - - - e - - - - - n - - d - - - e n - - - e - - - e
N D I F E F M D Z C D M Q Z K C E Y F C J M Y R N C W J C S Z R E X C H Z U N M X Z
- e - - - n - - - - - n - - - - - e d - - - e - - - e - n e - n d - e - e - -
N Z U C D R J X Y Y S M R T M E Y I F Z W D Y V Z V Y F Z U M R Z C R W N Z D Z J J
- e d - - - - n - - - - - e - - - - e d - - - - - d - - - e - n
X Z W G C H S M R N M D H N C M F Q C H Z J M X J Z W I E J Y U C F W D J N Z D I R

```

因为 NZ 是密文中常见的双字母,而 ZN 不出现,所以我们可以试着猜测

$$D_k(N) = h.$$

如果这个猜测正确,则由明文片段 ne-ndhe 可以猜测

$$D_k(C) = a.$$

于是,我们有

- - - - - e n d - - - - - a - - - e - a - - n e d h - - e - - - - - a - - - - -  
Y I F Q F M Z R W Q F Y V E C F M D Z P C V M R Z W N M D Z V E J B T X C D D U M J  
h - - - - - e a - - - e - a - - - a - - - n h a d - a - e n - - a - e - h - - e  
N D I F E F M D Z C D M Q Z K C E Y F C J M Y R N C W J C S Z R E X C H Z U N M X Z  
h e - a - n - - - - - n - - - - - e d - - - e - - - e - - n e a n d h e - e - -  
N Z U C D R J X Y Y S M R T M E Y I F Z W D Y V Z V Y F Z U M R Z C R W N Z D Z J J  
- e d - a - - - n h - - - h a - - - a - e - - - - e d - - - - - a - d - - h e - - n  
X Z W G C H S M R N M D H N C M F Q C H Z J M X J Z W I E J Y U C F W D J N Z D I R

在密文中,  $M$  是第二个最常出现的字母. 对于密文片段  $RNM$ , 我们已经猜测它对应明文  $nh-$ . 由此可知,  $h$  可能是某一个明文单词的首字母. 因此,  $M$  可能代表明文中的一个元音字母. 因为  $D_k(C) = a, D_k(Z) = e$ , 所以我们可以猜测

$$D_k(M) \in \{i, o\}.$$

又因为明文双字母  $ai$  比  $ao$  更常出现, 所以由密文双字母  $CM$ , 我们可以先猜测

$$D_k(M) = i.$$

于是,我们有

- - - - - i e n d - - - - - a - i - e - a - i n e d h i - e - - - - - a - - - i -  
Y I F Q F M Z R W Q F Y V E C F M D Z P C V M R Z W N M D Z V E J B T X C D D U M J  
h - - - - - i - e a - i - e - a - - - a - i - n h a d - a - e n - - a - e - h i - e  
N D I F E F M D Z C D M Q Z K C E Y F C J M Y R N C W J C S Z R E X C H Z U N M X Z  
h e - a - n - - - - - i n - i - - - - e d - - - e - - - e - i n e a n d h e - e - -  
N Z U C D R J X Y Y S M R T M E Y I F Z W D Y V Z V Y F Z U M R Z C R W N Z D Z J J  
- e d - a - - i n h i - - h a i - - a - e - i - - e d - - - - - a - d - - h e - - n  
X Z W G C H S M R N M D H N C M F Q C H Z J M X J Z W I E J Y U C F W D J N Z D I R

现在我们来确定哪个密文字母对应明文字母  $o$ . 因为  $o$  是一个常见的明文字母, 所以我们可以猜测

$$E_k(o) \in \{D, F, J, Y\}.$$

看起来, 最可能的是  $E_k(o) = Y$ . 否则, 由密文片段  $CFM$  或  $CJM$  将得到明文中的一长串元音  $aoi$ , 这是不太可能的. 因此, 我们猜测

$$D_k(Y) = o.$$

对于另外三个最常出现的密文字母  $D, F, J$ , 我们猜测

$$\{D_k(D), D_k(F), D_k(J)\} \subseteq \{r, s, t\}.$$

由于密文片段  $NMD$  出现了两次, 所以我们可以猜测

$$D_k(D) = s.$$

于是, 密文片段  $NMD$  对应明文  $his$ . 这与我们前面的猜测  $D_k(D) \in \{r, s, t\}$  是一致的. 对于密文片段  $HNCMF$ , 我们猜测它可能对应明文  $chair$ , 由此可知

$$D_k(F) = r, D_k(H) = c.$$

于是

$$D_k(J) = t.$$

现在,我们有

o-r-r-iend-ro--a-r-i-se-a-inedhise--t---a-s-s-i-t  
Y I F Q F M Z R W Q F Y V E C F M D Z P C V M R Z W N M D Z V E J B T X C D D U M J  
h-s-r-r-i-se-a-s-i-e-a-o-r-a-t-i-o-n-h-a-d-t-a-e-n--a-c-e-h-i-e  
N D I F E F M D Z C D M Q Z K C E Y F C J M Y R N C W J C S Z R E X C H Z U N M X Z  
h-e-a-s-n-t-o-o-i-n-i-o-r-e-d-s-o-e-o-r-e-i-n-e-a-n-d-h-e-s-e-t-t  
N Z U C D R J X Y Y S M R T M E Y I F Z W D Y V Z V Y F Z U M R Z C R W N Z D Z J J  
-e-d-a-c-i-n-h-i-s-c-h-a-i-r-a-c-e-t-i-t-e-d--t-o-a-r-d-s-t-h-e-s-n  
X Z W G C H S M R N M D H N C M F Q C H Z J M X J Z W I E J Y U C F W D J N Z D I R

利用与上述分析类似的方法,容易确定其余密文字母和明文字母的对应关系.最后,将得到的明文加上标点符号,我们得到明文:

Our friend from Paris examined his empty glass with surprise, as if evaporation had taken place while he wasn't looking. I poured some more wine and he settled back in his chair, face tilted up towards the sun. □

上面讨论的单表古典密码的统计分析方法是针对一般的单表置换密码的. 如果已知所用的密码体制,则相应的分析工作可能会更简单一些. 对于加法密码和乘法密码,只需得到一个密文字母所对应的明文字母,就可求得密钥. 对于一般的仿射密码,只需得到两个密文字母所对应的明文字母,就可求得密钥.

另外,对于加法密码和乘法密码,由于密钥量比较小,也可以通过穷举攻击求得密钥.

### 2.3.2 多表古典密码的统计分析

在多表古典密码的分析中,首先要确定密钥字的长度,也就是要确定所使用的加密表的个数,然后再分析确定具体的密钥.

下面我们以 Vigenere 密码为例来说明多表古典密码的分析方法.

确定密钥字的长度  $m$  的常用方法有 Kasiski 测试法(Kasiski test)和重合指数法(index of coincidence).

在明文中,如果两个相同的明文片段之间的距离  $d$  是密钥字长度  $m$  的倍数,则这两个明文片段所对应的密文片段一定是相同的. 反过来,如果密文中出现两个相同的密文片段(长度至少为 3),则它们对应的明文片段极有可能是相同的,当然也可能不同.

寻找密文中长度至少为 3 的相同的密文片段对,计算每对密文片段中的两个密文片段之间的距离. 这样,我们就得到了一些距离值  $d_1, d_2, d_3, \dots, d_i$ . 我们可以猜测密钥字的长度  $m$  可能是  $d_1, d_2, d_3, \dots, d_i$  的最大公因子. 这就是 Kasiski 测试法. 它是 Friedrich Kasiski 于 1863 年首先提出来的.

利用重合指数法可以进一步确定密钥字的长度是否为  $m$ . 重合指数的概念是 Wolfe Friedman 在 1920 年给出的.

**定义 2.1** 设  $x = x_1 x_2 \cdots x_n$  是一个长度为  $n$  的英文字母串.  $x$  的重合指数定义为  $x$  中的两个随机元素相同的概率, 记为  $I_c(x)$ . 假设英文字母 A, B, C,  $\cdots$ , Z 在  $x$  中的出现次数分别为  $f_0, f_1, f_2, \cdots, f_{25}$ . 显然, 从  $x$  中选取两个元素共有  $\begin{bmatrix} n \\ 2 \end{bmatrix}$  种方法, 选取的两个元素同时为第  $i$  个英文字母的情形共有  $\begin{bmatrix} f_i \\ 2 \end{bmatrix}$  种,  $0 \leq i \leq 25$ . 因此, 我们有

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

将表 2.6 中英文字母 A, B, C,  $\cdots$ , Z 的期望概率分别记为  $p_0, p_1, p_2, \cdots, p_{25}$ . 假设  $x$  为英文明文. 因为  $x$  中的两个随机元素同时为第  $i$  个英文字母的概率为  $p_i^2$ ,  $0 \leq i \leq 25$ , 所以我们有

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

如果  $x$  是由单表密码得到的密文, 则显然各个密文字母的期望概率是各个明文字母的期望概率的一个重新排列. 因此, 我们仍然有

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

假设  $y = y_1 y_2 \cdots y_n$ , 是由 Vigenere 密码得到的长度为  $n$  的密文. 将  $y$  按列排成一个  $m \times (n/m)$  的矩形阵列. 各行分别记为  $y_1, y_2, \cdots, y_m$ . 如果  $m$  确实是密钥字的长度, 则  $y_i$  中的各个密文字母都是由同一个加法密码得到的. 因此,

$$I_c(y_i) \approx 0.065, 1 \leq i \leq m.$$

另一方面, 如果  $m$  不是密钥字的长度, 则  $y_i$  中的各个密文字母将是由不同的加法密码得到的,  $y_i$  中的各个密文字母看起来更随机一些. 对于一个随机的英文字母串, 我们注意到

$$I_c \approx 26 \times (1/26)^2 = 1/26 = 0.038.$$

因为 0.065 和 0.038 相差较远, 所以我们一般能够确定密钥字的长度, 或者说我们能够确定由 Kasiski 测试法得到的密钥字的长度是否正确.

下面我们来讨论如何确定密钥字. 首先给出交互重合指数 (mutual index of coincidence) 的概念, 它可以用于确定密钥字.

**定义 2.2** 设  $x = x_1 x_2 \cdots x_n$  和  $y = y_1 y_2 \cdots y_{n'}$  是两个长度分别为  $n$  和  $n'$  的字母串.  $x$  和  $y$  的交互重合指数定义为  $x$  中的一个随机元素与  $y$  中的一个随机元素相同的概率, 记为  $M I_c(x, y)$ . 假设英文字母 A, B, C,  $\cdots$ , Z 在  $x$  和  $y$  中的出现次数分



别为  $f_0, f_1, f_2, \dots, f_{25}$  和  $f'_0, f'_1, f'_2, \dots, f'_{25}$ , 则容易看出

$$M I_c(x, y) = \frac{\sum_{i=0}^{25} f_i f'_i}{nn}.$$

假设我们已经确定了密钥字的长度  $m$ , 密文子串  $y_i$  中的各个密文字母都是由同一个加法密码得到的. 设密钥为  $k = (k_1, k_2, \dots, k_m)$ . 现在我们来估计  $M I_c(y_i, y_j)$  的值. 显然,  $y_i$  中的一个随机元素与  $y_j$  中的一个随机元素同时为第  $h$  个英文字母的概率为  $p_{h-k_i} p_{h-k_j}, 0 \leq h \leq 25$ , 这里下标中的运算为模 26 运算. 因此, 我们有

$$M I_c(y_i, y_j) \approx \sum_{h=0}^{25} p_{h-k_i} p_{h-k_j} = \sum_{h=0}^{25} p_h p_{h-k_i-k_j}.$$

注意到这个估计值仅依赖于  $(k_i - k_j) \bmod 26$ , 我们称  $(k_i - k_j) \bmod 26$  为  $y_i$  与  $y_j$  的相对位移(relative shift). 另外, 因为

$$\sum_{h=0}^{25} p_h p_{h+l} = \sum_{h=0}^{25} p_h p_{h-l},$$

所以相对位移  $l$  和  $26-l$  产生相同的  $M I_c$  的估计值.

表 2.9 列出了交互重合指数的估计值.

表 2.9 交互重合指数  $M I_c$  的估计值

相对位移	$M I_c$ 的估计值	相对位移	$M I_c$ 的估计值
0	0.065	7 (19)	0.039
1 (25)	0.039	8 (18)	0.034
2 (24)	0.032	9 (17)	0.034
3 (23)	0.034	10 (16)	0.038
4 (22)	0.044	11 (15)	0.045
5 (21)	0.033	12 (14)	0.039
6 (20)	0.036	13	0.043

从表 2.9 可以看出, 当相对位移不等于 0 时, 交互重合指数的估计值在 0.031 到 0.045 之间, 而当相对位移等于 0 时, 交互重合指数的估计值为 0.065. 注意到 0.045 与 0.065 相差较远, 我们可以利用这个事实来推测  $y_i$  与  $y_j$  的相对位移  $l = k_i - k_j$ . 设  $e_0 = 0, e_1 = 1, e_2 = 2, \dots, e_{25} = 25$ . 用  $e_g$  作为密钥按加法密码对  $y_j$  加密, 设密文为  $y_j^g, 0 \leq g \leq 25$ . 利用公式

$$M I_c(x, y_j^g) = \frac{\sum_{i=0}^{25} f_i f'_{i-g}}{nn}$$

容易计算交互重合指数  $M I_c(y_i, y_j^g), 0 \leq g \leq 25$ . 因为  $y_i$  与  $y_j^l$  的相对位移为 0, 所以

当  $g = l$  时,  $M I_c(y_i, y_j^g)$  应该接近 0.065. 而当  $g \neq l$  时,  $M I_c(y_i, y_j^g)$  应该在 0.031 到 0.045 之间变化. 通过这种方法, 我们可以获得任意两个密文字串  $y_i$  和  $y_j$  的相对位移. 对于加法密码, 因为只有 26 个可能的密钥, 所以可以很容易地通过穷举搜索得到密钥.

下面的例子说明了 Vigenere 密码的分析方法. 该例子取自 D. R. Stinson 的 “Cryptography: Theory and Practice” 一书.

例 2.6 设某一段明文经 Vigenere 密码加密后的密文为

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQRBW  
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK  
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX  
VRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR  
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJTB  
AMRVLCCRREMN DGLXRRIMGNSNRWCHRQHA EYEVT AQEBBI  
PEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHP  
WQAI IWXNRMGWOI I FKEE

可以看出, 密文片段 CHR 在密文中出现了 5 次, 每次出现的开始位置分别为 1, 166, 236, 276, 286. 第一次出现到其他各次出现的距离分别为 165, 235, 275, 285. 容易计算

$$\gcd(165, 235, 275, 285) = 5.$$

因此, 密钥字的长度很可能是 5.

下面我们来计算重合指数, 看看是否也能得出同样的结论. 根据定义 2.1 中的公式, 密文子串  $y_i$  的重合指数是容易计算的,  $1 \leq i \leq m$ ,  $m$  为密钥字的长度. 如果  $m = 1$ , 则重合指数为 0.045. 如果  $m = 2$ , 则两个重合指数分别为 0.046, 0.041. 如果  $m = 3$ , 则三个重合指数分别为 0.043, 0.050, 0.047. 如果  $m = 4$ , 则四个重合指数分别为 0.042, 0.039, 0.046, 0.040. 如果  $m = 5$ , 则五个重合指数分别为 0.063, 0.068, 0.069, 0.061, 0.072. 显然,  $m = 5$  时的各个重合指数最接近 0.065. 因此, 通过计算密文子串  $y_i$  的重合指数, 进一步证实了密钥字的长度是 5.

现在我们来讨论确定密钥字. 通过编制一个简单的计算机程序, 容易得到 260 个交互重合指数  $M I_c(y_i, y_j^g)$  的值,  $1 \leq i < j \leq 5, 0 \leq g \leq 25$ . 表 2.10 中列出了这些值.

对于每一对  $(i, j)$ , 找出接近 0.065 的  $M I_c(y_i, y_j^g)$  的值. 对于给定的  $(i, j)$ , 如果存在惟一的  $M I_c(y_i, y_j^g)$  的值接近 0.065, 则我们可以将相应的  $g$  视为  $y_i$  与  $y_j$  的相对位移的值. 在表 2.10 中, 用方框划出了 6 个接近 0.065 的交互重合指数的值. 因此, 我们可以猜测  $y_1$  与  $y_2$  的相对位移是 9,  $y_1$  与  $y_5$  的相对位移是 16,  $y_2$  与  $y_3$  的相对位移是 13,  $y_2$  与  $y_5$  的相对位移是 7,  $y_3$  与  $y_5$  的相对位移是 20,  $y_4$  与  $y_5$  的相对位移是 11. 假设  $k = (k_1, k_2, k_3, k_4, k_5)$  是 Vigenere 密码的密钥, 则我们有

表 2.10 密文字串的交互重合指数  $M I_c$  的估计值

$i$	$j$	$M I_c(y_i, y_j)$ 的值, $0 \leq g \leq 25$								
1	2	.028	.027	.028	.034	.039	.037	.026	.025	.052
		<u>.068</u>	.044	.026	.037	.043	.037	.043	.037	.028
		.041	.041	.034	.037	.051	.045	.042	.036	
1	3	.039	.033	.040	.034	.028	.053	.048	.033	.029
		.056	.050	.045	.039	.040	.036	.037	.032	.027
		.037	.036	.031	.037	.055	.029	.024	.037	
1	4	.034	.043	.025	.027	.038	.049	.040	.032	.029
		.034	.039	.044	.044	.034	.039	.045	.044	.037
		.055	.047	.032	.027	.039	.037	.039	.035	
1	5	.043	.033	.028	.046	.043	.044	.039	.031	.026
		.030	.036	.040	.041	.024	.019	.048	<u>.070</u>	.044
		.028	.038	.044	.043	.047	.033	.026	.046	
2	3	.046	.048	.041	.032	.036	.035	.036	.030	.024
		.039	.034	.029	.040	<u>.067</u>	.041	.033	.037	.045
		.033	.033	.027	.033	.045	.052	.042	.030	
2	4	.046	.034	.043	.044	.034	.031	.040	.045	.040
		.048	.044	.033	.024	.028	.042	.039	.026	.034
		.050	.035	.032	.040	.056	.043	.028	.028	
2	5	.033	.033	.036	.046	.026	.018	.043	<u>.080</u>	.050
		.029	.031	.045	.039	.037	.027	.026	.031	.039
		.040	.037	.041	.046	.045	.043	.035	.030	
3	4	.038	.036	.040	.033	.036	.060	.035	.041	.029
		.058	.035	.035	.034	.053	.030	.032	.035	.036
		.036	.028	.046	.032	.051	.032	.034	.030	
3	5	.035	.034	.034	.036	.030	.043	.043	.050	.025
		.041	.051	.050	.035	.032	.033	.033	.052	.031
		.027	.030	<u>.072</u>	.035	.034	.032	.043	.027	
4	5	.052	.038	.033	.038	.041	.043	.037	.048	.028
		.028	.036	<u>.061</u>	.033	.033	.032	.052	.034	.027
		.039	.043	.033	.027	.030	.039	.048	.035	

$$\begin{aligned}k_1 - k_2 &= 9 \bmod 26, \\k_1 - k_5 &= 16 \bmod 26, \\k_2 - k_3 &= 13 \bmod 26, \\k_2 - k_5 &= 7 \bmod 26, \\k_3 - k_5 &= 20 \bmod 26, \\k_4 - k_5 &= 11 \bmod 26.\end{aligned}$$

将  $k_2, k_3, k_4, k_5$  用  $k_1$  表示, 我们有

$$\begin{aligned}k_2 &= k_1 + 17 \bmod 26, \\k_3 &= k_1 + 4 \bmod 26, \\k_4 &= k_1 + 21 \bmod 26, \\k_5 &= k_1 + 10 \bmod 26.\end{aligned}$$

因此, 密钥  $k = (k_1, k_1 + 17, k_1 + 4, k_1 + 21, k_1 + 10)$ ,  $k_1 \in Z_{26}$ . 通过尝试  $k_1$  的每一个可能的取值, 不难确定密钥字为 JANET,  $k = (9, 0, 13, 4, 19)$ . 因此, 我们得到与密文相对应的明文(添加上适当的标点符号)为

The almond tree was in tentative blossom. The days were longer, often ending with magnificent evenings of corrugated pink skies. The hunting season was over, with hounds and guns put away for six months. The vineyards were busy again as the well-organized farmers treated their vines and the more lackadaisical neighbors hurried to do the pruning they should have done in November. □

## 习 题

- 2.1 设英文字母 A, B, C,  $\dots$ , Z 分别编码为 0, 1, 2,  $\dots$ , 25. 已知单表仿射加密变换为

$$c = 5m + 7 \bmod 26,$$

其中  $m$  表示明文,  $c$  表示密文. 试对明文 HELPME 加密.

- 2.2 设英文字母 A, B, C,  $\dots$ , Z 分别编码为 0, 1, 2,  $\dots$ , 25. 已知单表仿射加密变换为

$$c = 11m + 2 \bmod 26,$$

其中  $m$  表示明文,  $c$  表示密文. 试对密文 VMWZ 解密.

- 2.3 在仿射密码体制中, 明文空间  $\mathbf{M}$  和密文空间  $\mathbf{C}$  都是  $(Z_q)^n$ , 密钥空间  $\mathbf{K}$  由对偶  $(L, b)$  构成, 其中  $L$  是  $Z_q$  上的  $n$  阶可逆方阵,  $b \in (Z_q)^n$ . 对任意明文  $x \in \mathbf{M}$  和密钥  $(L, b) \in \mathbf{K}$ , 密文  $y = xL + b$ . 求仿射密码体制的密钥量. 这里假设  $q$  是素数,  $Z_q$  是有限域,  $(Z_q)^n$  是  $Z_q$  上的  $n$  维向量空间.

- 2.4 试证: 先做  $m$  个字母的 Hill 加密, 再做  $n$  个字母的 Hill 加密, 结果是长度为

$\text{lcm}(m, n)$  的 Hill 加密变换. 这里  $\text{lcm}(m, n)$  是  $m$  和  $n$  的最小公倍数.

- 2.5 设英文字母  $a, b, c, \dots, z$  分别编码为  $0, 1, 2, \dots, 25$ . 已知 Hill 密码中的明文分组长度为 2, 密钥  $K$  是  $Z_{26}$  上的一个 2 阶可逆方阵. 假设明文  $\text{friday}$  所对应的密文为  $\text{pqcfku}$ , 试求密钥  $K$ .

# 第 3 章    Shannon 理论

在 1949 年,C. E. Shannon 发表了“保密系统的通信理论”一文. 该文对密码学的科学研究具有重要影响. 本章介绍 Shannon 关于密码学的一些思想.

## 3.1    密码体制的数学模型

Shannon 以概率统计的观点研究了信息的传输和保密问题. 通信系统和保密系统的组成分别如图 3.1 和图 3.2 所示. 通信系统的设计目的是在信道有干扰的情况下,使接收的信息无差错或差错尽可能地小. 保密系统的设计目的是使窃听者即使完全准确地接收到了信道上的传输信号也无法恢复原始信息.

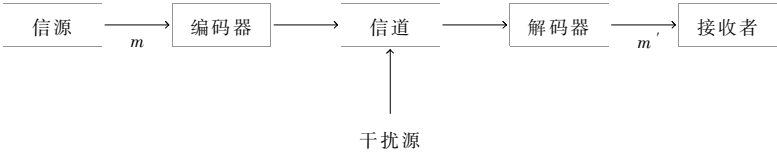


图 3.1    通信系统

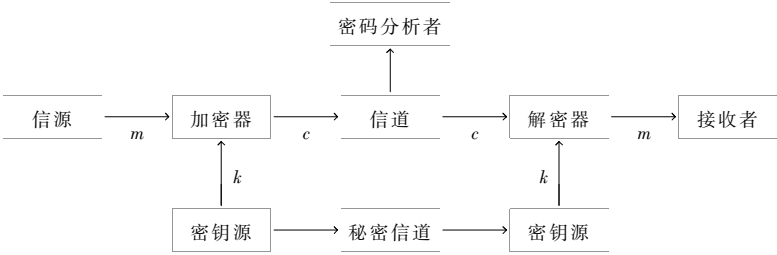


图 3.2    保密系统

在保密系统中,信源是信息的发送者. 离散信源可以产生字符或字符串. 设信源字母表为

$$X = \{a_i \mid i = 0, 1, 2, \cdots, q-1\},$$

其中  $q$  是一个正整数,表示信源字母表中字母的个数. 字母  $a_i$  的出现概率记为  $\Pr(a_i)$ ,  $0 \leq \Pr(a_i) \leq 1, 0 \leq i \leq q-1$ , 并且

$$\sum_{i=0}^{q-1} \Pr(a_i) = 1.$$

如果我们只考虑长为  $r$  的信源输出,则明文空间为