计算密码学

COMPUTATIONAL CRYPTOGRAPHY

走向数学丛书

卢开澄 著

责任编辑：孟实华　装帧设计：邱湘军

内容提要

保密通信是一个十分古老而又引人入胜的课题，在国防、商业、通信以及经济生活各方面有重要的意义。随着通信技术的不断发展，特别是计算机的广泛采用，密码的设计思想不断更新，密码体制在不断改进。

本书深入浅出地介绍了密码学的基本知识以及密码学的近代发展。作者不仅介绍了一些古典的加密方法，更着重讲述了70年代以来发展的 DES 数据加密标准和公开密钥体制，包括公钥体制的各种方案和与此有关的新的研究课题，如大数分解、数字签名、离散对数、知识证明、纠错码加密方法等，并有许多例供读者参考。书中还论述了数学的许多分（如概率统计、信息论、数论、置换群和有限理论、组合学以及算法复杂性理论等）的思想方法和结果（包括许多近代的理论数学成果）的相互交织，以及在保密通信领域中的应用。

ISBN7—5355—1584—3/G·1579

湘教（92）30 期　　定价：3.90 元

（湘）新登字 005 号

走向数学丛书

计算密码学

卢开澄 著

湖南教育出版社

前　言

王　元

从力学、物理学、天文学直到化学、生物学、经济学与工程技术，无不用到数学．一个人从入小学到大学毕业的十六年中，有十三、四年有数学课．可见数学之重要与其应用之广泛．

但提起数学，不少人仍觉得头痛，难以入门，甚至望而生畏．我以为要克服这个鸿沟，还是有可能的．近代数学难于接触，原因之一大概是由于其符号、语言与概念陌生，兼之近代数学的高度抽象与概括，难于了解与掌握．我想，如果知道讨论的对象的具体背景，则有可能掌握其实质．显然，一个非数学专业出身的人，要把数学专业的教科书都自修一遍，这在时间与精力上都不易做到．若停留在初等数学水平上，哪怕做了很多难题，似亦不会有助于对近代数学的了解．这就促使我们设想出一套“走向数学”小丛书，其中每本小册子尽量

用深入浅出的语言来讲述数学的某一问题或方面，使工程技术人员，非数学专业的大学生，甚至具有中学数学水平的人，亦能懂得书中全部或部分含义与内容．这对提高我国人民的数学修养与水平，可能会起些作用．显然要将一门数学深入浅出地讲出来，决非易事．首先要对这门数学有深入的研究与透彻的了解．从整体上说，我国的数学水平还不高，能否较好地完成这一任务还难说．但我了解很多数学家的积极性很高，他们愿意为“走向数学”撰稿．这很值得高兴与欢迎．

承蒙国家自然科学基金委员会、中国数学会数学传播委员会与湖南教育出版社支持，得以出版这套“走向数学”丛书，谨致以感谢．

前　言

在信息时代的今天，信息和其它资源一样成为人类重要的财富，而且信息本身还是时间，甚至于是生命．当前计算机被广泛应用而且日渐深入，几乎已无处不感到它的存在．但储存信息的计算机系统是很脆弱的，信息的传输则是依靠“不设防”的公共信道．信息作为一种资源被盗窃不同于其它的财富，首先不易被发现，而其后果可能更为严重．如何保护信息的安全？这个问题便尖锐地被提到议事日程上来．对它进行加密看来是有效而且可行的一种方法．它只需要付出很少的代价，便可得到满意的结果．

密码作为军事和政治斗争中的一种技术，已有悠久的历史，而且也有其不寻常的表现．但成为一门学科，还只是近几年的事．在计算机科学蓬勃发展的刺激下，70年代中期在这块园地开出两朵艳丽夺目的花，一是公钥密码系统，一个是数据加密标准DES．从此奠定了近代密码学的基础．在短短十多年时间里其发展可以说是异彩纷呈，令人目不暇接．数学在此过程中扮演了一个重要的角色．我们都听说数论中“1＋1”问题如何

如何，但数论在近代密码学中的贡献更令人神往. 读者将看到除数论外，信息论，概率统计，抽象代数中的群论和有限域理论，组合论，算法复杂性理论，编码理论，自动机理论，甚至于代数几何中的椭圆曲线等，都在密码学研究中找到了自己的位置. 这在其它学科中是不多见的现象. 这些内容饶有趣味，本书用通俗的方式讨论它们. 密码学的研究成果已不局限于通信保密上，如数字签名，身份验证等技术已是其它学科的基础.

本书分两个部分，密码学的基本概念和近代密码学的若干问题，有一些是我们工作的成果. 这样的划分其实是模糊的，比如数据加密标准无疑是近代密码学的一个方面，但放在了第一部分中，只是因为它的基本方法更接近于传统密码. 第二部分也不能简单地用公钥来概括. 近代密码学的突出特点是更多地依靠计算，所以本书命名为“计算密码学”. 作者认为传统密码和公钥密码是密码学中不可截然分开更不是对立的两个部分.

密码学作为一门学科非常年轻，但大有可为，这是现实的需要. 我国必须要有自己的数据加密标准，我们不能没有自己的密码系统. 作者希望本书对普及和推广这方面的知识会起到抛砖引玉的作用. 为了让更多的读者能接受讨论的内容，作者力求做到深入浅出. 但内容毕竟涉及到近代密码学的诸多方面，有相当的深度和难度，请读者能理解.

卢开澄

目　录

第一章　密码学若干基本概念

§1　引　论

密码作为一种技术已有上千年的历史. 自从人类有了战争，便自然产生了密码. 然而密码正式成为一门学科，还是近几年的事. 在计算机发展到网络的信息时代，信息本身就是时间，就是财富. 但信息存储于计算机系统中，信息的传输依赖于十分脆弱的公共信道. 信息的泄露不易被发现，但它造成的社会影响是巨大的. 所以，保护信息的安全是时代发展的必然要求，它已被迫切地提到日程上来.

密码是保护信息安全的有效而可行的方法. 它用很小的代价，为信息提供足够的安全保护. 在计算机蓬勃发展的刺激下，数据安全作为一个新的分支已活跃在计算机科学这个领域里. 不仅如此，它还是其他许多学科的基础和工具，被广泛地应用着.

数据加密标准 DES 和公钥密码体制，是 70 年代后半期在密码学园地上盛开的两朵奇葩，它们几乎是在相同的时间里提出的. DES 是 Data Encryption Standard 的缩写，由 IBM 公司研究并提出，1977 年经美国国家标准局批准，作为非机密机构保密通信用，最初预定服务期限十年. 至今，十年已经过去了，DES 还在“超期服役”中. 1976 年狄菲（W. Diffie）和赫尔曼（M. E. Hellman）在一篇著名的论文“New Directions in Cryptagraphy”中提出了公钥密码体制的构想，不久便推出公钥密码系统，可以毫不夸张地说：没有公钥密码和 DES 的研究，便没有近代密码学. 近代密码学的突出特点是更多地依靠于计算，公钥密码的研究异彩纷呈，当 Diffie 和 Hellman 提出他们十分卓越的思想时，还没有一个具体的实例. 但由于它的优势十分明显，所以，在这以后，各种公钥体制纷至沓来，真有点咄咄逼人的气势. 十五年过去了，应该说公钥尚未成功，还要继续努力. 这不能苛求于公钥本身，它毕竟才只有十几年. 十几岁对于一个人来说最多才是青年时期. 虽然如此，公钥密码的研究仍然光彩夺目. 除了计算机科学外，它还涉及数学中的数论、群论、有限域理论、信息论、编码理论、自动机理论、算法复杂性理论、概率统计，以及代数几何中的椭圆曲线等，这在其它学科中也是罕见的现象. 以上各方面在本书中都将一一论及.

讨论近代密码学无疑是本书的重点，公钥自然是中心内容，公钥是相对于传统的密码体制而存在的，所以在第一章里将简单地介绍一些传统的密码技术. 尽管如此，它的方法和引出的数学问题也是饶有趣味的.

我国必须要有自己的密码系统，也要有自己的数据加密标准，这是时代的需要. 而且唯此不能依靠进口，这是学科的性质所决定的. 开展密码的研究是当务之急，它除了依靠专业人

员外，群众性的研究也很重要，国外的经验也说明了这一点.

§2 保密通信是怎样进行的

若 A 要通过公共信道向 B 送去信息 m，由于公共信道缺乏足够的安全保护，信息 m 容易被第三者所窃取，甚至于被篡改，为此在 m 进入公共信道之前，先对它进行加密变换，得密文 c，即：

$$E: c = E_k(m)$$

其中 k 是参数，称为密钥，A 将密文 c 送给 B，B 收到后对 c 作解密变换，恢复 m，即：

$$D: m = D_k(c)$$

所以，加密变换 E 和解密变换 D 实际上是一对变换和它的逆，即：

$$D_k(E_k(m)) = m$$

相对于密文 c，m 称为明文.

加密通信的过程可用下图来表示：

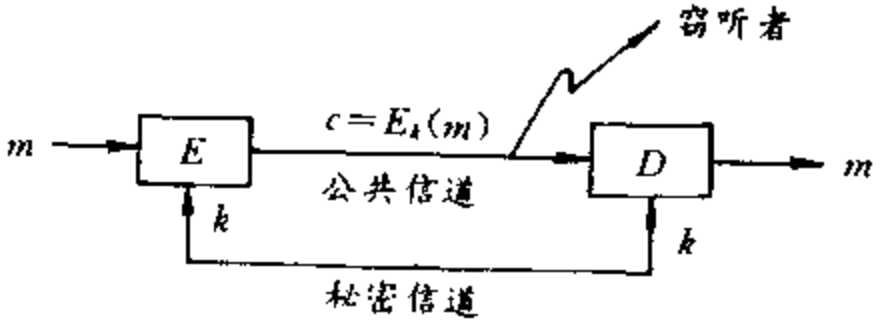

图 1.2.1

密钥 k 在保密通信中占有极其重要的地位. 它由通信双方秘密商定，只有他们双方掌握，第三者就是知道所用的加密算法 E 和解密算法 D，若不知道所用的密钥 k，也仍然无法获得明文 m，以此来达到保密的目的. 保密只是通信安全的一个目的. 还有一个目的是信息 m 的完整性，即要保证 B 收到的密文 c 不会被第三者篡改，即第三者若不掌握密钥 k，就无法伪造任何密文. 下面举例说明以上的概念.

例 1　凯撒 (Caesar) 密码　凯撒密码是将明文的每一个字母一律循环推移 k 位. 所以，凯撒密码也叫做单表密码，例如明文：

Secure message transmission is of extreme importance in information based society.

这段明文的意思是：在信息社会里，秘密通信是极其重要的.

现将明文字符通过凯撒密码后移 3 位加密得密文：

VHFXUH PHVVDJH WUDQVPLVVLRQ LV RI HAWUHPH LPSRUWDQFH LQ LQIRUPDWLRQ EDVHG VRFLHWB

这里凯撒变换是加密算法，k 是密钥. 本例 $k=3$. 可见凯撒密码并不安全，也就是说截到密文 c，在不知密钥的前提下也不难获得明文 m.

例 2　词组密钥密码　明文字母和密文的置换如下表：

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

即字母表 *abcde……uvwxyz* 分别和 *FIVESTAR……YZ* 对应，密钥为词组 *five stars*，这样明文：

Secure message transmission is of extreme importance in

information based society.

被加密成密文：

OSVQNS HSOOFAS PNFJOHBOOBKJ BO KT SXPNSHS BHLKNPFJVS BJ BJTKNHFPBKJ IFOSE OKVBSPY.

要对词组密钥密码系统的密钥采取强行(brute force)搜索的方法，则要面临26个字母的全排列的各种可能，根据Stirling公式：

$$n! = \sqrt{2n\pi}\left(\frac{n}{e}\right)^n$$

可求得 $26! \approx 4 \times 10^{26}$

若以每秒可搜索10^7种排列的超高速电子计算机进行穷举，需要的时间为

$$\begin{aligned} T &= 4 \times 10^{26}/(365 \times 24 \times 3600 \times 10^6) \\ &= 4 \times 10^{26}/(3.1536 \times 10^{12}) \\ &= 1.27 \times 10^{13}(\text{年}) \end{aligned}$$

显然，采用蛮攻击或强行搜索的方法是行不通的．在下一节里将介绍一种统计分析的方法．

例3 蒲内费厄(Playfair)密码

蒲内费厄是英国的密码学家．图1.2.2是个5×5方阵．

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

图1.2.2

矩阵中元素TSINGHUAVERYBCDFKLMOPQWXZ正好包含除了J以外的其它25个英文字母．前面12个字母是依TSINGHUA UNIVERSITY顺序取出前面未出现的字母，后面跟

以依英文字母表中前面未出现的英文字母. 加密算法由明文字母对 m_1m_2 而定. 设 m_1m_2 的密文为 c_1c_2. 加密法则如下:

(a) 若 m_1 和 m_2 在同一行，则 c_1 和 c_2 分别是 m_1 和 m_2 右边的字母，这里将第 1 列看作是在第 4 列的右边.

(b) 若 m_1 和 m_2 在同一列，则 c_1 和 c_2 分别为 m_1 和 m_2 下方的字母. 第 1 行看作是位于第 4 行的下方.

(c) 若 m_1 和 m_2 不在同一行或同一列，则 c_1 和 c_2 为矩形的两个顶点，且该矩形的其它两顶点为 m_1 和 m_2，且 c_1 和 m_1 同行，c_2 和 m_2 同行.

(d) 若 $m_1=m_2$，则在明文 m_1 和 m_2 之间插入一空字符(设为 X).

(e) 若明文字符数是奇数，则在明文末端加上一空字符.

例 4 利用图 1.2.2 的方阵，对明文

BEIJING CHINA

进行加密. 在这里 J 当作 I 处理，与图 1.2.2 对应的密钥为 *TSINGHUA UNIVERSITY* 先将明文分成两个字符一组

BE IX IX IN GC HI NA

分别加密得

DA NW NW NG ND AT IV

蒲内费厄密码加密的结果，一个字母对应的密文并不固定，这不同于词组密钥密码系统.

例 5 费尔南(Vernam)密码

费尔南密码假定明文 m 用 n 位的 0，1 符号串来表示，密钥 k 也是 0，1 符号串，设

$$m = m_1m_2\cdots m_n,\ m_i = 0 \text{ 或 } 1,\ i = 1,2,\cdots,n$$

$$k = k_1k_2,\cdots k_l,\ k_j = 0 \text{ 或 } 1,\ j = 1,2,\cdots,l$$

$$E_k(m) = c = c_1c_2\cdots c_n,\ c_i \equiv m_i + k_i(\mathrm{mod}\ 2)$$

$$i = 1,2,\cdots,n$$

所以，费尔南密码的弱点在于若已知密钥 k 的一组明文和它对应的密文，则费尔南密码便被攻破.

设密文

$$c_l = c_1^{(l)}c_2^{(l)}\cdots c_n^{(l)},\ l = 1,2$$

分别对应于明文

$$m_l = m_1^{(l)}m_2^{(l)}\cdots m_n^{(l)},\ l = 1,2$$

即 $$c_i^{(l)} = k_i \oplus m_i^{(l)},\quad l = 1,2,\ i = 1,2,\cdots,n$$

则
$$\begin{aligned} c_i^{(1)} \oplus c_i^{(2)} &= m_i^{(1)} \oplus k_i \oplus m_i^{(2)} \oplus k_i \\ &= m_i^{(1)} \oplus m_i^{(2)} \end{aligned}$$

若 $m_i^{(1)}$ 已知，则 $m_i^{(2)}$ 便可得到.

一般地，密钥 k 的长度 l 有限，可以周而复始重复地出现. 也可以用长度为 l_1 和 l_2 的两个密钥

$$k_1 = k_1^{(1)}k_2^{(1)}\cdots k_{l_1}^{(1)}$$

$$k_2 = k_1^{(2)}k_2^{(2)}\cdots k_{l_2}^{(2)}$$

只要 l_1 和 l_2 互素，由 k_1 和 k_2 可以产生周期为 l_1l_2 的密钥比特流

$$K = k_1k_2\cdots k_n,\quad n = l_1l_2$$

$$k_i = k_i^{(1)} \oplus k_i^{(2)},\quad i = 1,2,\cdots,l_1l_2$$

这里 $k_i^{(1)}$ 和 $k_i^{(2)}$ 都分别是由 k_1 和 k_2 产生的周期比特流.

若费尔南密码的密钥 k 是一组不重复的随机流，则这样的密码称之为一次一密. 一次一密密码是完全保密密码. 完全保密的概念见第一章 § 8 商农理论.

§ 3 统计分析法

在密码学中加密和破译是一对矛和盾，所以，了解破译技

术对于研究密码也是必不可少的. 在词组密钥的密码系统中，经过字母的变换使得明文面目全非，然而并非没有留下蛛丝马迹. 例如英文字母出现的频率差别很大，这就给密码分析者以可乘之机. 大量的统计表明，虽然统计的对象迥异，但统计结果表明，各个字母各自出现的频率却惊人地接近.

下面是一组统计结果

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		

图 1.3.1 是它们的频率图：

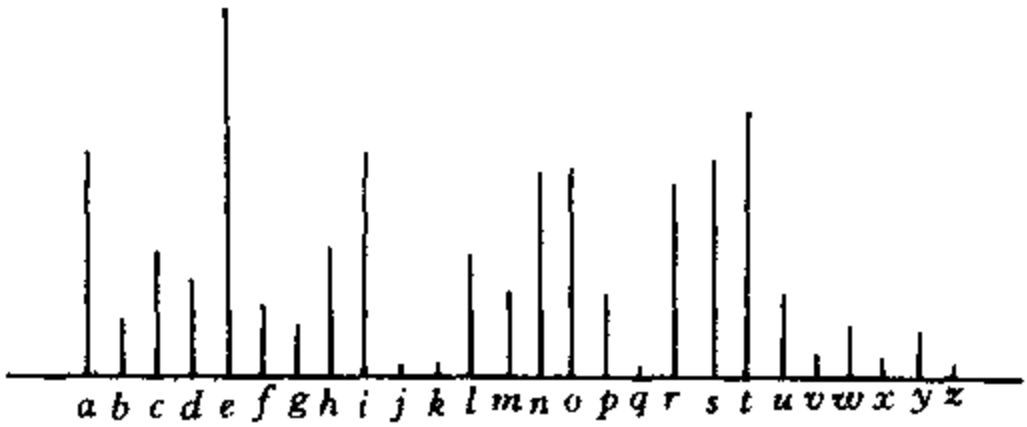

图 1.3.1

从频率图可以看出以下几个比较明显的差异：

1. 字母 e，t，a，o，n，i，r，s，h 的频率较高，其中 e 尤为突出.

2. 字母 d，l，u，c，m 是中频.