实用加解密技术

——BASIC与C++语言实现

（美）Gilbert Held 著

刘乃琦 晏 华 译

电子科技大学出版社

内 容 提 要

本书从实用性的角度，针对信息（数据）的加解密问题进行了详细讨论。论述了密码学的基本原理及概念，保密领域内若干重要术语，加解密实用技术和应用实例，阐述了单表替换、多表替换、随机数替换型加解密技术及其实现，以及公钥加密体制及其语言实现。并且以 C++和 BASIC 两类常用计算机编程语言来实现各种加解密算法，若干实例都具有实用价值。

本书内容由浅入深，在诸多实例和程序实现的基础上详细介绍了微机应用领域中常见的加解密技术，并给出了算法步骤和编程提示。尤其是书中的许多思路和编程技巧，对从事加解密技术应用和计算机安全、保密学研究和应用的人员以及高等院校学生都具有很大的参考价值。

图书在版编目（CIP）数据

实用加解密技术：BASIC 与 C++语言实现/（美）赫尔德（Held，G.）著：刘乃琦，晏华译.—成都：电子科技大学出版社，2000.9

ISBN 7－81065－528－0

Ⅰ.实… Ⅱ.①赫…②刘…③晏… Ⅲ.①密码-加密②密码-解密译码 Ⅳ.TN918.4

中国版本图书馆 CIP 数据核字（2000）第 48818 号

实用加解密技术——BASIC 与 C++语言实现

（美） Gilbert Held 著

刘乃琦 晏 华 译

出 版：电子科技大学出版社出版 （成都建设北路二段四号，邮编：610054）

责任编辑：王仕德

发 行：新华书店经销

印 刷：四川导向印务有限公司

开 本：787×1092 1/16 **印张** 20 **字数** 484 千字

版 次：2000 年 9 月第一版

印 次：2000 年 9 月第一次印刷

书 号：ISBN 7—81065—528—0/TP・348

印 数：1—4000 册

定 价：36.00 元

译 者 序

在信息社会中，信息（数据）的安全是人们面临的一个重要问题，随着社会和人类对信息的依赖性日益增强，信息的安全已经成为未来与人们生存和发展休戚相关的问题。信息安全领域中的一个重要分支是信息的保密。保密学的范围如今已大大扩展，既包含了传统密码学的领域，也扩展到防止信息被窃取、扩散、复制和滥用等方面。对数据的加解密、技术实现及其应用，一直是用户所关心的重要问题。

本书从实用性的角度，针对信息（数据）的加解密问题进行了详细讨论，介绍了密码学的基本原理及概念、保密领域内若干重要术语、加解密实用技术，列举了应用实例，并且以 C++和 BASIC 两类常用计算机编程语言来实现各种加解密算法，书中若干实例都具有实用价值。

本书内容分为八章：第一章首先对密码学、加解密技术、现代计算机应用和电子函件传输中的关键概念和术语进行了解释，给读者建立了一个基础知识结构；第二章基于单字母替换概念，讨论了该类型加解密技术的实现，并给出程序实例；第三、四两章分别基于关键字和替换原则，阐述了加解密技术；第五章和第六章分别讨论了多字母替换型和随机数替换型加解密技术及其实现；第七章讨论了加解密技术的程序实现和开发；第八章讨论了公钥加密体制及其语言实现。

本书内容由浅入深，在诸多实例和程序实现的基础上详细介绍了微机应用领域中常见的加解密技术，并给出了算法步骤和编程提示，尤其是书中的许多思路和编程技巧，对从事加解密技术应用和计算机安全、保密学研究和应用的人员、高等院校学生都具有很大的参考价值。

译 者

2000 年 7 月

目　录

第一章 技术与术语

本章旨在介绍密码学领域内的重要术语，让读者获取和理解该领域的知识结构，提供开发和使用实际加密技术的能力，从而将通过不同电子函件系统传输的数据与信息的真实含义隐藏起来。本章所涉及的大多数技术对一个训练有素的分析员来说是完整的，然而，在选择不同的信息安全保护级别时，他们也可能要花费很多时间才能理解这些数据的真实含义。实际上，本书讨论的加密技术需要 600 亿次试验才能破解并获取数据的真实含义。每一次试验都需要对截获的消息进行打印，而对 10 行消息的理解，一个人可能要扫视 6 000 亿行数据。

美国国家安全局（NSA）、英国情报局 MI5 和前苏联克格勃都采用了超级计算机系统来对截获的消息进行处理，以获得其真实含义。但很少有商务机构肯花费如此大的费用来构建这些软件和硬件系统，并完成这类操作。本书提供了更高级的技术，以形成一种机制，来防止无意的、别有用心的和非法入侵者获取用户通信的真实含义。

为了给读者提供密码学领域中的相关术语，我们通过各种详细实例，讨论了不同的术语及其含义。首先，我们回顾和复习了密码的主要类型，并将在后继章节中对它们进行详细讨论。在适当时候，还将讨论与传统密码不同的方法，不去讨论传统密码的进展，而是展示其在传统方法基础上的实际应用。

本章第二个目的是讨论并设计开发短小例程和程序，用户可以选择C++语言或者BASIC语言来编程和进行文件名变换。

一、密码与编码

很多读者容易混淆和模糊的一个主要领域是密码系统和编码系统，虽然两者都可以用于隐藏传输信息，但是它们采用的是不同的替换技术。

在密码系统中，明文表示被加密以前的一组数据，在处理时并不注重它们实际的含义。而在编码系统中，单词或者词组被其他的单词或词组所替换，从而隐藏了它们的原义。所以，消息 FIRE THE CORPORATE LAWYER 经过简单加密（见后所述），传输时变为：GJSF UIF OPSQPSBUF MBXZFS，而以编码方式传输，消息可以被替换为：SELL THE CORPORATE DONKEY, 这里，单词 SELL 对应于 FIRE，而 DONKEY 则对应于 LAWYER。

对于一条消息，许多编码系统一般不会改变太多的术语含义，因此，对消息内容本身就提供了一种很直接的暗示。而且，一系列消息的展现，也会提供一组合理的推测，从而使有的人可能确定编码消息的含义。使用编码的另一个障碍，尤其是当它们很少使用时，

要求预先熟悉和掌握所有的替换码。某些军用和外交专用的编码系统基于一种编码本（密码本），这种编码本含有成千上万条术语和对应的编码。编码本的准备不仅困难和耗时，而且，一旦丢失或者泄露，就形同噩梦缠身，因而不得不用新的编码本来替换。虽然军方和外交领域仍然在某种特殊情况下使用编码系统，但有效的编码应用需要在原来提供基本消息编码和解码的编码系统基础上进行改进和开发。相比之下，大部分密码系统却只需要记住一个或者多个称为密钥的信息项，就可以对消息进行加密和解密。实际情况是，加密和解密操作比基于传统编码本的编码和解码操作更容易实现，特别是针对较长的消息内容。此外，许多密码系统可比编码系统获得更高的消息安全保护级别，因此，本书将重点讨论不同的加密技术。

附一：

如果读者是一个电影爱好者，可能还记得一部战争影片："最长的一天，虎、虎、虎"，以及内容相似的电影，读者可能注意到了当时消息的传送，其显现的内容是毫无意义的。比如消息内容是："Paul has a long arm,"（鲍尔有一条长手臂），"The soup is in thekitchen,"（肥皂在厨房里），以及"Ralph needs shoes,"（拉尔夫需要鞋）等。它们表示了一种已经编码的消息，其中的单词或者词组已经由另外的单词或者词组所替换，以便隐藏消息的真实含义。在另一部影片中，读者很可能记得一个士兵或者水兵使用一种仪器来截获消息，并且制作一种包含消息内容的穿孔纸带。纸带接着被放入一台专用的纸带阅读机，该机器阅读送入的纸带并产生新的纸带，这个士兵或者水兵再阅读新纸带上的消息，继而走过房间并说道："我们得把这个消息送给司令"等等。

这些电影都展示了专门解密机器的使用，这些机器的设计目的是读取已加密的消息，重新产生消息的原始文本（即明文）。穿孔在纸带上被截获的消息包含了一系列看似随机选取的字符，比如，"QAFRT…"，这个字符序列实际上代表了一个已加密的消息。在这个消息中，每一个明文字符都按照预先确定的一个算法用一个密文字符来替换它。

二、密码术语

如前所述，密码系统是一种替换系统，替换过程的发生针对单个字符或者成组字符，而不管它们的实际含义。实际的加密过程可以通过硬件和软件来完成，也可以借助一人或者多人的脑力，以人工的方式完成。

1. 明文与密文

密码系统的使用，是针对一组消息或者文本。原始的、未改动的消息或文本的内容被称为明文（plaintext 或者 cleartext）。通过加密系统，名文消息的真实含义被隐藏了，这个隐藏过程被称为加密（enciphering），而最后得出的结果则被称为密文（enciphered text）。

2. 加密

图 1.1 展示了一种加密过程，以框图形式描述。图中，明文 x 通过加密过程 E 变换为密文 y，构成一个函数：$y=Ek(x)$，这里 k 称为密钥。加密过程可被认为是一种算法，它在基于密钥 k 的基础上对明文进行运算。这里，密钥定义了加密算法的运算规则，不同的密钥会对同一个明文产生不同的密文。

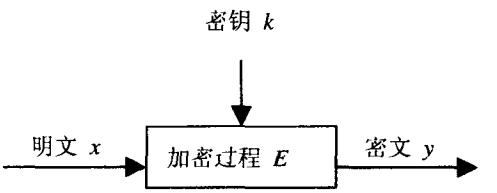

图 1.1 加密过程

例如，通过构建一个模 2 算法来实现一个现代密钥，完成加密过程。读者可能已经生疏了取模算法，这里，再进行一个简单的复习。采用模 n 算法进行加法时，将被加数除以 n，保留余数作为模 n 算法的结果。例如，我们熟悉的十进制运算，以模 10 进行的 6 加 8 的运算的结果是 4。表 1.1 列出了基于模 2 的四种可能的数位位置。

表 1.1 模 2 加法

0	0	1	1
0	1	0	1
0	1	1	0

为了展示采用模 2 加运算的加密过程，假定对单词 HELP 加密，它采用 7 位 ASCII 码形式存放，并假定密钥为单词 BURP，这样，表 1.2 就展示了这个加密过程。

表 1.2 采用模 2 加法的加密过程

明文	H	E	L	P
	1001000	1000101	1001100	1010000
密钥	B	U	R	P
	1000010	1010101	1010010	1010000
加密结果	0001010	0010000	0011110	0000000
	LF	DLE	RS	NUL

可见，通过模 2 加的加密过程，明文 HELP 被变换为一个 28 位的字符序列，表示成四个控制字符：换行符（LF）、数据连接回退符（DLE）、记录间隔（RS）和空字符（NUL）。如果有人截获了这个加密文本，他只能得到 LE、DLE、RS、NUL 这样的字符序列，而不

是明文 HELP。

3. 解密

将一个密文变换回或者恢复为它原来的明文的过程称为解密。图 1.2 以框图形式展示了解密过程。解密过程 D 被认为是加密过程的逆过程，需要有一种算法将密文变换回它原来的明文。解密过程也由一个密钥 k 来控制，它作用于密文 y 而产生明文 x，满足公式：$x=Dk(y)$。

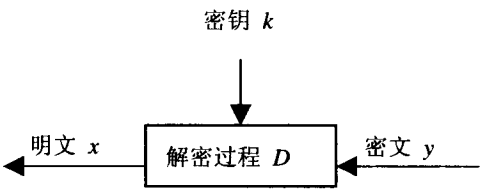

图 1.2 解密过程

通过对前面加密的消息进行解密，可以清楚地理解解密过程。例如，假定我们收到了消息 LF DLE RS NUL，而且也掌握了密钥 BURP，我们则可以采用模 2 减法来还原明文消息。这里，需要了解模 2 减法的运算过程。

在模 2 减法运算中，如果被减数大于减数，结果就是减法所得数。如果被减数小于减数，则在做减法之前，先将下一位基数值加到被减数上，然后再做减法。例如，对十进制运算，9−5 的结果是 4，而 5−9 的结果则是 6，因为被减数 5 上先加上了一个十进制的基值 10，然后再减去 9 得到 6。

由于前述的加密过程采用的是模 2 加法，这里，我们则用模 2 减法来解密。表 1.3 列出了模 2 减法的所有四种可能的位组合。

表 1.3 模 2 减法

0	0	1	1
0	1	0	1
0	1	1	0

熟悉了模 2 减法，我们再来看看解密过程。如果我们是消息的接收者，接收到了已经加密的消息 LF DLE RS NUL，而且也掌握了密钥 BURP，解密过程则很容易实现，如表 1.4 所示。

表 1.4 采用模 2 减法进行的解密

加密文本	LF	DLE	RS	NUL
	0001010	0010000	0011110	0000000
密钥	B	U	R	P
	1000010	1010101	1010010	1010000
解密文本	1001000	1000101	1001100	1010000

4. 密钥与密钥空间

从图 1.1 和图 1.2 中我们看到，用于加解密的密钥是控制加解密算法运算的关键。即使

某人获得了已加密消息的副本，也了解消息加密运算的算法，但他也必须使用正确的密钥才能成功破译加密过的消息。因此，构成密钥的所有可能值的数量被称为一个算法的密钥空间，这个密钥空间对一个加密算法来说是非常重要的。具有有限密钥空间的算法是非常容易采用穷举法不断摸索而破译的，它可以对截获的消息采用不同的密钥试验，最后破译加密消息。例如，用 9 个十进制数序列构成一个密钥，其密钥空间数量为 10^{10}，或者说有 10 亿种密钥值。

虽然采用计算机可以很容易地产生所有 10 位数的排列，但不能指望人都能够记住这些序列。因此，如果我们想开发一种密码能够适合人工处理，就不得不在密钥空间和实际应用之间进行折衷。此外，从实用观点出发，人们容易记住单词和词组，而不是数字组合。本书中很多密码系统都基于采用单词或词组的字符密钥，这样，就比使用数字序列作密钥更加增强了记忆。尽管通过个人计算机可以自动地进行加密过程，但在某种情况下，为了适应程序处理，也采用单词或者词组，而不是数字序列。采用字母或者字母数字组合的密钥（不单是数字序列）极大地扩展了密钥空间。例如，两位十进制数字的密钥可以提供多达 10×10=100 种惟一的密钥。两个大写字母和两个小写字母组成的密钥序列可以达到 26×26=676 个惟一的密钥，而两个字母数字（字母限制为单一的大写或者小写）可以组成 36×36=1296 个惟一的密钥。

表 1.5 展示了一个密钥空间，该空间由字母、数字和字符组成，长度可以从 1 到 20 位，表中字段宽度（Field Width）这一列表示密钥中的字符个数。注意，除非特别指明，当采用字母或者字母数字作密钥时，我们把密钥空间的讨论局限在所有的大写或者小写字母上。同时采用大、小写字母可以大大增大密钥空间，当然，也存在潜在问题。即虽然单词和词组很容易记忆，但其中哪些是大写字母，哪些是小写字母却很容易忘记。因此，本书采用的密钥通常限制为单一模式字母（要么大写、要么小写）。例如，密钥长度为 8 位，可以用单词 computer，此时，可以键入 computer 或者 COMPUTER，但不能使用 Computer，除非编写一个程序来分辨字母的大小写，并把所有字母都转变为一种模式。

表 1.5

字段宽度	数字组合	字母组合	字母数字组合
1	1.00000E+01	2.60000E+01	3.60000E+01
2	1.00000E+02	6.76000E+02	1.29600E+03
3	1.00000E+03	1.75760E+04	4.66560E+04
4	1.00000E+04	4.56976E+05	1.67962E+06
5	1.00000E+05	1.18814E+07	6.04662E+02
6	1.00000E+06	3.08916E+08	2.17678E+09
7	1.00000E+07	8.03181E+09	7.83642E+10
8	1.00000E+08	2.08827E+11	2.82111E+12
9	1.00000E+09	5.42950E+12	1.01560E+14
10	1.00000E+10	1.41167E+14	3.65616E+15
11	1.00000E+11	3.67034E+15	1.31622E+17
12	1.00000E+12	6.54290E+16	4.73838E+18

13	1.00000E+13	2.48115E+18	1.70582E+20
14	1.00000E+14	6.45100E+19	6.14094E+21
15	1.00000E+15	1.67726E+21	2.21074E+23
16	1.00000E+16	4.36087E+22	7.95866E+24
17	1.00000E+17	1.13383E+24	2.86512E+26
18	1.00000E+18	2.94795E+25	1.03144E+28
19	1.00000E+19	7.66467E+26	3.71319E+29
20	1.00000E+20	1.99281E+28	1.33675E+31

让我们观察整个表 1.5，其中，E+6 代表 100 万，E+9 代表 10 亿，一个 6 位数的字母数字组合可以产生 21.7 亿种密钥（这里，还限制字母要么全是大写字母，要么全是小写字母）。读者还可以注意到，一个 9 位数段的密钥，当只限定为数字时，其所有数字组合也达到了 10 亿种。于是，我们就可以采用较少的密钥字符来维持一个大的密钥空间，也可以通过采用字母数字组合（而不仅仅采用单独的数字或者字母）来增大密钥空间。

5. 加密系统类型

目前，加密系统（也称密码体系）的类型有两类：私钥和公钥系统。前述采用同一个密钥完成加解密操作的系统被称为采用私钥的加密系统。这里，密钥必须保持一种“私有”性，不能够泄露给那些想要接收加密信息的人。此外，一旦密钥被泄露，即意味着凡是采用该密钥加密的消息都可能泄密。私钥系统通常也被称为“对称（symmetric）”密钥系统，对称的含义表示同样一个密钥被用于加密和解密两种处理过程。私钥系统存在的最大问题在于密钥的生成和分配，例如，采用私钥加密系统的双方，如果需要传送加密消息，那么双方只需要采用一个单一的密钥。但是，如果有三方需要彼此之间传送加密消息，就应当有一个公共的密钥。然而，如果这个密钥被泄露，三方之间传送的任何消息都可能泄密。于是，很多机构采用公钥加密系统，即在不同的人之间采用不同的密钥。这样，在三方或者三个合作机构之间，就需要三个不同的密钥。要连接四方或者四个机构，则需要四个密钥。

类似地，如果对 n 个用户或者机构，所需要的单独的密钥数量为 2^n-2。这对于用户数或机构数较少的情况还不成问题，但是当用户数增加时，涉及密钥生成和分配的注册过程就变得很难管理。特别在使用现代电子函件系统时，需要与成百上千的人通信，此时，采用私钥的加密系统很明显不适宜了。而且，考虑通过浏览器访问的“万维网”（WWW）上进行电子商务活动的安全性，则私钥加密系统的使用就更不可能。幸好数学家们开发了另一类加密系统，极大地缓减了前述的与私钥加密系统有关的密钥生成与分配问题，这一类加密系统就是公钥加密系统。

公钥加密系统在 20 世纪 70 年代后期提出，数学家们发现了某些数学关系，即用于加密一段消息的一个密钥，不能再用于解密，而需要用另一个密钥（它形成一个配对密钥）才能解密。这种数学关系需要采用两个密钥，一个作为私钥，另一个可以公开提供给任何一个人，甚至可以在网络站点的主页上公布，这个密钥被称为公钥。

在公钥加密系统中，想要发送加密消息给他人或机构的人，要取得接收方的公钥，并用它来加密消息，接收方采用他们自己的私钥来解密这个加密消息。这样，采用公钥加密系统的人或者机构，就只需要保留两个密钥，一个公钥和一个私钥，而不管系统中有多少人和机构。那么，有 n 个用户的系统，总的密钥数量有 $2n$ 个。因此，当 n 很大时，所需要的密钥数量就从原来私钥系统的 2^n-2 个减少了。

公钥加密系统的使用，提供了单向的加密和解密的密钥，它取决于所用密钥的类型，即，公钥用于加密，而私钥用于解密。所以，公钥加密系统也被称为非对称（asymmetric）加密系统。

按照数学家的观点，密文 $E_k(x)$ 是采用一个公钥 (k_p) 来加密，这个公钥与一个私钥 (k_r) 具有一种数学关系，不像在私钥系统中，$D_k[E_k(x)]$ 可以引起原始明文的重构，如果不掌握 k_r，要从 k_p 中计算 $D_k(x)$ 是很难的。然而，当知道私钥 (k_r) 时，$D_k(x)$ 就很容易计算。采用公钥系统的关键，是人们产生他们自己的公钥和私钥的能力。这样，他们就可以自由地将公钥分配给他们愿意以安全方式通信的人。通过采用指数和大质数，可以形成非常长的公钥和私钥。虽然其他人也可能采用类似的算法来产生公钥和私钥，企图破译另一些人加密的消息，然而，获得这个长质数因子的时间甚至需要用超级计算机计算十余年。因此，可以认为，适当地选取公钥和私钥对，在今后可以对消息的安全性提供一种非常高的保护级别。

6. 密钥生成、管理与分配

如果所选择的密钥组合很容易被猜测出来，一个能产生千万组密钥组合的加密系统可能是非常脆弱的。例如，生日日期、家庭住址等最好不要作为密钥，因为采用穷举法的破译者，总是采用它们作首选试探，而一个毫无经验的人却往往采用这些字符组合作为一个密钥。在理想情况下，密钥的产生应当基于某种随机过程。然而，如果一个人正在旅途中，要用笔记本电脑发送或接收私人信息，而这些信息又按照预定密钥加了密，这时一个人很愿意选择并记住字符组 PCATV，而不是 Q4R51，因为前者比后者更容易被猜测到。读者可以考虑采用随机过程来产生密钥，用于预定的日期时间，并把它们存放在一个计算机文件中，而这个文件又采用另外的密钥加密。然后，使用者被要求在一定的时间期限内记住一个密钥，以便查找其他的密钥，同时给每一个密钥的使用规定一个有效的时间期限，这是与密钥管理有关的许多技术中的一种。

密钥管理是针对密钥的生成、维护和分配的过程，它还负责确定当一个密码系统或者系统的密钥丢失或者被泄露的情况下需要采取的各种措施。

由于组织机构的不同和各种因素的变化，密钥管理的过程以及与此相关的功能也不同。这些因素可能包括掌握加密密钥知识的雇员的人数，他们所在的地理分布和企业内部函件服务的整体性等等。另外，密钥系统的类型，是公钥还是私钥，都将极大地影响密钥分配的方法。有的机构将公钥公布在他们的 Web 站点上，或者通过内部办公函件以类似计算机口令的方式分配。另一些组织则通过注册函件来分发密钥，通过相互合作采用当前密钥加密的电子函件系统，或者以明文方式分发密钥。显然，以明文方式或者以当前密钥方式加密的新私钥的重复分发，会给组织机构带来潜在的泄密危险。不过，我们必须注意，很多

机构只是简单地采用一种机制来保持个人之间的通信，而不去寻找和选择一种详尽的方法以保证他们之间的通信不会被破坏。虽然有些复杂精巧的分配密钥的方法将在后继章节中阐述，但我们仍然将重点先放在实际的、有效的密钥分配方法上。

不管新密钥分配的方法如何，密钥总是要不断分配的，因为雇员可能离开公司，也可能丢失他们的密钥，或者把含有密钥的文本放在不该放的地方，以及做了失密的事，事后才意识到，等等。

附二:

或许最令人感兴趣的公共安全泄密事件，是发生在当时美国总统叠放在自己外衣口袋里的文件被拍照。文件当时是以这样一种方式折叠的，拍照后，一个密钥清楚地显现在照片上，并刊登在报纸上。万幸的是，涉及国家安全的这个关键字只表示了与一种专门通信类型有关的类别，并没有反映用于加密通信的密钥。但是，却给涉及国家安全的许多人员带来了麻烦。对他们来说，关键字的公开曝光导致了大量的、无报酬的超时加班，要求他们迅速地改换已经暴露的、用于通信类型分类的关键字。

在某些联邦机构内，私钥通常一个季度改换一次；而另一些机构则每月、每周，甚至每天改换密钥一次。这种频繁改换密钥的情况，通常在数据内容经常被加密隐藏，而且这些数据受到潜在威胁的情况下才是值得的。对于后者很难给予界定，因为很多人并不知道他们的电子交谈已经被窃听了，等他们知道时已经晚了。至于潜在威胁，我们必须在密钥改换和在一定时期内开发和分配新密钥之间进行平衡。如果我们设计的系统实现起来很困难或者必须花很大的努力才能有效地使用，很多用户就不会使用这个系统。因此，“保持简练”这个古老的军事格言是值得参考的，在密钥生成和分配方法的选择中也应当注意。

三、密码类型

在大学图书馆或公共图书馆中，我们可以找到许多阐述密码系统的书籍，例如，“破译者”（David Kahn, Macmillian,1969），“密码分析”（Helen F.Gaines, Dover Publications, 1956）都对密码系统作了很好的描述，特别讨论了这些系统的历史应用。最近出版的“计算机网络安全”（D.W.Davies & W.L.Price）一书，以数学为基础向读者展示了最新开发的密码系统结构。这些书的内容包含了数百种密码（加密）系统，从最基本的手工加密过程，到需要采用计算机的非常复杂的密码系统。通常，每一种密码系统都可以被分类成两种基本类型：替换型密码（substitution）和换位型密码（transposition）。

下面我们将简单介绍密码分类，并讨论用于开发专用类型密码系统的若干密码系统，并重点考虑用于明文密文变换的算法和密钥空间，以便使加密文本可以用于许多电子函件系统中。对替换型密码和换位型密码系统的讨论基于私有密钥（私钥），当然，开发基于公

钥的替换型和换位型密码系统也是十分可行的，前者于 20 世纪 70 年代推出，后者的历史甚至可以追溯到凯撒（Caesar）时代。从历史的观点来看，我们所要讨论的密码系统都是基于私钥的。因此，在这一章中，我们将阐述加密系统的发展，主要检测替换型密码和换位型密码系统，及其与私钥的关系。

1．替换密码系统

在替换型密码系统中，把一种算法运算作用在出现的明文字母序列上，使每一个明文字母都由一个密文字母所替换。在简单替换型密码系统中，替换过程是不能改变的。也就是说，每一个明文字母都由对应的同样的密文字母所替换。例如，众所周知的“凯撒密码”，在这种密码系统中，密文代表了一种固定的字母替换，将明文 KILL ALL THE LAWYERS 加密为 LJMM BMM UIF MBXZFST，这里用一个字母位置替换作为加密算法。

把上述替换的明文和密文进行比较，我们可以发现两者之间的关系。如果将明文和密文的字母都限制在 26 个大写英文字母中，对于简单单字母的凯撒密码替换，明文(P)和密文(C)之间的关系就很明显了。在这个例子中，密文的字母是由明文字母表中字符错一位而得到的，其关系可以描述成 $P_A=C_B$ ，即明文中的字母 A 就对应于密文中的字母 B。这样就有

明文字母表(P): ABCDEFGHIJKLMNOPQRSTUVWXYZ

密文字母表(C): BCDEFGHIJKLMNOPQRSTUVWXYZA

在可变替换型密码中，密文对明文的替换会导致另一种可能性，这种情况出现的概率很高，即明文中重复的字母可以由不同的密文字母所替换。这样，就需要一个以上的字母映射来把明文变换成密文，这种方式被称为多字母替换。

作为一种实现方式，一个随机数或者伪随机数序列可以用来产生可变的密码替换。由密钥产生的伪随机数序列构成的可变替换，对于开发商务加密系统，如众所周知的数据加密标准（DES）都是一种基础。在后继章节中，我们将讨论单字母（单表）替换、多字母（多表）替换，以及采用随机数的加密系统。

附三：

第二次世界大战前夕，美国军事信号组织成功地建造了一种密码机，称为 PURPLE，可以复制日本的一种加密机的加密操作。截获日本的传输消息被送入 PURPLE，产生出明文信息，同时被翻译成英文，然后以称为 MAGIC 的编码方式发送给美国总统罗斯福，以及有关的高级内阁成员和军队的高级将领们。虽然曾经有若干来自东京，并传送到位于华盛顿的日本大使馆的紧急消息被截获和破译，但是，这些仍然属于绝密的情报并没有阻止珍珠港事件的发生。因为很简单，日本人从来就没有传送过任何要攻击珍珠港的消息。无论什么原因，都缺乏一种明显的攻击消息，在美国政府中滋长了一种自满情绪。具有审阅已经加密消息的能力，而且这些消息很重要，但并不意味着总是能够理解消息中某些扩展的含义。在很多情况下，阅读已加了密的消息的能力，并不能代替对消息内容的分析。例如，1942 年 6 月，美国海军情报所解码了截获的日本情报，注意到一个日本机构把中途岛（Midway）作为了其邮局地址，通过集结美国海军对抗了日本的威胁，中途岛战役最终代表了太平洋战争的一个转折点。

2. 换位密码系统

顾名思义，换位型加密系统重新安排消息中的字符顺序，一个最基本的换位系统可以简单移动（调整）明文的字符位置，使位置 n 映射到位置 n+1，而位置 n+1 又映射为位置 n。这样，消息 KILL ALL THE LAWYERS 就被加密成为 IKLL LAT LEH ALYWRES。在前一个例子中，很明显可以推断出明文的含义。作为明文字符换位的方式，有的加密系统采用一种算法将映射字母或者映射字母序列中的字符移位。换位加密系统的各种情况请见后继章节讨论。

四、电子函件传输问题

上面，我们已经简略地回顾了明文字母表和密文字母表有关的内容，明文字母表代表了用于产生明文消息的字符集中的所有可能的字符（简称明文表），密文字母表代表了在加密过程中采用的所有可能的加密字符（简称密文表）。我们必须注意明文字符集和产生密文字符集的加密过程，以便成功地将被加密的消息通过电子函件传输。事实上，大多数计算机字符集采用 8 位二进制表示一个字符，那么就有 2^8=256 种惟一的字符能够被用来表示明文字符集和密文字符集。幸好，很多电子函件系统把所传输的字符限制为 7 位，而把第 8 位作为传输字节的奇偶校验位。这里，字符传输的限制是 2^7，即 128 种字符。那么，许多不限制密文字符集的密码系统，如 DES 算法，就不能直接地通过某些电子函件系统传输数据。

传输随机产生的密文，应当能以 8 位字符集表示 256 种惟一的字符，并通过 7 位的电子函件系统发送。那么，就需要对每个 8 位字符进行一种变换。最早实现这种变换的计算机程序称为 UUENCODE，它将三个 8 位字符转换成四个 7 位字符。而另一个程序 UUDECODE 则将通过 7 位电子函件传输的每一组的四个 7 位字符再重新转换回原来的三个 8 位字符。虽然我们可以用这两个程序或者类似的软件来通过 7 位的电子函件系统传输 8 位的字符集，但是，所有的接收者也必须具有与编码程序相对应的解码程序。

如果要通过某种电子函件系统来传输密文，而又不需要接收者必须具有 7～8 位字符的转换程序，我们就必须对明文字符集和用来产生密文的算法作出一定的限制。这些限制我们将在后继章节中讨论。

五、例程与程序

为了实际展示加密过程，本书提供了很多例程来实现不同的加密和解密运算。也为了适应各类不同的读者，程序实例采用 C++和 QuickBASIC 两种编程语言加以描述。对于用 C++语言编程的模块，作者采用了 Microsoft Visual C++编译器以产生 C++的可执行程序，采用 Microsoft QuickBASIC 编译器来产生 BASIC 语言的子程序和执行程序。

选择 C++语言作为程序模块编程的原因，首先在于 C++是一种功能非常强大的编程语言，对不同的加密算法提供了频繁的数据处理能力。其次，C++是一种最容易获得的、标准的编译器，由它产生的程序模块可以由读者移植到由其他软件厂商开发的 C++语言编译器上。不过，如果读者没有 C++编译器，书中每一道程序也采用了 QuickBASIC 语言编程。此外，采用这两种语言编程的程序模块，它们的源程序和可执行程序都存放在与本书配套的 CD－ROM 盘上，这就提供了更加灵活的方式。读者可以提取源程序，并将它应用到诸如 Visual C++或者 Visual BASIC 语言编译器上，从而产生与 Windows 兼容的程序。在 MS－DOS 环境下执行书中提供的二进制模块，将对不同的加密技术产生不同的密文。读者还可以修改上述的源程序，以满足自己特殊的需求。

选择 QuickBASIC 语言的原因也首先在于它是每一个 DOS 系统（MS-DOS 或者 PC-DOS）都具有的语言版本，每个读者都可以非常容易地使用书中提供的子程序和源程序，并通过 BASIC 语言解释程序来执行源程序。读者也可以修改这些源程序，选择若干子程序模块，重新构建满足自己特殊需要的加解密程序。其次，所有 DOS 操作系统盘中都包含了 QBASIC 语言解释器，它支持本书中所有的程序和子程序，无需作任何修改。此外，大多数读者都具有 BASIC 语言和编程的知识和基础，能够很好地理论联系实际。QBASIC 编译器能够产生目标代码，允许独立运行，可执行程序也已经存放在 CD-ROM 中。这样，读者可以直接执行程序，无需对源程序进行任何修改，也不需要一定要将源程序调入解释程序中执行。当然，读者不能修改可执行程序，当读者对源程序的某些功能进行改动后，与原来的可执行程序就不一定一样了。

六、文件命名规则

要使用配套的 CD-ROM 盘，需要了解本书中和存放在光盘上程序的命名规则。扩展名为.CPP 的文件代表 C++语言的源文件，在本书中，它们是子程序和程序的清单。扩展名为.BAS 的文件是 BASIC 语言源文件，也是源程序代码清单。扩展名为.DAT 的文件是数据文件，它们包含了在本书中出现的程序中使用的明文和密文消息，它们也存放在 CD-ROM 中，第四类文件是.EXE 文件，它们是可执行文件，读者可以在 DOS 平台或者 Windows 环境下直接执行。

为了使更多的读者使用本书，上述的文件在 CD-ROM 中分别存放在两个目录下，名称为 C 的目录包含了 C++语言的源程序和可执行程序，BASIC 目录下则包含了 BASIC 语言的源程序和可执行程序。虽然很容易地从扩展名来分辨 BASIC 语言和 C++语言的源程序，但是它们的可执行程序都带有同样的扩展名.EXE。因此，必须注意两种语言各自的可执行程序存放的目录。附录 A 给出了存放在 CD-ROM 上文件的清单，包括它们的名称和解释。表 1.6 给出了本书中文件的命名规则。