密码学引论

MI MA XUE YIN LUN

王天芹　编著

河南大学出版社

河南大学教材出版基金资助

MI MA XUE YIN LUN

密码学引论

编著　王天芹

河南大学出版社

中国·开封

图书在版编目(CIP)数据

密码学引论/王天芹编著. —开封:河南大学出版社,2008.9
ISBN 978-7-81091-866-4

Ⅰ.密… Ⅱ.王… Ⅲ.密码—理论—高等学校—教材 Ⅳ.TN918.1

中国版本图书馆 CIP 数据核字(2008)第 140494 号

责任编辑 朱建伟
责任校对 牛 纬
封面设计 马 龙

出版发行 河南大学出版社
地址:河南省开封市明伦街 85 号 邮编:475001
电话:0378-2825001(营销部) 网址:www.hupress.com
排 版 郑州市今日文教印制有限公司
印 刷 河南省诚和印制有限公司
版 次 2008 年 9 月第 1 版 **印 次** 2008 年 9 月第 1 次印刷
开 本 787mm×1092mm 1/16 **印 张** 9.25
字 数 211 千字 **印 数** 1—2000 册
定 价 22.00 元

内容简介

本书是在讲授密码学选修课讲义的基础上编写而成的. 主要内容包括:密码学的基本概念与古典密码,密码学的信息论基础,密码学的数学基础,私钥密码体制及一些有代表性的分组密码算法,公钥密码体制及与公钥有关的若干算法,数字签名,Hash 函数及算法,密钥分配技术和身份识别技术.

本书可作为高等院校计算机、数学等相关专业本科生或研究生的密码学教材或参考书,也可供从事信息安全专业的科技人员参考.

目　录

第1章 引 言

密码学是一门涉及数学、计算机科学等的交叉学科，以研究秘密通信为目的. 近年来，随着计算机和通信网络的广泛应用，信息的安全性问题已成为全社会关注的问题. 因此，作为信息安全基石的密码学，引起了数学和计算机科学工作者日益浓厚的兴趣，其研究十分活跃.

密码学主要由密码编码学和密码分析学两个分支组成. 密码编码学的主要任务是寻求产生安全性高的有效密码算法和密码协议，以满足对信息进行加密或认证的要求；密码分析学的主要任务是破译密码算法和密码协议或伪造认证信息. 两者相互对立，但在发展中又相互促进. 算法是密码学研究的重点内容，包括数据加密算法、数字签名算法、消息摘要算法和密钥管理协议等. 这些算法通过直接对信息进行运算，从而保护信息的安全特性，即通过加密变换保护信息的机密性，通过消息摘要变换检测信息的完整性，通过数字签名保护信息的抗否认性等.

1.1 密码学的产生与发展

密码学的产生可以追溯到古代. 例如，罗马军队通信曾使用恺撒密码. 这一时期，密码的设计没有理论依据，学者们常常凭借直觉和信念进行密码设计和分析. 在 1948 年和 1949 年，Shannon 分别发表了两篇论文《通信的数学原理》和《保密系统的通信理论》，为密码学建立了理论基础. 从此，密码学成为一门科学. 近 30 年来，密码学不仅从理论上得到了快速发展，而且也得到了广泛的社会应用. 密码技术不仅用于网上传送数据的加解密，也用于信息的认证以及各类安全协议（如 IPSec，SSL）和安全应用（如安全电子邮件，安全电子交易）中.

密码学的发展主要经历了四个阶段.

第一阶段：1948 年以前

这一阶段的密码学家往往是凭直觉和信念进行密码设计和分析，可以认为密码学是一门艺术而非科学.

第二阶段：1949～1975 年

这一阶段是对称密码学的早期发展时期. 最有代表性的工作是 Shannon 于 1949 年发表的论文《保密系统的通信理论》，该文为对称密码学建立了理论基础. 从此密码学成为一门科学.

第三阶段：1976～1996 年

这一阶段密码学得到快速发展.其标志性事件:一是 Diffie 与 Hellman 于 1976 年发表的论文《密码编码学的新方向》,该文导致了一场密码学上的革命,他们首次证明了在发送端和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;二是美国于 1977 所制订的数据加密标准 DES.这两件事标志着现代密码学的诞生.

第四阶段:1997 年~

这一阶段是应用密码学的发展时期.密码学得到了广泛的社会应用.

1.2 基本概念

密码学的基本目的是使两个人(通常被称为 Alice 和 Bob)在一个不安全的信道上安全地传输信息,而敌手 Oscar 却无法知道通信的内容.这样不安全的信道在现实生活中是普遍存在的,比如电话线或计算机网络.Alice 发送给 Bob 的信息称为明文(plaintext).Alice 用预先确定好的密钥(key)对明文进行变换,变换的过程称为加密,其逆过程称为解密.对明文进行加密的结果称为密文(ciphertext).对明文进行加密时所采用的一组规则称为加密算法,对密文进行解密时所采用的一组规则称为解密算法.Alice 将密文通过信道发送给 Bob.对 Oscar 而言,他在信道上只能得到 Alice 发给 Bob 的密文却无法知道其对应的明文;但对接收者 Bob 来说,由于他事先知道密钥,可以对密文进行解密,从而获得明文.

这个观点可以用数学方式描述如下:

定义 1.1 一个密码体制是满足以下条件的五元组$(\mathcal{P},\mathcal{C},\mathcal{K},\mathcal{E},\mathcal{D})$:

1. $\mathcal{P}$为明文空间,表示所有可能的明文组成的有限集.
2. $\mathcal{C}$为密文空间,表示所有可能的密文组成的有限集.
3. $\mathcal{K}$为密钥空间,表示所有可能的密钥组成的有限集.其中每一个密钥 $K\in\mathcal{K}$ 均由加密密钥 K_e 和解密密钥 K_d 组成.
4. $\mathcal{E},\mathcal{D}$ 分别为加密算法和解密算法.对于任意的 $K\in\mathcal{K}$,都存在一个加密规则 $e_K\in\mathcal{E}$ 和相应的解密规则 $d_K\in\mathcal{D}$,并且对每个 $e_K:\mathcal{P}\to\mathcal{C}$ 和 $d_K:\mathcal{C}\to\mathcal{P}$,对任意的明文 $x\in\mathcal{P}$,均有 $d_K(e_K(x))=x$.

在定义 1.1 中,性质 4 是最重要的.它说明,如果用 e_K 对明文 x 进行加密,用 d_K 对相的密文进行解密,则得到明文 x.显然,加密函数 e_K 必须是一个单射函数,否则不能正确解密.例如,如果 $y=e_K(x_1)=e_K(x_2)$,这里 $x_1\neq x_2$,则 Bob 就无法判断 y 应解密为 x_1 还是 x_2.注意到,如果 $\mathcal{P}=\mathcal{C}$,则每个加密函数都是一个排列(置换).

Alice 和 Bob 通过某一密码体制进行通信的过程参见图 1.1.

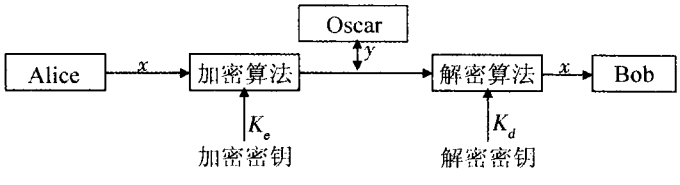

图 1.1 保密通信模型

如果一个密码体制的加密密钥和解密密钥相同，或由其中的一个很容易推导出另一个，则称为私钥密码体制或对称密码体制；否则，称为公钥密码体制或非对称密码体制.

根据对明文的划分和密钥的使用方法不同，可以将私钥密码体制分为分组密码和流密码(又称序列密码)两类. 分组密码将明文 M 划分为一系列称之为分组的明文块 M_1，M_2，…，对每一个 M_i 都用同一个密钥 K_e 进行加密；流密码将明文 M 划分为一系列的字符或位 m_1，m_2，…，对每一个 m_i 用密钥序列的第 i 个分量 K_{e_i} 进行加密.

密码分析是指在不知道密钥的情况下，试图根据已知信息确定明文或密钥的过程. 对密码进行分析的尝试称为攻击. 如果敌手能够攻击成功，则称相应的密码体制是可破译的；否则，称其为不可破译的. 一般情况下，我们都假设敌手 Oscar 知道正在使用的密码体制，即敌手已有密码算法及其实现的详细资料，秘密完全寓于密钥中. 这个假设通常称为 Kerckhoffs 假设，是由荷兰学者 A. Kerckhoffs 于 19 世纪提出的. 当然，如果 Oscar 不知道具体的密码体制，那么完成密码分析将更困难. 但我们不能将密码体制的安全性建立在敌手 Oscar 不知道具体密码体制的假定下. 因此，我们的目标是设计在 Kerchhoffs 假设下安全的密码体制.

敌手的攻击方法主要有三种：

穷举法(又称强力攻击法)：通过对密钥空间进行搜索来破译密码；

统计分析法：通过对明文和密文的统计规律进行分析来破译密码；

数学分析法：针对加密算法的数学依据，通过数学关系式求解未知量来破译密码.

根据敌手可利用的数据来分类，常见的攻击类型有四种：

唯密文攻击(ciphertext only attack)：敌手只有密文串 y；

已知明文攻击(known plaintext attack)：敌手掌握明文串 x 和对应的密文串 y；

选择明文攻击(chosen plaintext attack)：敌手可获得对加密机的临时访问权限，这样他可以选择明文串 x，获得对应的密文串 y；

选择密文攻击(chosen ciphertext attack)：敌手可获得对解密机的临时访问权限，这样他可以选择密文串 y，获得对应的明文串 x.

在每种情况下，敌手的目标都是确定正在使用的密钥或待破译密文所对应的明文. 显然，以上四种类型的攻击强度依次增大. 如果一个密码体制能抵抗选择明文攻击，那么它当然能够抵抗唯密文攻击和已知明文攻击.

密码系统的安全性依赖于破译该系统的困难程度. 实际上，所有实用的密码系统理论上都是可破译的. 如果破译所需的计算能力和时间是现实所不能实现的，则称这样的密码体制是计算上安全的. 本书将要介绍的若干密码破译技术，目的是帮助读者加深对密码算法的理解，关于密码分析的进一步讨论可参阅相应文献.

1.3 古典密码体制及其破译

古典密码体制由基于字符的密码算法构成，采用手工或机械操作实现加/解密. 不同的密码算法是字符之间的相互代替或是互相之间换位，或者是这两种方法的组合. 现代密码体制虽然更为复杂，但基本原理仍然是替代和置换. 因此，回顾和研究这些古典密码体

制，对于理解、设计和分析现代密码具有借鉴价值.

1.3.1 移位密码

本小节介绍移位密码(shift cipher)，其基础是数论中的模运算. 首先给出有关模运算的基本定义.

定义 1.2 假设 a 和 b 均为整数，m 是一正整数. 若 m 整除 $b-a$，则可将其表示为 $a\equiv b(\mathrm{mod}\ m)$. 读作"$a$ 与 b 模 m 同余"，正整数 m 称为模数. 假如用 m 分别去除 a 和 b，可得到相应的商和余数，余数是在 0 与 $m-1$ 之间，即可将 a 与 b 分别表示为

$$a=q_1m+r_1,b=q_2m+r_2,0\leqslant r_1\leqslant m-1,0\leqslant r_2\leqslant m-1.$$

可以看出，$a\equiv b(\mathrm{mod}\ m)$当且仅当 $r_1=r_2$. 上述的余数 r_1 可用记号 $a \bmod m$ 来表示. 因此，$a\equiv b(\mathrm{mod}\ m)$当且仅当 $a \bmod m=b \bmod m$.

令 Z_m 表示集合$\{0,1,\cdots,m-1\}$，在其上定义加法和乘法，类似于普通实数域上的加法和乘法，所不同的只是所得的值是取模以后的余数. 例如，在 Z_{16} 上计算 11×13. 因为 $11\times13=143=8\times16+15$，故在 Z_{16} 上，$11\times13=15$.

在 Z_m 中减去一个元素，定义 $a-b$ 为：$(a+m-b) \bmod m$.

定义 1.3 移位密码.

令 $\mathcal{P}=\mathcal{C}=\mathcal{K}=Z_{26}$，对 $0\leqslant K\leqslant25$，定义

$$e_K(x)=(x+K) \bmod 26,$$
$$d_K(y)=(y-K) \bmod 26,$$

其中 $x,y\in Z_{26}$.

因为英文有 26 个字母，故一般定义在 Z_{26} 上. 很容易验证移位密码满足前面所定义的密码体制的条件，即对任意的 $x\in Z_{26}$，有 $d_K(e_K(x))=x$.

注：若取 $K=3$，则此密码体制通常称为恺撒密码(Caesar cipher)，它是已知最早的替代密码.

使用移位密码可以加密普通的英文句子. 下面给出一个实例.

例 1.1 假设移位密码的密钥为 $K=11$，明文为：

wewillmeetatmidnight

首先，将明文中的字母对应于相应的整数，得到如下的数字串：

22	4	22	8	11	11	12	4	4	19
0	19	12	8	3	13	8	6	7	19

然后，将每一个数都与 11 相加后取模 26，可得：

7	15	7	19	22	22	23	15	15	4
11	4	23	19	14	24	19	17	18	4

最后，再将其转换为相应的字母串，即得密文：

HPHTWWXPPELEXTOYTRSE

要对密文进行解密，只需要执行相应的逆过程即可. 即首先将密文转换为数字，用每个数减去 11 后取模 26 运算，最后将相应数字再转换为字母可得明文.

注:以上例子中,我们使用小写字母来表示明文,而使用大写字母来表示密文.为讨论方便,后面仍然使用这种约定.

移位密码(模26)是不安全的,因为密钥空间太小,只有26个可能的密钥,很容易通过穷举所有可能密钥得到有意义的明文.我们给出一个例子.

例1.2 设有如下密文串:

JBCRCBQRWCRVNBJENBWRWN

依次试验所有可能的解密密钥 $d_0, d_1, \cdots$,得到不同字母串:

jbcrclqrwcrvnbjenbwrwn
iabqbkpqvbqumaidmavqvm
hzapajopuaptlzhclzupul
gyzozinotzoskygbkytotk
fxynyhmnsynrjxfajxsnsj
ewxmxglmrxmqiweziwrmri
dvwlwfklqwlphvdyhvqlqh
cuvkvejkpvkogucxgupkpg
btujudijoujnftbwftojof
astitchintimesavesnine

至此,已可以得出有意义的明文,相应的密钥 $K=9$.(明文串为"a stitch in time saves nine",中文意思是"小洞不补,大洞吃苦")

使用上述方法计算明文平均只需试验 $26/2=13$ 次即可.

上面的例子表明,一个密码体制安全的必要条件是能够抵抗穷尽密钥搜索攻击,也就是说,密钥空间必须足够大.(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)

1.3.2 维吉尼亚密码

在移位密码中,一旦密钥被选定,则每个明文字符对应唯一的密文字符,这种密码体制一般被称为单表代换密码.下面我们将介绍一个非单表代换密码体制——维吉尼亚密码,它是16世纪的法国人 Blaise de Vigenere 发明的.

定义1.4 维吉尼亚密码.

设 m 是一个选定的正整数.定义 $\mathcal{P}=\mathcal{C}=\mathcal{K}=(Z_{26})^m$.对任意的密钥 $K=(k_1,k_2,\cdots,k_m)$ 定义:

$$e_K(x_1,x_2,\cdots,x_m)=(x_1+k_1,x_2+k_2,\cdots,x_m+k_m),$$
$$d_K(y_1,y_2,\cdots,y_m)=(y_1-k_1,y_2-k_2,\cdots,y_m-k_m).$$

以上运算都是在 Z_{26} 上进行的.

下面是一个小例子.

例1.3 假设 $m=6$,密钥字为"CIPHER",其对应于数字串 $K=(2,8,15,7,4,17)$.假设明文为:

thiscryptosystemisnotsecure

将明文串转化为对应的数字，每 6 个为一组，使用密钥字进行模 26 加法运算，如下所示：

19	7	8	18	2	17	24	15	19	14	18	24	18	19
2	8	15	7	4	17	2	8	15	7	4	17	2	8
21	15	23	25	6	8	0	23	8	21	22	15	20	1
4	12	8	18	13	14	19	18	4	2	20	17	4	
15	7	4	17	2	8	15	7	4	17	2	8	15	
19	19	12	9	15	22	8	25	8	19	22	25	19	

相应的密文为：

VPXZGIAXIIVWPUBTTMJPWIZTWZT

解密时，我们使用相同的密钥字，但进行的是模 26 减运算，这里不再给出.

可以看出，维吉尼亚密码的密钥空间大小为 26^m，所以，即使 m 的值很小，使用穷尽密钥搜索方法也需要很长的时间.

在一个密钥字长度为 m 的维吉尼亚密码中，一个字母可以被映射为 m 个字母中的某一个，称其为多表代换密码体制. 一般来说，多表代换密码比单表代换更为安全.

1.3.3 希尔密码

前面介绍的密码都是以单个字母作为代换对象的. 如果每次对多个字母进行代换就是多字母代换密码. 本小节介绍一种多字母代换密码——希尔密码(Hill cipher). 这种密码体制是 Lester S. Hill 在 1929 年提出的. 设 m 是一正整数，定义 $\mathcal{P}=\mathcal{C}=(Z_{26})^m$. 希尔密码的主要思想是取 m 个明文符号的 m 个线性组合，得到 m 个密文符号. 不同的是，这种变换是在 Z_{26} 上进行的.

例如，设 $m=2$，每一个明文元素使用 $x=(x_1,x_2)$ 来表示，同样，密文元素使用 $y=(y_1,y_2)$ 来表示，则 y_1,y_2 将被表示为 x_1,x_2 的线性组合. 可以取

$$y_1=11x_1+3x_2,$$
$$y_2=8x_1+7x_2.$$

使用矩阵，可将上式简写为：

$$(y_1,y_2)=(x_1,x_2)\begin{pmatrix}11 & 8\\ 3 & 7\end{pmatrix}.$$

密钥 K 一般取为一个 $m\times m$ 矩阵，记为 $K=(k_{i,j})$. 对明文 $x=(x_1,x_2,\cdots,x_m)\in\mathcal{P}$ 以及 $K\in\mathcal{K}$，按照如下方法来计算 $y=e_K(x)=(y_1,y_2,\cdots,y_m)$：

$$(y_1,y_2,\cdots,y_m)=(x_1,x_2,\cdots,x_m)\begin{pmatrix}k_{1,1} & k_{1,2} & \cdots & k_{1,m}\\ k_{2,1} & k_{2,2} & \cdots & k_{2,m}\\ \cdots & \cdots & \cdots & \cdots\\ k_{m,1} & k_{m,2} & \cdots & k_{m,m}\end{pmatrix},$$

也可直接表示为 $y=xK$.

从上面的加密变换可以看出，密文是通过对明文进行线性变换得出的. 解密变换使用

K 的逆矩阵 K^{-1} 来进行，相应的明文应该为 $x=yK^{-1}$.

下面给出一个希尔密码的具体应用例子.

例 1.4 假设密钥为：

$$K=\begin{pmatrix}11 & 8\\ 3 & 7\end{pmatrix}.$$

其逆矩阵为：

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

设要加密的明文为“july”，可将明文转化为如下两个加密单元：(9,20)(对应于“ju”)和(11,24)(对应于“ly”)，分别对其进行加密变换如下：

$$(9,20)\begin{pmatrix}11 & 8\\ 3 & 7\end{pmatrix}=(99+60,72+140)=(3,4),$$

$$(11,24)\begin{pmatrix}11 & 8\\ 3 & 7\end{pmatrix}=(121+72,88+168)=(11,22).$$

因此，密文为 DELW. 要解密密文，作如下的计算：

$$(3,4)\begin{pmatrix}7 & 18\\ 23 & 11\end{pmatrix}=(9,20),(11,22)\begin{pmatrix}7 & 18\\ 23 & 11\end{pmatrix}=(11,24).$$

这样即可得到正确的明文.

由前面的分析可以看出，如果密钥 K 可逆，则存在解密算法. 事实上，K 可逆是完成解密的必要条件.(此结论来自基本的线性代数知识，在此我们不给出证明)因此，我们只对可逆矩阵 K 感兴趣.

下面我们给出 Z_{26} 上希尔密码的具体描述.

定义 1.5 希尔密码.

设 $m\geqslant 2$ 为取定的正整数，$\mathscr{P}=\mathscr{C}=(Z_{26})^m$，$\mathscr{K}=\{$定义在 Z_{26} 上的 $m\times m$ 可逆矩阵$\}$. 对任意的密钥 K，定义

$$e_K(x)=xK, d_K(y)=yK^{-1}.$$

以上运算都是在 Z_{26} 上进行的.

1.3.4 置换密码

前面讨论的密码体制都是替代密码，即明文字母被不同的字母代替. 置换密码(permutation cipher)的思想是保持明文的所有字母不变，只是改变明文字母的位置.

定义 1.6 置换密码.

令 m 为一正整数. $\mathscr{P}=\mathscr{C}=(Z_{26})^m$，$\mathscr{K}$ 由所有定义在集合$\{1,2,\cdots,m\}$上的置换组成. 对任意的密钥(置换)π，定义

$$e_\pi(x_1,x_2,\cdots,x_m)=(x_{\pi(1)},x_{\pi(2)},\cdots,x_{\pi(m)}),$$

$$d_\pi(y_1,y_2,\cdots,y_m)=(y_{\pi^{-1}(1)},y_{\pi^{-1}(2)},\cdots,y_{\pi^{-1}(m)}).$$

这里，π^{-1} 为置换 π 的逆置换.

正如在代换密码中所讨论的一样，为方便起见，我们可以认为 $\mathscr{P}$ 和 $\mathscr{C}$ 定义在 26 个英

文字母表上.

下面给出一个具体的例子.

例 1.5 设 $m=6$,密钥为如下的置换 π:

1	2	3	4	5	6
3	5	1	6	4	2

注意,上表的第一行是关于 $x(1\leqslant x\leqslant 6)$ 值的列表,第二行是其相应的置换 $\pi(x)$. 逆置换 π^{-1} 如下:

1	2	3	4	5	6
3	6	1	5	2	4

假设给出的明文是:

shesellsseashellsbytheseashore

首先,将明文字母分为每 6 个一组:

shesel | lsseas | hellsb | ythese | ashore

对每组的 6 个字母使用加密变换 π,得:

EESLSH | SALSES | LSHBLE | HSYEET | HRAEOS

因此,最后得到的密文为:

EESLSH SALSES LSHBLE HSYEET HRAEOS

解密过程和加密过程一样,只不过使用的是逆置换 π^{-1},这里不再给出具体过程.

事实上,置换密码是希尔密码的特例. 给定集合 $\{1,2,\cdots,m\}$ 的一个置换 π,可按如下方法定义一个与 π 关联的 $m\times m$ 置换矩阵 $K_\pi=(k_{i,j})$:

$$k_{i,j}=\begin{cases}1, & \text{若 } i=\pi(j);\\ 0, & \text{其他情形.}\end{cases}$$

(一个置换矩阵是指每行每列都恰好只有一个 1,其他都是 0 的矩阵. 一个置换矩阵可以通过对单位矩阵进行行置换和列置换而得到)

容易看出使用矩阵 K_π 为密钥的希尔密码事实上等价于使用密钥 π 进行加密的置换密码,并且还有 $K_\pi^{-1}=K_{\pi^{-1}}$,即 K_π 的逆矩阵是 π^{-1} 关联的置换矩阵. 这说明二者的解密变换也是等价的.

对于前面例子中的置换 π,其关联置换矩阵为:

$$K_\pi=\begin{pmatrix}0&0&1&0&0&0\\0&0&0&0&0&1\\1&0&0&0&0&0\\0&0&0&0&1&0\\0&1&0&0&0&0\\0&0&0&1&0&0\end{pmatrix},$$

并且有:

$$K_{\pi^{-1}}=\begin{pmatrix}0&0&1&0&0&0\\0&0&0&0&1&0\\1&0&0&0&0&0\\0&0&0&0&0&1\\0&0&0&1&0&0\\0&1&0&0&0&0\end{pmatrix}.$$

很容易验证以上两矩阵的乘积恰为单位矩阵.

1.3.5 古典密码体制分析

许多密码分析方法都利用了英文语言的统计特性.目前已经有许多从各种小说、杂志和报纸上统计的26个英文字母出现的频率.表1.1的统计数据由Beker和Piper给出.

表1.1 26个英文字母出现的概率

字母	概率	字母	概率
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

另外,考虑两字母组或者三字母组也是很有用的.30个最常见的两字母组(按出现次数递减排序)为:TH, HE, IN, ER, AN, RE, DE, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, TI, IS, ET, IT, AR, TE, SE, HI, OF.12个最常见的三字母组(按出现次数递减排序)为:THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR,DTH.

简单的单表代换密码,如移位密码极易破译.仅统计得出密文中最高频率字母再与上述统计表中字母对应即可确定位移量.另外,由于密钥量很小,也很容易用穷举密钥搜索来破译.

采用唯密文攻击法分析希尔密码是很难的,但是如果采用已知明文攻击,则很容易破译希尔密码.假定敌手已经确定 m 的值,并且至少有 m 个不同的明—密文对:

$$x_j=(x_{1,j},x_{2,j},\cdots,x_{m,j}),y_j=(y_{1,j},y_{2,j},\cdots,y_{m,j}),$$

其中 $y_j=e_K(x_j),1\leqslant j\leqslant m$.如果我们定义两个 $m\times m$ 矩阵 $X=(x_{i,j})$ 和 $Y=(y_{i,j})$,则有矩阵方程 $Y=XK$,其中 $m\times m$ 矩阵 K 是未知密钥.假如矩阵 X 刚好是可逆的,则敌手Oscar可计算出 $K=X^{-1}Y$,从而破译希尔密码.(如果 X 不可逆,则必须重新选择 m 个明—密文

对)

下面给出一个例子.

例 1.6 假设明文"friday"利用 $m=2$ 的希尔密码加密,得到密文为"PQCFKU". 我们有

$$e_K(5,17)=(15,16),e_K(8,3)=(2,5),e_K(0,24)=(10,20),$$

使用前两个明—密文对,可得矩阵方程:

$$\begin{pmatrix}15 & 16\\ 2 & 5\end{pmatrix}=\begin{pmatrix}5 & 17\\ 8 & 3\end{pmatrix}K.$$

容易计算

$$\begin{pmatrix}5 & 17\\ 8 & 3\end{pmatrix}^{-1}=\begin{pmatrix}9 & 1\\ 2 & 15\end{pmatrix}.$$

因此

$$K=\begin{pmatrix}9 & 1\\ 2 & 15\end{pmatrix}\begin{pmatrix}15 & 16\\ 2 & 5\end{pmatrix}=\begin{pmatrix}7 & 19\\ 8 & 3\end{pmatrix}.$$

这个结果可以使用第三个明—密文对进行验证.

如果敌手不知道 m 的值,该如何攻击呢? 假定 m 不是太大,可以尝试 $m=2,3,\cdots$,直到发现密钥为止. 如果猜测的 m 值不正确,则可利用其余的明—密文对进行验证查出. 运用这种方法,即使在不知道 m 值的情况下也可以破译希尔密码.

多表代换密码的破译要比单表代换密码的破译困难得多. 下面给出分析维吉尼亚密码的一些方法. 首先必须确定密钥字的长度 m,这里我们介绍两种方法:Kasiski 测试法和重合指数法.

Kasiski 测试法由 Friedrich Kasiski 在 1863 年给出. 它是基于这样一个事实:两个相同的明文段,如果它们的位置间距为 x,其中 $x\equiv 0 \bmod m$,则将加密成相同的密文段. 反过来,如果在密文中观察到两个相同的长度至少为 3 的密文段,则它们很可能对应了相同的明文段.

Kasiski 测试过程如下:搜索长度至少为 3 的相同密文段,记录这些相同密文段起始点之间的距离,假如得到如下几个距离 $d_1,d_2,\cdots$,那么可以猜则 m 为这些 d_i 的最大公因数的因子.

求 m 值进一步的方法是使用所谓的重合指数法,这一概念由 Wolfe Friedman 在 1920 年提出. 其定义如下:

定义 1.7 设 $x=x_1x_2\cdots x_n$ 是一个含有 n 个字符的字符串,x 的重合指数记为$I_c(x)$,定义为 x 中两个随机元素相同的概率.

假设 $f_0,f_1,\cdots,f_{25}$分别表示 A,B,…,Z 在 x 中出现的频率,共有 $\binom{n}{2}$种方法来选择 x 中任意两个元素.(二项式系数 $\binom{n}{k}=\frac{n!}{k!\ (n-k)!}$表示从 n 个元素中取出 k 个元素的方式的个数)对每一个 $0\leqslant i\leqslant 25$,共有 $\binom{f_i}{2}$种方法使得所选的两个元素都为 i. 因此有如下公

式：

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

假设 x 是英语文本串，记表 1.1 中字母 A,B,…,Z 出现的概率分别为 $p_0,p_1,\cdots,p_{25}$. 我们期望：

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

上式成立主要是因为两个随机元素都是 A 的概率为 p_0^2，两个随机元素都为 B 的概率为 p_1^2，等等. 如果 x 是通过单表变换密码而得来的，则各个概率将被置换，但量 $\sum_{i=0}^{25}p_i^2$ 将不会改变.

假设我们使用维吉尼亚密码加密的密文串为 $y=y_1y_2\cdots y_n$. 将串 y 分割为 m 个长度相等的子串，分别为 $Y_1,Y_2,\cdots,Y_m$，分割方式为：以列的形式写出密文，组成一个 $m\times(n/m)$ 矩阵，矩阵的每一行对应于子串 Y_i，$1\leqslant i\leqslant m$. 如果 m 是正确的密钥长度，则每一个 $I_c(Y_i)$ 大约等于 0.065. 反过来，如果 m 不是密钥长度，那么子串 Y_i 更像一随机串，因为它们是用不同的密钥通过移位密码加密来得到的. 注意到，对于一个完全随机串，有

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

0.065 和 0.038 这两个值相差充分大，可以用来确定正确的密钥长度.（或者说，进一步确认用 Kasiski 测试所猜测的密钥长度）

这里我们用一个例子来说明这两种方法.

例 1.7 已知使用维吉尼亚密码加密获得如下密文：

CHREEVO AHMAERATBI AXX WTNXBEEOPH BSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUT EMND CMGTSX MXBTUI AD NGMGP SRELX NJ ELX
VRVPRTUL HDNQ WTWDT YGBPHXTFALJHASSV BFXNG LLCHR
ZBWEL EKMSJ JKNBH W RJGNMG JSGL XFE YP HAG NRB IEQ JT
AM RVLCRREM NDG LXRR IMG NSN RWCHRQHAEYEVTAQEBBI
PEEWE VKAKOE WADR EMXM TBJJ CHR TKDN VRZCHRCLQOHP
WQAIIWXNRMGWOIIFKEE

首先使用 Kasiski 测试法. 密文串 CHR 共出现在 5 个位置，开始位置分别为 1,166,236,276 和 286，其距离分别为 165,235,275 和 285. 这四个整数的最大公约数为 5，故我们猜测密钥字的长度很可能为 5.

我们再使用重合指数法确认这一猜测. 当 $m=1$ 时，重合指数为 0.045；当 $m=2$ 时，两个重合指数分别为 0.046 和 0.041；当 $m=3$ 时，分别为 0.043,0.050 和 0.047；当 $m=4$ 时，分别为 0.042，0.039，0.046 和 0.040；当 $m=5$ 时，分别为 0.063,0.068,0.069,0.061 和 0.072. 这些值为密钥字的长度为 5 提供了强有力的证据.

假设密钥字长度 m 已知，那么如何确定密钥字呢？通常采用重合互指数法.