密码学引论

[荷兰] H.C.A.范蒂尔博格 著

四川省电子学会·机电部三十所

密码学引论

〔荷兰〕H.C.A.范蒂尔博格 著
吴兴龙 宋云生 译
龚奇敏 刘向武 校

四川省电子学会、机电部三十所

1990年·成都

An Introduction to Cryptology

Henk C.A. van Tilborg

Kluwer Academic Publishers, 1988

密码学引论

[荷兰]H.C.A.范蒂尔博格 著

吴兴龙 宋云生 译
龚奇敏 刘向武 校

四川省电子学会、机电部三十所出版
(成都市.810 信箱)
四川智力开发研究所电脑照排部排版、文献情报开发部印刷

机电部三十所情报室

译者序

近二十年来，密码学获得了空前的普及和发展。究其原因，表面上直接受了两件事的影响：其一是美国在 70 年代中公开征集并发布实用的密码算法作为数据加密标准(DES)，其二是美国学者 W.迪菲和 M.E.赫尔曼在他们的《密码学的新方向》一文中首次提出了公开密钥密码体制的新思想。但更确切地说，是信息时代的需要。在信息社会里，人类生活的所有领域每日每时都产生着大量信息，其中最重要或机密的信息在存贮和交换中无疑都需要保护，这种保护既包括传统的抗窃听功能，又增加了现代的抗欺骗及窜改等功能。人们认识到，加密处理是实现这类保护的理想途径，因此密码学冲破了原来狭隘的军事、外交圈子，迅速走向全社会，它吸引着政府部门，民间团体及学术界众多的专家、教授、工程师及其他业余爱好者，从而在 70 年代中出现了直接导致密码学空前繁荣的上述两件事。

今天，密码学的繁荣势头有增无减。国际上，自 80 年代初以来，在欧洲和美国分别有专门的世界性密码年会，除此之外，在数学、计算机科学、信息论及其他电子工程等相关学科的学术会议中，也包含有关密码学的专题讨论及优秀文章。学者和专家们发表了数以千计的论文，还写了十多部密码学及应用专著，其中不乏优秀作品，如：Beker，H 和 F.Piper 的《密码体制，通信的保护》(1982 年)，Denning，D.E.R 的《密码学和数据安全》(1982 年)，Meyer.C.H 和 S.M.Mathyas 的《密码学：计算机安全方面的新领域》(1982 年)，R.A.Rueppel 的《序列密码的分析与设计》(1986 年) 等。1984 年初，国际标准化组织 (ISO) 还专门成立了数据加密分技术委员会，每年举行若干次工作组会议和年会，与其他分技术委员会合作讨论和制订除了密码算法以外的所有加密技术的国际标准。

在我国，随着社会经济的发展，人们对密码学的需求正与日俱增。广大专业人员及学校师生都希望有一本密码学基础方面的教科书。我们认为 H.C.A.范蒂尔博格著的《密码学引论》(An Introduction to Cryptology) 这本书就是密码学入门方面的好书。在该书中，作者对传统密码体制特别是对公开密钥体制及与其有关的理论问题都作了深入浅出的介绍，书末附有初等数论和有限域理论的知识，大部分章节都给了一些习题。该书不仅将读者引入密码学的大门，而且为读者指出了进一步学习和研究的方向。由于我们水平有限，译校错误难免，请批评指正。

本书出版过程中，得到赵芝、苏曼波、周恺、陈琴等同志大力帮助，深表感谢。

龚奇敏

1990 年 4 月

目 次

第一章 引 言

密码术（cryptology）是一门研究密码体制的科学。它可以分为密码编码学（cryptography）和密码分析学（cryptanalysis）两个学科。密码编码学涉及密码体制本身的设计，而密码分析学则研究密码体制的破译。此时，我们还不打算给一种密码体制下正式的定义，有关定义问题将在本章的后面论述。先假设读者对什么是密码体制已有一个正确的直观概念。

为什么大家都想使用密码体制？这可能有两个理由：

保密：当传送或存储数据时，人们都不想让第三者理解传送或存储信息的内容。

鉴别：此特性是签名的。报文接收者想证明报文确是来自某人，而不是来自其它人（即使报文的原发者事后想否认也不行）。

几个世纪以来（参见文献〔Kah67〕），密码体制一直是由军方和外交部门使用的。如今，由于计算机控制的通信系统在工业和民用部门的广泛应用，人们常常要求用密码体制对数据进行特殊的保护。

由于可以将存储数据看成是数据在时间域内的传送，因此在讨论数据存储和（或）数据传送时，我们将始终使用传送这一术语。在我们能描述由香农（C.E.Shannon）概括的传统密码体制〔Shn49〕之前，需要对语言和文本之类的概念给予更形式地叙述。

令 A 为一个有限集合，我们称它为"字母表"。$|A|$将表示 A 的基数。我们常常将$Z_q=\{0, 1, \cdots, q-1\}$ 用作为字母表，对字母表中的元素运算按模 q 进行（参见§A.3开头和§B.2）。用 $\{a, b, \cdots z\}$ 来表示 Z_{26}。在当代实际情况中，q 常常为2或2的幂。Z_q 中 n 个字母的连结将称为一个 $n-gram$，并用 $\underline{a}=(a_0, a_1, \cdots a_{n-1})$ 表示。Z_q 中所有 $n-gram$ 的集合将用 Z_q^n 来表示。

由 Z_q 中若干个字母组成的报文是 $Z_q^*:=\bigcup_{n\geq 0} Z_q^n$ 中的一个元素。一种"语言"便是 Z_q^* 的一个子集。在程序语言中，这一子集是由递归规则严格定义的。然而在口语情况下，这些规则就不十分严格了。这里，我们选择一种概率统计方法。

由 Z_q 中 n 个字母组成的报文称为"明文"。它的有限或无限"明文源 S"分别是随机变量的一个有限和无限序列，即

$$(M_0, M_1, \cdots M_{r-1})$$

和

$$(M_0, M_1, M_2, \cdots),$$

它们是由事件出现的概率描述的。因此，对所有可能的事件

$$(m_0, m_1, \cdots, m_{n-1}) \in Z_q^n,$$

$$Pr_{plain}\{Mj=m_0, M_{j+1}=m_1, \cdots\cdots M_{j+n-1}=m_{n-1}\}, j>0.$$

在 $j=0$ 的情况下，我们只需要写 $Pr_{plain}\{(m_0, m_1, \cdots\cdots m_{n-1})\}$。当然，这些概率必须满足某些明显的关系式：

ⅰ 对所有的$(m_0,m_1,\cdots\cdots m_{n-1})$，$Pr_{plain}\{(m_0,m_1,\cdots\cdots m_{n-1})\}>0$.

ⅱ $\sum_{(m_0,m_1\cdots m_{n-1})} Pr_{plain}\{(m_0,m_1\cdots m_{n-1})\}=1$

ⅲ Kolmogorov 的相容条件:

$$\sum_{m_n,m_{n+1}\cdots m_{l-1}} Pr_{plain}\{(m_0,m_1,\cdots,m_{l-1})\}=Pr_{plain}\{(m_0,m_1,\cdots,m_{n-1})\},\ l>n$$

例 1.1: S 产生独立、同分布的 1-grams，即 $P(t)$，$0<t<q$. 因此，

$$Pr_{plain}\{(m_0,m_1,\cdots,m_{n-1})\}=P(m_0)P(m_1)\cdots P(m_{n-1}),n>1.$$

表 1.1 给出英文报文中字母的分布情况（参见文献〔Mey82〕，表 12-1）. 在这一模型中，我们有

$$Pr_{plain}\{(run)\}=Pr_{plain}\{(urn)\}=P(r)P(u)P(n)=0.0612\times 0.0271\times 0.0709=1.18*10^{-4}$$

a	0.0804	h	0.0549	o	0.0760	u	0.0271
b	0.0154	i	0.0726	p	0.0200	v	0.0099
c	0.0306	j	0.0016	q	0.0011	w	0.0192
d	0.0399	k	0.0067	r	0.0612	x	0.0019
e	0.1251	l	0.0414	s	0.0654	y	0.0173
f	0.0230	m	0.0253	t	0.0925	z	0.0009
g	0.0196	n	0.0709				

表 1.1 英语中 1-grams 的概率分布

it	a	b	c	d	e	f	g	h	i	j	k	l	m
a	0.0011	0.0193	0.0388	0.0469	0.0020	0.0100	0.0233	0.0020	0.0480	0.0020	0.0103	0.1052	0.0281
b	0.0931	0.0057	0.0016	0.0008	0.3219	0.0000	0.0000	0.0000	0.0605	0.0057	0.0000	0.1242	0.0049
c	0.1202	0.0000	0.0196	0.0004	0.1707	0.0000	0.0000	0.1277	0.0761	0.0000	0.0324	0.0369	0.0015
d	0.1044	0.0020	0.0026	0.0218	0.3778	0.0007	0.0132	0.0007	0.1803	0.0033	0.0000	0.0125	0.0178
e	0.0660	0.0036	0.0433	0.1194	0.0438	0.0142	0.0125	0.0021	0.0158	0.0005	0.0036	0.0456	0.0340
f	0.0838	0.0000	0.0000	0.0000	0.1283	0.0924	0.0000	0.0000	0.1608	0.0000	0.0000	0.0299	0.0009
g	0.1078	0.0000	0.0000	0.0018	0.2394	0.0000	0.0177	0.1281	0.0839	0.0000	0.0000	0.0203	0.0027
h	0.1769	0.0005	0.0014	0.0008	0.5623	0.0000	0.0000	0.0005	0.1167	0.0000	0.0000	0.0016	0.0016
i	0.0380	0.0082	0.0767	0.0459	0.0437	0.0129	0.0280	0.0002	0.0016	0.0000	0.0050	0.0567	0.0297
j	0.1259	0.0000	0.0000	0.0000	0.1818	0.0000	0.0000	0.0000	0.0350	0.0000	0.0000	0.0000	0.0000
k	0.0395	0.0028	0.0000	0.0028	0.5282	0.0028	0.0000	0.0198	0.1582	0.0000	0.0113	0.0198	0.0028
l	0.1342	0.0019	0.0022	0.0736	0.1918	0.0105	0.0108	0.0000	0.1521	0.0000	0.0079	0.1413	0.0082
m	0.1822	0.0337	0.0026	0.0000	0.2975	0.0010	0.0000	0.0000	0.1345	0.0000	0.0000	0.0010	0.0654
n	0.0550	0.0004	0.0621	0.1681	0.1212	0.0102	0.1391	0.0013	0.0665	0.0009	0.0066	0.0073	0.0104
o	0.0085	0.0101	0.0162	0.0231	0.0037	0.1299	0.0082	0.0025	0.0092	0.0014	0.0078	0.0416	0.0706
p	0.1359	0.0000	0.0006	0.0000	0.1747	0.0000	0.0000	0.0237	0.0423	0.0000	0.0000	0.0812	0.0073
q	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
r	0.1026	0.0033	0.0172	0.0282	0.2795	0.0031	0.0175	0.0017	0.1181	0.0000	0.0205	0.0164	0.0303
s	0.0604	0.0012	0.0284	0.0027	0.1795	0.0024	0.0000	0.0561	0.1177	0.0000	0.0091	0.0145	0.0112
t	0.0619	0.0003	0.0036	0.0002	0.1417	0.0007	0.0002	0.3512	0.1406	0.0000	0.0000	0.0101	0.0044
u	0.0344	0.0415	0.0491	0.0243	0.0434	0.0052	0.0382	0.0010	0.0258	0.0000	0.0014	0.1097	0.0329
v	0.0749	0.0000	0.0000	0.0023	0.6014	0.0000	0.0000	0.0000	0.2569	0.0000	0.0000	0.0000	0.0012
w	0.2291	0.0008	0.0000	0.0032	0.1942	0.0000	0.0000	0.1422	0.2104	0.0000	0.0000	0.0041	0.0000
x	0.0672	0.0000	0.1119	0.0000	0.1269	0.0000	0.0000	0.0075	0.1119	0.0000	0.0000	0.0000	0.0075
y	0.0586	0.0034	0.0103	0.0069	0.2897	0.0000	0.0000	0.0000	0.0690	0.0000	0.0034	0.0172	0.0379
z	0.2278	0.0000	0.0000	0.0000	0.4557	0.0000	0.0000	0.0000	0.2152	0.0000	0.0000	0.0127	0.0000

it	n	o	p	q	r	s	t	u	v	w	x	y	z
a	0.1878	0.0008	0.0222	0.0000	0.1180	0.1001	0.1574	0.0137	0.0212	0.0057	0.0026	0.0312	0.0023
b	0.0000	0.0964	0.0000	0.0000	0.0662	0.0229	0.0049	0.0727	0.0016	0.0000	0.0000	0.1168	0.0000
c	0.0011	0.2283	0.0000	0.0004	0.0426	0.0087	0.0893	0.0347	0.0000	0.0000	0.0000	0.0004	0.0000
d	0.0053	0.0733	0.0000	0.0007	0.0324	0.0495	0.0013	0.0601	0.0099	0.0040	0.0000	0.0264	0.0000
e	0.1381	0.0040	0.0192	0.0034	0.1927	0.1231	0.0404	0.0048	0.0215	0.0205	0.0152	0.0121	0.0004
f	0.0009	0.2789	0.0000	0.0000	0.1215	0.0026	0.0496	0.0462	0.0000	0.0000	0.0000	0.0043	0.0000
g	0.0451	0.1140	0.0000	0.0000	0.1325	0.0256	0.0247	0.0512	0.0000	0.0000	0.0000	0.0053	0.0000
h	0.0038	0.0786	0.0000	0.0000	0.0153	0.0027	0.0233	0.0085	0.0000	0.0011	0.0000	0.0041	0.0000
i	0.2498	0.0893	0.0100	0.0008	0.0342	0.1194	0.1135	0.0011	0.0250	0.0000	0.0023	0.0002	0.0079
j	0.0000	0.3147	0.0000	0.0000	0.0070	0.0000	0.0000	0.3357	0.0000	0.0000	0.0000	0.0000	0.0000
k	0.0565	0.0198	0.0000	0.0000	0.0085	0.1102	0.0028	0.0028	0.0000	0.0000	0.0000	0.0113	0.0000
l	0.0004	0.0778	0.0041	0.0000	0.0034	0.0389	0.0254	0.0269	0.0056	0.0011	0.0000	0.0819	0.0000
m	0.0042	0.1246	0.0722	0.0000	0.0026	0.0244	0.0005	0.0337	0.0005	0.0000	0.0000	0.0192	0.0000
n	0.0194	0.0528	0.0004	0.0007	0.0011	0.0751	0.1641	0.0124	0.0068	0.0018	0.0002	0.0157	0.0004
o	0.2190	0.0222	0.0292	0.0000	0.1530	0.0357	0.0396	0.0947	0.0334	0.0345	0.0012	0.0041	0.0004
p	0.0006	0.1511	0.0581	0.0000	0.2306	0.0180	0.0287	0.0457	0.0000	0.0000	0.0000	0.0017	0.0000
q	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
r	0.0325	0.1114	0.0055	0.0000	0.0212	0.0655	0.0596	0.0192	0.0142	0.0017	0.0002	0.0306	0.0000
s	0.0021	0.0706	0.0386	0.0009	0.0027	0.0836	0.2483	0.0579	0.0000	0.0039	0.0000	0.0081	0.0000
t	0.0015	0.1229	0.0003	0.0000	0.0479	0.0418	0.0213	0.0195	0.0005	0.0088	0.0000	0.0203	0.0005
u	0.1517	0.0019	0.0386	0.0000	0.1460	0.1221	0.1255	0.0029	0.0014	0.0000	0.0010	0.0014	0.0005
v	0.0000	0.0530	0.0000	0.0000	0.0000	0.0023	0.0000	0.0012	0.0012	0.0000	0.0000	0.0058	0.0000
w	0.0357	0.1292	0.0000	0.0000	0.0106	0.0366	0.0016	0.0000	0.0000	0.0000	0.0000	0.0024	0.0000
x	0.0000	0.0075	0.3507	0.0000	0.0000	0.0000	0.1716	0.0000	0.0000	0.0000	0.0373	0.0000	0.0000
y	0.0172	0.2207	0.0310	0.0000	0.0310	0.1517	0.0172	0.0138	0.0000	0.0103	0.0000	0.0069	0.0034
z	0.0000	0.0506	0.0000	0.0000	0.0000	0.0000	0.0000	0.0127	0.0000	0.0000	0.0000	0.0000	0.0253

表 *1.2* 英语的转移概率 p $(it|s)$

例 1.2：S 产生独立、同分布的 2-grams，即 $P(s, t)$，$0<s, t<q$。因此，

$$Pr_{plain}\{(m_0,m_1,\cdots m_{2n-1})\}=P(m_0,m_1)P(m_2,m_3)\cdots P(m_{2n-2},m_{2n-1}) \quad n>1$$

英文报文中 2-grams 的分布情况可参阅文献〔Kon81，表 2、3、4〕。当然，我们可以按此作出 3-grams 和更多 grams 的分布表。在下面这个例子中，我们给出另一种不同的方法。

例 1.3：信源 S 是由转移矩阵 $P=(P(t|s))$ $0<s, t<q$ 和平衡分布 $P\_\_=(P(0), P(1), \cdots, P(q-1))$ 的马尔可夫（Markov）链产生的 1-grams，因此，

$$Pr_{plain}\{(m_0, m_1, \cdots, m_{n-1})\}=P(m_0)P(m_1|m_0)P(m_2|m_1)\cdots P(m_{n-1}|m_{n-2})$$

这里，P 是一个 $q\times q$ 的马尔可夫矩阵，它的 q 个横行是概率分布（$P(t|s)$ 表示符号 S 后面是符号 t 的概率）。向量 $P\_\_$ 是特征值为 1 的 P 的特征向量的概率分布。

令 P 和 $P\_\_$ 由表 1.2 和 1.3 给出。

现在，我们得到更实际的概率

$$Pr_{plain}\{(run)\}=0.0751\times 0.0192\times 0.1517=2.19\times 10^{-4}$$

$$Pr_{plain}\{(urn)\}=0.0272\times 0.1460\times 0.0325\approx 1.29\times 10^{-4}$$

$$Pr_{plain}\{(nru)\}=0.0814\times 0.0011\times 0.0192\approx 1.72\times 10^{-6}$$

a	0.0723	h	0.0402	o	0.0716	u	0.0272
b	0.0060	i	0.0787	p	0.0161	v	0.0117
c	0.0282	j	0.0006	q	0.0007	w	0.0078
d	0.0483	k	0.0064	r	0.0751	x	0.0030
e	0.1566	l	0.0396	s	0.0715	y	0.0168
f	0.0167	m	0.0236	t	0.0773	z	0.0010
g	0.0216	n	0.0814				

表 1.3 英语的平衡分布

注意，在上述三个例子中，其模型都是平稳的，即 $Pr_{plain}\{Mj=m_0, M_{j+1}=m_1\cdots\cdots M_{j+n-1}=m_{n-1}\}$ 与 j 无关。在一份报文的中间，可认为此特性成立，但在其它情况下，例如一封信开头的日期等就并非如此。

既然我们已对什么是明文源作了说明，下面就可讨论图 1.1.

"密码变换" E 是 Z_q^* 到 $Z_{q'}^*$ 的一一映射．在大多数情况下，q 将等于 q'．为了避免数据扩展，E 也常常将 n-grams 映射到 n-grams．令 m 为某一明文，则 $c=E(m)$ 称为密文．密码体制 E 是密码变换 $E=\{E_k|k\in K\}$ 的集合．指数集 K 称为"密钥空间"，其元素称为"密钥"．

由于 E_k 是一一映射的，因此其逆必然存在．我们用 D_k 来表示．当然，E 表示"加密"，D 表示"脱密"．我们有

$$\forall_m \forall_{k\in K}[D_k(E_k(m))=m]. \tag{1.1}$$

如果信源 A 想把数据 m 发送给信宿 B，那么他将用密码变换 E_k 把 m 加密成密文 c．A 和 B 都知道通过"保密信道"特殊选择的密钥．该信道可以是信使，也可以是 A 和 B 事先商定选择的密钥 k．B 通过计算

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

就可以脱密密文 c．一般情况下，同一种密码体制可长期使用．但必须经常更换密钥，以确保数据的安全．与传输线路连接的密码分析者可以是：

i) 被动的（窃听）：密码分析者试图根据密文 c（和已掌握的其它情况）求得明文 m（甚至求得密钥 k）．

ii) 主动的（窜改）：密码分析者试图主动处理正在传送的数据．例如，他可以编写自己的密文、重新传送过时的密文等等．

一般来说，密码分析有以下三种：

i) 仅知密文攻击：密码分析者只知道一段密文（常常还有报文的上下文关系）．

ii) 已知明文攻击：密码分析者已知一段对应于明文的密文．如果一种体制能抵制这一类型的攻击，人们就不必破坏脱密的报文．

iii) 选择明文攻击：密码分析者有一段他或她选择的对应于密文的明文．后几章将要讨论的公开密钥密码体制必须抵制这一类型的攻击．

对传统密码体制的一般介绍（如图 1.4 所示）到此结束．下面我们将讨论几种特殊的密码体制．

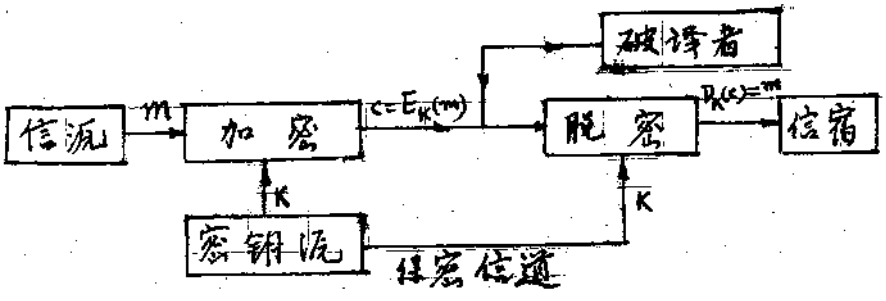


图 1.4 传统密码体制

第二章　古典密码体制

2.1　凯撒、单表代替、维吉尼亚密码

本章将讨论几种古典密码体制。为便于进一步阅读，建议对这一问题感兴趣的读者参阅文献〔Bek82〕、〔Den82〕、〔Kah67〕、〔Kon81〕、〔Mey82〕。在这些古典密码中，有一种称为 Julius Caesar（朱利叶斯·凯撒）的密码。凯撒密码用我们的术语定义为：

$$E_k(i)=(i+k)\ mod\ q, 0\leqslant i<q, \tag{2.1}$$

$$E=\{E_k|0\leqslant k<q\}, \tag{2.2}$$

式中 $i \bmod n$ 表示唯一满足 $j\equiv i \bmod n$，且 $0<j<n$ 的整数 j。在这种情况下，密钥空间 K 是集合 $\{0, 1, \cdots, q-1\}$，且 $D_k=E_{q-k}$。对大多数实用的字母表，密码分析者通过对所有 q 个可能的密钥进行尝试就能轻而易举地破译这种体制。这种尝试称为"穷尽密钥搜索"。例如，当 $q=26$，并用 $\{a, b, \cdots, z\}$ 作为字母表时，人们只需检验 26 种可能性。在表 2.1 中，可找到对密文 IYBABZ 的密码分析。

0	IYBABZ	6	OEHGHF	11	TJMLMK	16	YORQRP	21	DTWVWU
1	JZCBCA	7	PFIHIG	12	UKNMNL	17	ZPSRSQ	22	EUXWXV
2	KADCDB	8	QGJIJH	13	VLONOM	18	AQTSTR	23	FVYXYW
3	LBEDEC	9	RHKJKI	14	WMPOPN	19	BRUTUS	24	GWZYZX
4	MCFEFD	10	SILKLJ	15	XNQPQO	20	CSVUVT	25	HXAZAY
5	NDGFGE								

表 2.1 凯撒密文 IYBABZ 的密码分析

单表代替：在这种情况下，$K=S_q$，即所有 $\{0, 1, \cdots, q-1\}$ 置换的集合。密码体制 E 定义为

$$E=\{E_\pi|\pi\in S_q\} \tag{2.3}$$

式中

$$E_\pi(i)=\pi(i),\ 0\leqslant i<q \tag{2.4}$$

脱密函数 D_π 由 $D_\pi=E_{\pi^{-1}}$ 给出，因为

$$D_\pi(E_\pi(i))=E_{\pi^{-1}}(E_\pi(i)=E_{\pi^{-1}}(\pi(i))=\pi^{-1}(\pi(i))=i\quad 0\leqslant i\leqslant q$$

该体制不存在密钥空间小的问题。实际上，$|S_q|=q!$ 例如，如 $q=26$，密钥空间的大小就是 $26!\approx 4.03*10^{26}$。然而，该体制确实很好地证明了：密钥空间大并不能使人相信该体制是保密的！恰恰相反，人们只要对密文中的字母频率进行简单计算，并与表 1.2 中的字母频率进行比较，就能借助 π，很快发现明文中使用频率最高的字母的映象。此后，

要填入其它的字母就一点也不困难了．在下面这个例子中，我们已经知道表 2.2 中的密文和有关明文的附加信息，即明文讨论"bidirectional communication theory"．密码分析将证明是很容易的．

zhjeo	ndlze	hicle	osiol	digic	lmhzq	zolyi	zehdp	zhjeo	ndlze
hycdh	hipvs	uczyc	dhzhj	eondl	zehge	moyik	zhjpm	ihyig	gldiz
gizyd	ppsdo	iyizr	iosye	nnmhz	ydize	hicle	osceu	irioq	igyoz
vigic	ineol	flhio	dpydg	izhuc	zyciu	eeone	olzhj	eondi	zehge
moyig	zhjpm	ihyil	dycei	ciogl	dizgi	zydpp	siciq	zolyi	zehej
iczgz	hjpmi	hyizg	ikaol	ggiqv	sqzol	yiiqi	odhgj	eondi	zehxm
dhizj	zlguc	zycyd	hehps	viqio	zriqz	jicip	duejy	dmgdp	ziszg
evgio	riqqz	gizhf	mzgcz	hfici	idopz	ioydm	gijoe	nicip	diiol
jjiyi	zhvze	pefsd	hqgey	zepef	syenn	mhzyd	izehi	cieos	giing
iecdr	iuzqi	daapz	ydize	hgqmi	ieici	jdyji	cdipz	rzhfv	izhfg
doiys	icizo	dyize	hggem	oyige	jzhje	ondiz	ehucz	yczhj	pmihy
iidyc	eicio	zhdpp	aeggz	vpiqz	oiyiz	ehgic	iaoig	ihiad	aioqi
gyzvi	gicly	dgiej	vzqzo	iyize	hdpye	nnmhz	ydize	hicie	osdaa
pziqi	eicig	eyzdp	vicdr	zemoe	jneht	isg..			

表 2.2 用单表代替获得的密文

假设单词"communication"将在明文中出现．我们在密文中寻找 13 个连续的字母串，其中字 1＝字母 8，字母 2＝字母 12，字母 3＝字母 4，字母 6＝字母 13，字母 7＝字母 11．实际上，我们在密文中三次找到字母串"yennmhzydizch"．这给出了以下有关 π 的信息．

$$\begin{array}{cccccccc} c & o & m & u & n & i & a & t \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ y & e & n & m & h & z & d & i \end{array} \tag{2.5}$$

假设单词"direction"也出现在明文中．根据 (2.5)，我们需要在密文中寻找形式为"＊z＊＊yezch"的字母串，这证明"qzolyizch"出现 4 次，并给出

$$\begin{array}{ccc} d & r & e \\ \downarrow & \downarrow & \downarrow \\ q & o & l \end{array} \tag{2.6}$$

如果把 (2.5) 和 (2.6) 代入密文，我们便很容易完整地得到 π．例如，报文是这样开始的：

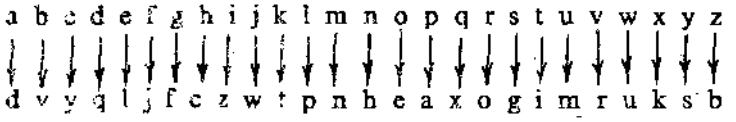
in＊ormationt＊eor＊treat＊t＊eynid…，

这显然是从

information theory treats the unid(irectional…)

得来的．

这便给出了字母 f、h、y 和 s 的 π 映象．用这样的方法进行计算，人们就很容易完整地得到 π．



维吉尼亚密码体制（以法国人 B.de Vigenere 命名，他于 1586 年写了《论密码》一书，书中含有这种体制更难的形式）是由 r 个凯撒密码组成，它们周期性地被使用，更严格地说，该体制定义为：

$$E=\{E_{(k_0,k_1,\cdots k_{r-1})}|(k_0,k_1,\cdots k_{r-1})\in K=Z_q^r\} \tag{2.7}$$

和

$$E_{(k_0,k_1,\cdots k_{r-1})}\ (m_0,m_1,m_2,\cdots)\ =\ (C_0,C_1,C_2) \tag{2.8}$$

而

$$C_i=(m_i+k_{imodr})\ mod\ q \tag{2.9}$$

		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
0	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
1	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
2	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
3	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
4	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
5	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
6	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
7	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
8	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
9	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
10	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
11	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
12	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
13	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
14	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
15	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
16	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
17	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
18	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
19	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
20	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
21	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
22	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
23	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
24	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
25	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

表 2.3 维吉尼亚表

例 2.1：我们用 $\{a, b, \cdots, z\}$ 表示 $\{0, 1, \cdots, 25\}$。所谓的"维吉尼亚表"（见表 2.3）在进行加密或脱密时是一种很有用的工具。用密钥"michael"得到以下加密过程：

明文　a c r y p t o s y s t e m o f t e n i s a c o m p r o m i s e b e t w e e n . . .

密钥　m i c h a e l m i c h a e l m i c h a e l m i c h a e l m i c h a e l m i c . . .

密文　m k t f p x z e g u a e q z r b g u i w l o w o w r s x u a g i e x h q m p . . .

由于英文中有冗余度，因此人们可以通过选择现有单词作为密钥，大大减小有效密钥空间的规模．用一个亲戚的名字会使加密的保密性或多或少降为零．

当然也可以周期地使用 r 个单表代替，而不是在维吉尼亚密码体制中周期地使用 r 个凯撒密码．这种体制称为"多表代替"体制．好几个世纪都没有人能找到破开这一体制的有效方法．其主要原因是人们没有确定密钥长度 r 的技术．一旦知道 r，就能通过将字母 i，$r+i$，$2r+i$，…，$0<i<r$ 组合在一起，求得 r 种单表代替，并能分别逐个破译这 r 种单表代替．1863 年，普鲁士军官 F.W.Kasiski 用一种称为"重合码分布"的技术，解决了求密钥长度 r 这一问题．下面我们来讨论这种方法．

2.2　重合码分布

假设 $M_i=(M_{i,0}, M_{i,1}, \cdots M_{i,n-1})$，$i=1$，2 为 Z_q 上两个独立、同分布的随机变量序列．即：

$$Pr_{plain}\{M_{ij}=m\}=p(m),\ 0\leqslant m<q \quad i=1,\ 2 \quad 0\leqslant j<n \tag{2.10}$$

我们对重合码的数量感兴趣，它定义为：

$$K[\underline{M}_1,\underline{M}_2]=|\{0\leqslant j<n|M_{1,j}=M_{2,j}\}| \tag{2.11}$$

显然，

$$Pr_{plain}\{M_{1,j}=M_{2,j}\}=\sum_m Pr_{plain}\{M_{1,j}=M_{2,j}=m\}=\sum_m P^2 \quad (m) \tag{2.12}$$

令 G 为 S_q 的一个子集，且假定有 G 上的一个随机变量$\prod$，其分布由下式给出：

$$Pr_{key}\{\prod=\pi\}=q(\pi) \tag{2.13}$$

令 M_1，$i=1$，2 分别用 G 中的单表代替 π_1，π_2 来加密．它们相应的映象是 $\underline{C}_1=(c_{1,0},C_{1,1},\cdots, C_{1,n-1})$ 和 $\underline{C}_2=(C_{2,0}, C_{2,1}, \cdots, C_{2,n-1})$．当然，对 i=1，2，且 $0<j<n$．

$$\mathrm{Pr}_{cipher}\{C_{i,j}=c\}=\sum_{\pi\in G}\mathrm{q}(\pi)\cdot \mathrm{P}(\pi^{-1} \quad (c)) \tag{2.14}$$

我们必须考虑以下两种可能性：

H_0：$\underline{M}_1$ 和 $\underline{M}_2$ 用同一种单表代替 π（概率 q（π））来加密。

H_1：$\underline{M}_1$ 和 $\underline{M}_2$ 分别用两个独立选择的单表代替 π_1 和 π_2（概率分别为 q（π_1）和 q（π_2））来加密。

于是，

$$Pr_{cipher}\{C_{1j}=C_{2j}|H_0\}=\sum_c Pr_{cipher} \quad \{C_{1,j}=C_{2,j}=c|H_0\}=$$

$$\sum_{\pi\in G}q(\pi)\sum_c p^2 \quad (\pi^{-1} \quad (c))=\sum_{\pi\in G}q(\pi)\sum_m p^2 \quad (m)=\sum_m p^2 \quad (m \tag{2.15}$$

而，

$$Pr_{cipher}\{C_{1j}=C_{2j}|H_1\}=\sum_c Pr_{cipher}\{C_{1,j}=C_{2,j}=c|H_1\}=$$

$$\sum_{\pi_1\pi_2\in G} q(\pi_1)q(\pi_2)\sum_c p\ (\pi^{-1}{}_1\ \ (c))\ \ p\ (\pi^{-1}{}_2\ \ (c))\ \ =$$

$$\sum_c\{\sum_{\pi_1\in G} q(\pi_1)p\ (\pi^{-1}{}_1\ \ (c))\ \}.\{\{\sum_{\pi_2\in G} q(\pi_2)p\ (\pi^{-1}{}_2\ \ (c))\ \}=$$

$$\sum_c\{\sum_{\pi\in G} q(\pi)p\ (\pi^{-1}\ \ (c))\ \}^2\ \ =\sum_c Pr^2{}_{cipher}\{C=c\} \tag{2.16}$$

例 2.2：借助表 1.1，对一英文，我们按公式（2.15）得值 0.06875．如果我们取 $G=S_{26}$ 或 G 由 26 种凯撒加密所组成，且假设 G 中的所有元素都同等可能，那么（2.16）得到的值为 $1/26\approx 0.03846$．因此在假设 H_0 的情况下，K〔C_1，C_2〕的期望值大约为 $0.06875n$；在假设 H_1 的情况下，K〔C_1，C_2〕的期望值则大约为 $0.03846n$．

ubsyv	kmhvy	rrtsb	bcrds	ndwrt	shxmb	ufrmx	gabnv	mircc	wcruc
amlyz	brvfw	lvvml	yzwaq	spyog	ssiec	hbgcu	bsvyc	zqrcw	rmhvc
xgooy	vcyds	pomtq	fpyqk	gbvme	rucad	lcafl	rsuqj	rbhce	qesfc
ehuoq	mdsto	rcdoy	meqqw	aglgo	vggsm	dabbl	gztbb	qyfwb	xwmgf
powgz	tycll	osrkg	fahuo	vqfog	swruq	nvpwf	vrnmp	qqgss	latgr
mqubs	vyczq	rswcj	deowq	qroih	gdspd	tbffn	xwgzt	bbqyf	wbxus
mbgsx	gqgjr	matqn	xslxm	oohcd	wiohg	rhuop	yicsm	eseoy	sxwug
blwcd	jrnhg	ehvxk	sugus	refvr	oepxw	rbgyg	grpvm	yhuop	yfseo
jdqqg	srzuc	yykwm	bqkrb	ecpss	jaulm	skyla	sgyfw	bxxfq	ceiwc
qafds	fmjrg	mbqoc	zpgoo	gssle	rhoxm	fvroj	dqqgd	lyfzv	fmisp
rsree	ceofw	fvrsv	yohvy	rqecb	bgcec	ssrda	fzkxg	abjrm	atwap
psqbp	oiyov	bdtcd	wakpj	bcfcm	zxsqs	vcohv	yrqfv	rzvce	sadty
hseni	qofvl	lqfvr	meqqc	simbu	frmxg	abnvg	myahx	mamhv	yrrts
bbcyb	dysib	fcgri	qaqvk	pzqvn	fmmgf	bpqmz	yrlwr	iczyr	iqmfy
jyecp	sejsf	betfm	jrllc	zdhdx	mssgr	lpubb	xiamu	rkrbf	vrsvy
ohvyr	qtolo	fcqbb	lwcdj	rnelp	frqmq	fseoh	dafnz	lpucq	yjdut
goily	waexc	etvfl	rkdvm	ejasu	kzgai	ekpyo	hvfmr	usfcl	yfwaq
pcmjv	xkeqb	vdejp	wfzpy	kquev	pooyv	tsevv	xkdaf	zohrl	cfder
uggsg	yxhvw	iqqeh	oraqg	gyafu	qudlc	nwqsv	cohvy	ryxqb	wqszw
pkxga	bgrlm	dmukw	zqsak	tnxwr	nxfqf	rcyjf	gjovc	fmcsg	yxtbb
xfqgb	mmyxf	rverv	cakrb	mueoi	brceo	eatgr	vlafs	qzegd	csdam
ubqsz	gpinv	wricn	vvcmr	lurml	hzyhg	rwpkx	gabfy	jrtsq	yqgzo
adfct	oisss	dguya	cphlz	mamzl	kpsqg	jsxfd	stkvb	fcgrl	aaaze
rgoog	sslci	nxxgf	wrcxf	qprre	tuchb	sdpwp	deraf	vxafu	qudlc
rfroc	afwhx	cqqlc	biqes	qlcrt	ssbic	qbgbs	nkcsd	lcpcz	sryzh
uxmkm	zvckp	qogov	yeqbw	tydsq	gmrih	uowsn	rbwml	mbgyr	cfvrw
ehafo	oiybw	bevyx	wapps	qbppe	qqlcb	iqesq	lcrts	qsvco	hrnxp
mbfsr	dafzk	xgabt	ylqrf	bwxfq	rbwml	mbgds	rtsfe	fbaav	xelfo
xsqyx	huofc	toiss	sdcsr	lpswa	glgoy	grldd	srkgr	ucacs	dncgr
efuan	vwyds	ntssf	sdeej	ycfds	dfvri	ifmjv	yypmz	vxjjg	samle
asfdl	cadcy	wgfsq	svcoh	vycim	arvcd	dcxdt	cnion	skubn	xxrnh
uohmy	wakrr	mbvwe	jfvrn	mafog	yvkar	vpmam	hvyrg	eiacx	yazrs
xmoqh	blbua	zohgm	hevcy	chrbx	fqsfd	exxwf	rqczh	bpxfq	ibwml
mbgbi	jmhvy	rqtwc	krbov	uakcp	qhxxg	zibew	jkwak	fmghf	sbuqs
xcxgy	svxxm	fvvri	pxnbn	mduqn	dmmzw	usgfb	fbfib	fnoox	fqgek
fjqca	n								

表 2.4 用维吉尼亚密码体制得到的密文