第一部分　加解密

01.　古典密码

厚 德 健 行

MXDXBVTZWVMXNSPBQXLIMSCCSGXSCJXBOVQXCJZMOJZCVC
TVWJCZAAXZBCSSCJXBQCJZCOJZCNSPOXBXSBTVWJC
JZDXGXXMOZQMSCSCJXBOVQXCJZMOJZCNSPJZHGXXMOSPLH
JZDXZAAXZBXHCSCJXTCSGXSCJXBOVQX

Lewis Carroll的明码文本, *爱丽丝梦游仙境*

这个答案不会像你第一遍草率的看到这些字符想的一样困难。这些字符, 像一些人很容易猜到的，是用密码书写的，确切的说, 它们传达了一个意思...

——Edgar Allan Poe, *The Gold Bug*

密码的含义

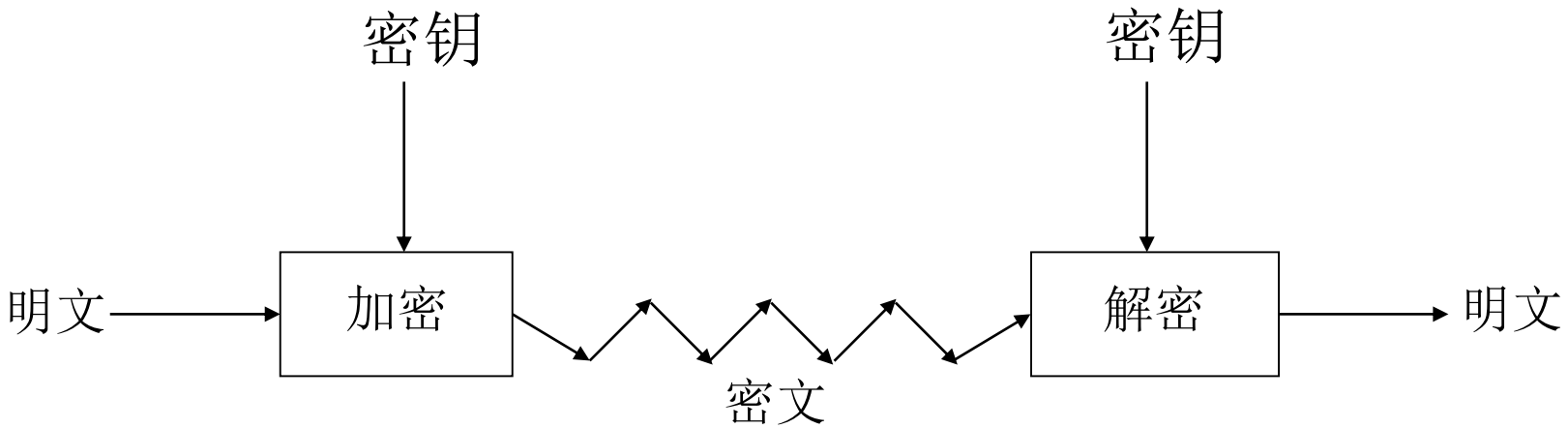
- 密码学cryptology — 是编制和破译"密码"的科学与技术。
- 密码编码学cryptography—研究"密码"的编制。
- 密码分析学cryptanalysis—研究"密码"的破译。
- 密码crypto—包含以上几个方面(甚至更多)。

如何表示密码

- 一个**密码**或者**密码体制**是用于**加密**数据(原始数据为**明文**)的
- 对原始数据**加密**的结果是**密文**
- 通过**解密**将密文恢复成**明文**
- 在密码体制的加密和解密过程中要使用**密钥**
- 在**对称密钥**密码中，加密和解密过程使用相同的密钥
- 在**公钥密码**中，加密和解密过程中使用不相同的密钥。

密码含义

- 基本原则
  - 假定密码体制内部的所有细节都被攻击者获知
  - 唯有密钥没有被攻击者掌握。
  - 这就是说，密码算法是没有秘密的。
- 也被称为**Kerckhoffs （克霍夫斯）准则**
- 我们为什么要作这样的假设？
  - 经验表明一旦保密的密码算法公开使用，就很难做到继续保密
  - 软硬件技术可分析秘密算法
  - 最好事先发现密码的缺陷



对称密钥加密的示意图

简单代替密码

- 明文: fourscoreandsevenyearsago
- 密文:

明文	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	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

- 密文:

  IRXUVFRUHDAGVHYHABHDUVDIR
- 这种移动3位的简单替代密码称为"凯撒密码"

凯撒密码的解密

- 假设某段密文采用的是凯撒密码
- 密文: VSRQJHEREVTXDUHSDQWU

明文	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	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

- 明文: spongebobsquarepants

变型一：非简单替代密码

- "移动n位"，$n \in \{0,1,2,\ldots,25\}$
- 那么密钥是：n
- 例如: key = 7

明文：	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：	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

密码分析 I: All穷举法

- 加密时使用了简单替换密码(移动n)
  - 但密钥不可知，
- 假设密文为: CSYEVIXIVQMREXIH
- 如何找到密钥n？
- 密钥只有26种可能—尝试所有可能的密钥!
- 穷举搜索密钥
- 得到结果: key = 4

变型二：弗吉尼亚密码

- “多表简单替换"，$n \in \{0,1,2,\ldots,25\}$
- 那么密钥是：$k_1, k_2, \ldots, k_m, k_i \in n$
- 加密

$e_k(p_1, p_2, \ldots, p_m) = (p_1+k_1, p_2+k_2, \ldots, p_m+k_m) \pmod{26}$

- 解密

$d_k(c_1, c_2, \ldots, c_m) = (c_1-k_1, c_2-k_2, \ldots, c_m-k_m) \pmod{26}$

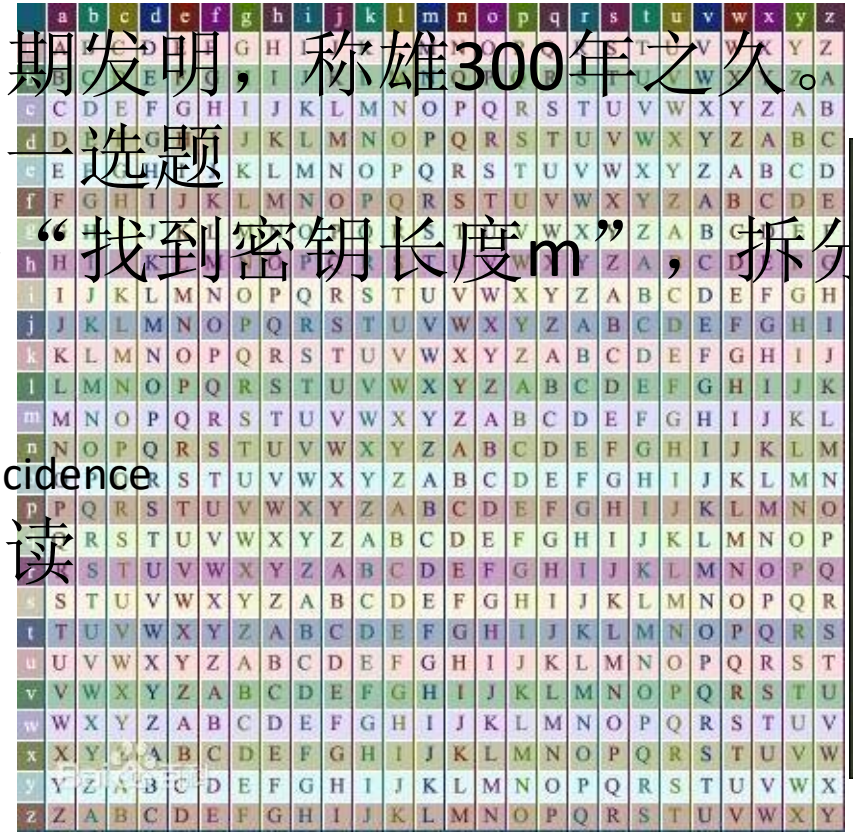
明文：	C	R	Y	P	T	O	G	R	A	P	H	Y
密钥：	L	U	C	K	L	U	C	K	L	U	C	K
密文：	N	L	A	Z	E	I	I	B	L	J	J	I

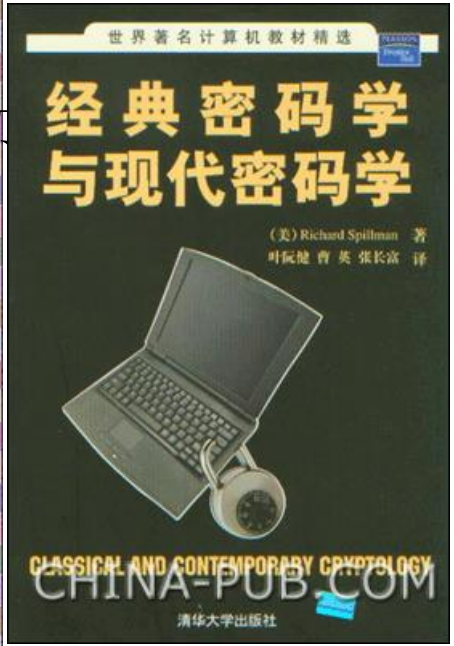
变型二：弗吉尼亚密码

- 16世纪中期发明，称雄300年之久。
- 我们实验一选题
- 破解思路“找到密钥长度m”，拆分
- 方法：

  Kasisky test

  Index of coincidence
- 进一步阅读





变型三：简单替代密码

- 密钥是全表字母的重排列
- 密钥不是字母表的移动
- 例如：

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

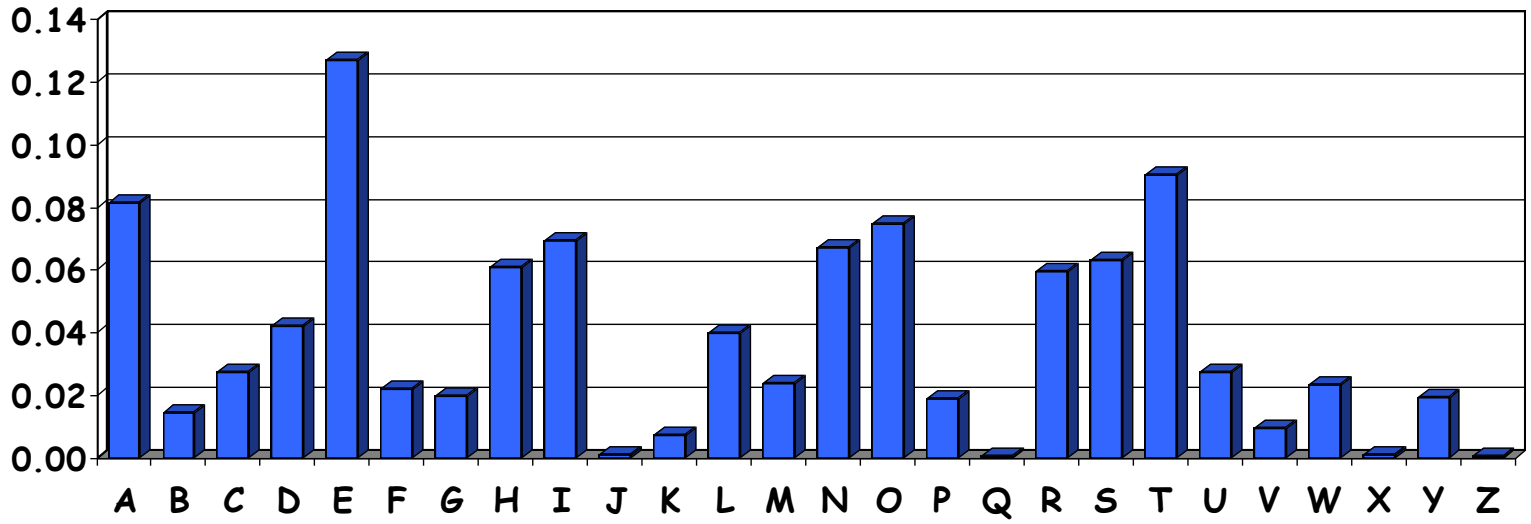
- 那么有 $26! \cong 4*10^{26}$ (2^{88})种可能的密钥!

密码分析II: 更好的办法

- 假设攻击者截获一段密文，并知道其是由简单替换加密所得
- 但不一定是由移动n位变形得到的
- 是否能找出下面给定密文的密钥：

PBFPVYFBQXZTYFPBFEQJHDXXQVAPTPQJKTOYQWIPBVWLXTOXBTF
XQWAXBVCXQWAXFQJVWLEQNTOZQGGQLFXQWAKVWLXQWAEBIPBFXFQ
VXGTVJVWLBTPQWAEBFPBFHCVLXBQUFEVWLXGDPEQVPQGVPPBFTIXPFHXZ
HVFAGFOTHFEFBQUFTDHZBQPOTHXTYFTODXQHFTDPTOGHFQPBQWAQJJ
TODXQHFOQPWTBDHHIXQVAPBFZQHCFWPFHPBFIPBQWKFABVYYDZBOTHP
BQPQJTQOTOGHFQAPBFEQJHDXXQVAVXEBQPEFZBVFOJIWFFACFCCFHQW
AUVWFLQHGFXVAFXQHFUFHILTTAVWAFFAWTEVOITDHFHFQAITIXPFHXA
FQHEFZQWGFLVWPTOFFA

- 不可能尝试所有的2^{88}个密钥
- 有更好的办法吗？
- 英文字母频率统计



密码分析 II

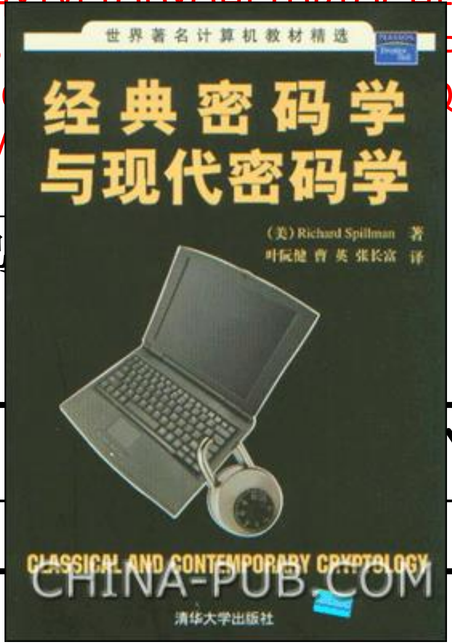
- 密文:

PBFPVYFBQXZTYFPBFEQJHDXXQVAPTPQJKTOYQWIPBVWLXTOXBTFXQWAXBVCXQWAXFQJVW
LEQNTOZQGGQLFXQWAKVWLXQWAEBIPBFXFQVXGTVJVWLBTPQWAEBFPBFHCVLXBQUFEVW
LXGDPEQVPQGVPPBFTIXPFHXZHVFAGFOTHFEFBQUFTDHZBQPOTH
QPBQWAQJJTODXQHFOQPWTBDHHIXQVAPBFZQHCFWPFHPBFIPB ... P
QJTQOTOGHFQAPBFEQJHDXXQVAVXEBQPEFZBVFOJIWFFACFCCFH ... Q
HFUFHILTTAVWAFFAWTEVOITDHFHFQAITIXPFHXAFQHEFZQWGFLV

- 使用下面的信息解密密文(进一步阅

密文字母频率统计:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	Y	Z
21	26	6	10	12	51	10	25	10	9	3	10	0	1	15	28	42	0	6	8



密码分析: 术语

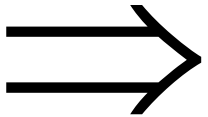
- 如果已知最好的攻击方法所需要的工作量与穷举密钥相当，则称一个密码系统是安全的
  - 即穷举秘钥搜索
- 如果存在比穷举更好的攻击方法，则此密码系统是有缺陷的
- 但是这样一来，一个不安全密码系统可能比一个看来安全的密码系统更难被破译。
  - 为什么呢？

双重置换密码

- 明文: attackxatxdawn

	col 1	col 2	col 3
row 1	a	t	t
row 2	a	c	k
row 3	x	a	t
row 4	x	d	a
row 5	w	n	x

进行列置换



	col 1	col 3	col 2
row 3	x	t	a
row 5	w	x	n
row 1	a	t	t
row 4	x	a	d
row 2	a	k	c

- 密文: xtawxnattxadakc
- 密钥: 由矩阵的大小和行列置换规则组成 如：(3,5,1,4,2) 和 (1,3,2)

简单的完美密码："一次一密"

e=000 h=001 i=010 k=011 l=100 r=101 s=110 t=111

加密规则：明文 ⊕ 密钥 = 密文

	h	e	i	l	h	i	t	l	e	r
明文：	001	000	010	100	001	010	111	100	000	101
密钥：	111	101	110	101	111	100	000	101	110	000
密文：	110	101	100	001	110	110	111	001	110	101
	s	r	l	h	s	s	t	h	s	r

"一次一密"的解密

e=000 h=001 i=010 k=011 l=100 r=101 s=110 t=111

解密规划：密文 ⊕ 密钥 = 明文

	s	r	l	h	s	s	t	h	s	r
密文:	110	101	100	001	110	110	111	001	110	101
密钥:	111	101	110	101	111	100	000	101	110	000
明文:	001	000	010	100	001	010	111	100	000	101
	h	e	i	l	h	i	t	l	e	r

双重间谍(Alice)假意提供"密钥":

	s	r	l	h	s	s	t	h	s	r
密文:	110	101	100	001	110	110	111	001	110	101
"密钥":	101	111	000	101	111	100	000	101	110	000
"明文":	011	010	100	100	001	010	111	100	000	101
	k	i	l	l	h	i	t	l	e	r

e=000 h=001 i=010 k=011 l=100 r=101 s=110 t=111

发送消息的Alice被抓获，其提供密钥是：

	s	r	l	h	s	s	t	h	s	r
密文:	110	101	100	001	110	110	111	001	110	101
"密钥":	111	101	000	011	101	110	001	011	101	101
"明文":	001	000	100	010	011	000	110	010	011	000
	h	e	l	i	k	e	s	i	k	e

e=000 h=001 i=010 k=011 l=100 r=101 s=110 t=111

关于"一次一密"方案的总结

- 可证明安全…
  - 密文没有泄露与明文相关的任何信息
  - 所有的"明文"几率均等
- …但是，只有当被正确使用
  - 密钥必须是随机的，且只使用一次
  - 密钥只有发送者和接收者知道

注意：密钥的长度与明文长度相同

- 这样的话，为何一次一密的密钥只能使用一次呢？

二十世纪早期

- 一战——Zimmerman 电报
- 在1929，美国国务卿Henry L. Stimson——"绅士不偷看别人的书信"
- 二战——密码分析的"黄金时代"
  - 中途岛和珊瑚海的胜利
  - 日本Purple密码 (MAGIC编码)
  - 德国Enigma密码 (ULTRA编码)

二战后---密码历史

- 香农 — 信息论科学之父
- 计算机带来的变革 — 大量数据
- 数据加密标准 (DES), 70’s
- 公钥密钥编码学, 70’s
- CRYPTO会议, 80’s
- 高级加密标准 (AES), 90’s
- **密码技术延伸到各个领域**

香农

- 信息论建立者
- 1949 论文: *Comm. Thy. of Secrecy Systems*
- 密码设计两原则：**混淆**和**扩散**
  - **混淆**—隐藏密钥和密文之间的关系
  - **扩散**—将明文中的统计信息散布到整个密文中
- 一次一密可证明是安全的
- 一次一密只采用了混淆，而双重置换密码只采用了扩散

密码分析学的分类

对于Trudy来说从信息可用的角度

- 唯密文
- 已知明文
- 选择明文
  - "午餐时间攻击"
  - 协议可能加密选定的明文
- **自适应**选择明文
- 密钥相关攻击
- 前向搜索 (仅对公钥加密)
- 等等.....

练习

1.假定下面截取的是某个经典电报密码本加密方案中的一部分被解密的密码本片段。

123 once
199 or
202 maybe
221 twice
233 time
332 upon
451 a

请解密如下密文：

242, 554, 650, 464, 532, 749, 567

假设有如下附加序列用于加密该消息：

119, 222, 199, 231, 333, 547, 346

Once upon a time or maybe twice.

练习

2.假设Alice用一种安全的加密方案加密了一条消息，该方案使用40位的密钥。Trudy知道密文和加密算法，但是她不知道明文和密钥。Trudy计划实施一次穷举式检索攻击，也就是说，她打算尝试每一个可能的密钥，直到她能够找到那个正确的密钥。请思考如下问题：

a. 平均而言，Trudy在找到那个正确的密钥之前要尝试多少个密钥呢？

2^{39}

b. Trudy如何才能够知道她找到了那个正确的密钥呢？注意：对于Trudy来说，因为有太多的选择，所以不可能手工去检查每一个密钥，她必须有一些自动化的手段来判定假设的密钥正确与否。

设置自动化的统计分析方法，判断用密钥还原出的明文是否符合英语的“字母频率”逻辑。

练习

3.加密如下消息：

we are all together

用4行乘4列的双换位密码(即本书中描述的那类双换位密码)进行加密，使用的行置换如下：

(1, 2, 3, 4) → (2, 4, 1, 3)

使用的列置换如下：

(1, 2, 3, 4) → (3, 1, 2, 4)

lealethrawergtoe

关注我，下节内容更精彩：02：对称密码