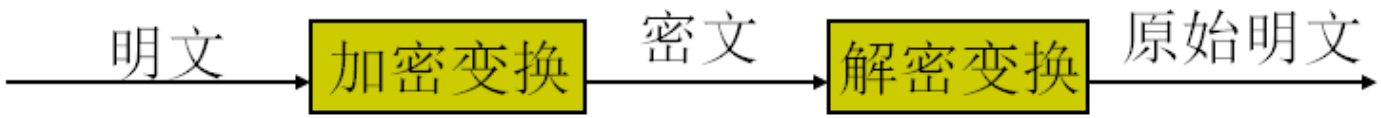
应用密码学（第二讲）
— 经典密码学

林东岱

信息安全国家重点实验室

2022年9月

什么是密码?



$$E(M)=C, \quad D(C)=M, \quad D(E(M))=M$$

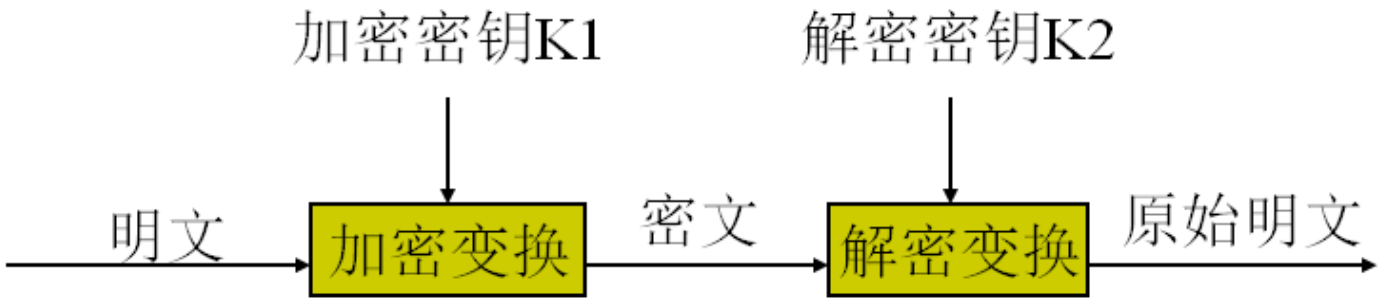
明文：通信的过程中,发送者想给接收者发送的消息。

加密：用某种方法伪装消息以隐藏其内容的过程。

密文：被加密的消息称为密文(ciphertext)。

解密：把密文转变成明文的过程。

算法和密钥



$E_{K1}(M)=C, \quad D_{K2}(C)=M, \quad D_{K2}(E_{K1}(M))=M$

本节概要

第一部分

替换密码体制

几个实例...

- 黑话
- 《红灯记》中的密电码
- 跳舞的小人
- Caesar密码
- .....

替换密码

- 凯撒(Caesar)密码

  明文: **Caesar cipher is a shift substitution**

  密文: **FDHVDU FLSKHU LV D VKLIW VXEVWLWXWLRQ**

- 移位密码(shift substitution cipher)

  加密变换为:

  $Ek(i)=(i+k) \bmod q = j,\ 0 \leqslant i, j < q,\ K = \{k \mid 0 \leqslant k < q\}$

  解密变换为:

  $Dk(j)=Eq-k(j) = (j+q-k) \bmod q = (j-k) \bmod q = i$

替换密码

- 乘数密码(multiplicative cipher)

  又称采样密码，加密变换：

  $E_k(i) = ik \bmod q = j \quad 0 \le j < q$

- 仿射密码（affine cipher）

  选取k_1, k_2两个参数，其中$(k_1, 26) = 1$，令

  $c = k_1 m + k_2 \bmod 26$

替换密码

- Playfair密码(一战英国)
  - 若m_1和m_2在同一行，则密文c_1和c_2分别紧靠m_1, m_2右端的字母。
  - 若m_1和m_2在同一列，则密文c_1和c_2分别是紧靠m_1, m_2下方的字母。
  - 若m_1和m_2不在同一行，也不在同一列，则c_1和c_2是出m_1, m_2确定的矩形的其他两角的字母，并c_1和m_1，c_2和m_2同行。
  - 若m_1=m_2，则插入空字母(比如Q)于重复字母之间。
  - 若明文字母数为奇数，将空字母Q加在明文的末端。

密码字为:
FIVESTARS

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

多表替换密码

- 维吉尼亚(Vigenere)密码

  明文: **The mark of the immature man is that he wants to die nobly for a cause while the mark of the mature man is that he wants to live humbly for one**

  密钥: **cipherkey** （$k = k_1 k_2 \ldots k_n$ ）

  密文: **vpt teiu sd vpt pqdkxstm bhr zc xfcb wl arxxq vw spi eyfja ndy e tky qg ewppv dlc oigr sw dlc oiibvv wel ka ioek ri ucviz xf vmtg pjtfci jmt wcl**

本节概要

第二部分

置换密码体制

换位密码

密文：卑者高者鄙的尚的是通是墓卑行高志鄙证尚铭

明文：

换位密码

密文：卑者高者鄙的尚的是通是墓卑行高志鄙证尚铭

明文：

卑	鄙	是	卑	鄙
者	的	通	行	证
高	尚	是	高	尚
者	的	墓	志	铭

Hill密码

- 基本思想是将t个明文字母通过线性变换转换为t个密文字母:

$$M = m_1 m_2 \ldots m_t,\ E_k(M) = c_1 c_2 \ldots c_t$$

$$\begin{aligned}
c_1 &= k_{11} m_1 + k_{12} m_2 + \ldots + k_{1t} m_t \\
c_2 &= k_{21} m_1 + k_{22} m_2 + \ldots + k_{2l} m_t \\
&\ldots\ldots \\
c_t &= k_{t1} m_1 + k_{t2} m_2 + \ldots + k_{tt} m_t
\end{aligned}$$

ADFGX密码(一战德国)

- 初始矩阵

	A	D	F	G	X
A	p	g	c	e	n
D	b	q	o	z	r
F	s	l	a	f	t
G	m	d	v	i	w
X	k	u	y	x	h

- 每一个明文字母用它所在行和列的标记代替，如s变成了*FA*，z变成了*DG*，*比如明文Kaiser Wilhelm变换为：*

  XA FF GG FA AG DX GX GG FD XX AG FD GA

ADFGX密码(一战德国)

- 选择关键字，比如Rhein，并用关键词标记矩阵的列

R	H	E	I	N
X	A	F	F	G
G	F	A	A	G
D	X	G	X	G
G	F	D	X	X
A	G	F	D	G
A				

ADFGX密码(一战德国)

- 重新调整列，使列的标记按字母表顺序排列

E	H	I	N	R
F	A	F	G	X
A	F	A	G	G
G	X	X	G	D
D	F	X	X	G
F	G	D	G	A
				A

- 按列向下读字母(忽略标记)可得到密文如下：

  FAGDFAFXFGFAXXDGGGXGXGDGAA

本节概要

第三部分

经典密码体制分析

英文字母频率的分布

字母	频率	字母	频率
A	8.176	N	6.749
B	1.492	O	7.507
C	2.782	P	1.929
D	4.253	Q	0.095
E	12.703	R	5.987
F	2.228	S	6.327
G	2.015	T	9.056
H	6.094	U	2.758
I	6.996	V	0.978
J	0.153	W	2.360
K	0.772	X	0.150
L	4.025	Y	1.974
M	2.406	Z	0.074

统计特性

- 英文字母的统计特性

极高频率字母组	E
次高频率字母组	T A O I N S H R
中等频率字母组	D L
低频率字母组	C U M W F G Y P B
甚低高频率字母组	V K J X Q Z

- 双字母组合的统计特性: TH HE IN EB AN
- 三字母组合的统特性: THE ING AND HER
- 英文单词的开始字母：E S D T
- 英文单词的结尾字母：T A S W

破译举例

- 密文： YKHLBA JCZ SVIJ JZB LZVHI JCZ VHJ DR IZXKHLBA VSS RDHEI DR YVJV LBXSKYLBA YLALJVS IFZZXC CVI LEFHDNZY EVBTRDSY JCZ FHLEVHT HZVIDB RDH JCLI CVI WZZB JCZ VYNZBJ DR ELXHDZSZXJHDBLXI JCZ XDEFSZQLJT DR JCZ RKBXJLDBI JCVJ XVB BDP WZ FZHRDHEZY WT JCZ EVXCLBZ CVI HLIZB YHVEVJLXVSST VI V HZIKSJ DR JCLI JCVJ PZHZ DBXZ XDBILYZHZY IZXKHZ VHZ BDP WHZVMVWSZ

A	B	C	D	E	F	G	H	I	J	K	L	M
5	24	19	23	12	7	0	24	21	29	6	20	1
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3	0	3	1	11	14	9	0	27	5	17	12	45

破译举例 I

极高频率字母组:	Z
次高频率字母组:	J V B H D I L C
中等频率字母组:	X S E Y R
低频率字母组:	T F K A W N P
甚低频率字母组:	M Q G O U

- 密文字母Z的频率最高，它一定是明文字母E。
- 在英语中只有一个单字母单词A，因此可以断定密文字母V对应于明文字母A。
- 三字母JCZ的频率最高，因此它一定就是THE。密文字母J对应于明文字母T，密文字母C对应于明文字母H。
- 密文字母J的频率处于第二位，进一步证明了其对应于明文字母为T。

破译举例 II

- 考察双字母单词VI。因为已知V对应于A，根据英语知识，只可能是AN，AS，AM，AT。首先它不是AN．否则因其后有冠词A而语法不通。又因J对应于T，故又不是AT，只能是AS或AM。明文字母M属于低频字母，而密文字母I后于高频字母，因此密文字母I对应于明文字母S，密文VI的明文为AS。
- 考察三字母单词VSS。因为已知V的明文为A，在英语中A后面接两个相同字母的单词只有从L，因此密文字母S对应于明文字母L。
- 在三字母单词VHZ中，因为已知V的明文为A，Z的明文为E，根据英语知识它只能是ARE或AGE。因为H在密文中属于高频字母，G在明文字母中属于低频字母，故H的明文为R。仿此分析三字母单词JZB，可知密文字母B对应于明文字母N，JZB的明文为TEN。
- 分析四字母单词JCLI，可知密文字母L对应于明文字母I。分析四字母单词WZZB，可知密文字母W对应于明文字母B。
- 由双字母单词WT可知密文字母T对应于明文字母Y。
- 由密文HZVIDB可推出密文字母D对应于明文字母O。

破译举例 III

- 双字母单词DR的频率很高，已知D的明文是O，则R的明文一定是F。
- 由三字母组BDP可推出密文字母P对应于明文字母W。
- 在密文DBXZ中，因为已知D，B，Z的明文，故可推出密文字母X对应于明文字母C。
- 从密文EVBT可推出密文字母E对应于明文字母M。
- 从密文IFZZXC可推出密文字母F对应于明文字母P。
- 从密文FZHRDHEZY可推出密文字母Y对应于明文字母D。
- 从密文JZXCBDSDAT可推出密文字母A对应于明文字母G。
- 同时注意到三字母词尾LBA的频率较高。进一步证明这一推断是正确的。
- 从密文YKHLBA可推出密文字母K对应于明文字母U。
- 从密文LEFHDNZY可推出密文字母N对应于明文字母V。
- 最后从WHZVMVPZHZ可知M对应于K。

破译举例

- 极高频率字母组 Z
- 次高频率字母组 J V B H D I L C
- 中等频率字母组 X S E Y R
- 低频率字母组 T F K A W N P
- 甚低频率字母组 M Q G O U
- 明文：DURING THE LAST TEN YEARS THE ART OF SECURING ALL FORMS OF DATA INCLUDING DIGITAL SPEECH HAS IMPROVED MANYFOLD THE PRIMARY REASON FOR THIS HAS BEEN THE ADVENT OF MICROELECTRONICS THE COMPLEXITY OF THE FUNCTION THAT CAN NOW BE PERFORMED BY THE MACHINE HAS RISEN DRAMATICALLY AS A RESULT OF THIS RECENT DEVELOPMENT IN TECHNOLOGY MANY OF THE CIPHER SYSTEM THAT WERE ONCE CONSIDERED SECURE ARE NOW BREAKABLE

Vigénere密码分析

1. 猜测密钥长度
   - Kasiski测试
   - 重合指数（index of coincidence）
2. 确定密钥字
   - 互重合指数（mutual index of coincidence）

定义 1 (重合指数)

设$X = x_1x_2\cdots x_n$是由n个字符组成的字符串，我们定义X的重合指数为X中任选两个字符，它们相同的概率，记为$I_c(X)$.

重合指数的计算

- 假定$f_0, f_1, \cdots, f_{25}$分别表示字母$A, B, \cdots, Z$在X中出现的频数（个数），则按定义

$$I_c(X) = \sum_{i=0}^{25} \frac{\binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)}$$

其中符号$\binom{m}{n}$表示在m个事物中取n个的组合数.

- 假定在字符串X中，字母$A, B, \cdots, Z$出现的频率分别是$p_0, p_1, \cdots, p_{25}$, 那么

$$I_c(X) = \sum_{i=0}^{25} p_i^2$$

- 容易计算，当X为自然语言时，$I_x(X) = 0.065$, 而当X是随机字符串时，$I_c = 26 \times (\frac{1}{26})^2 = 0.038$.

互重合指数（Mutual Index of coincidence）

定义 2 (互重合指数（Mutual Index of coincidence)

假定$X = x_1x_2\cdots x_m$, $Y = y_1y_2\cdots y_n$是两个长度分别为m和n的字符串。我们定义X和Y的互重合指数为X中的一个随机元素和Y中的一个随机元素相等的概率，记为$MI_c(X, Y)$.

如果将字母$A, B, \cdots, Z$在X和Y中出现的频数分别用$f_0, f_1, \cdots, f_{25}$和$g_0, g_1, \cdots, g_{25}$表示，那么

$$MI_c(X, Y) = \sum_{i=0}^{25} \frac{f_i}{m} \cdot \frac{g_i}{n} = \frac{\sum_{i=0}^{25} f_i g_i}{mn}$$

期望的互重合指数

相对位移	MI_c的期望值	相对位移	MI_c的期望值
0	0.065	7(19)	0.039
1(25)	0.039	8(18)	0.034
2(24)	0.032	9(17)	0.034
3(23)	0.034	10(16)	0.038
4(22)	0.044	11(15)	0.045
5(21)	0.033	12(14)	0.039
6(20)	0.036	13	0.043

应用密码学作业（一） I

1. 试写一篇利用密码理论或技术解决科学或实际中安全问题的小论文，要求不能重复已有解决方案， 必要时给出实现（该题本学期有效，通过者可在最终成绩中加5分）。
2. 试求Hill密码的密钥空间大小。
3. 下面有两段密文，第一个使用替换密码加密，第二个使用Vigenére密码加密，试确定它们的明文，要求写出你解密消息的具体步骤，包括统计分析和计算。
    1. 替换密码:
    EMGLOSUDCGDNCUSWYSFHNSFCYKDPUMLWGYICOXYSIPJCK
    QPKUGKMGOLICGINCGACKSNISACYKZSCKXECJCKSHYSXCG
    OIDPKZCNKSHICGIWYGKKGKGOLDSILKGOIUSIGLEDSPWZU
    GFZCCNDGYYSFUSZCNXEOJNCGYEOWEUPXEZGACGNFGLKNS
    ACIGOIYCKXCJUCIUZCFZCCNDGYYSFEUEKUZCSOCFZCCNC
    （提示： F 解密为 w）

应用密码学作业（一） II

2. Vigenére密码:
KCCPKBGUFDPHQTYAVINRRTMVGRKDNBVFDETDGILTXRGUD
DKOTFMBPVGEGLTGCKQRACQCWDNAWCRXIZAKFTLEWRPTYC
QKYVXCHKFTPONCQQRHJVAJUWETMCMSPKQDYHJVDAHCTRL
SVSKCGCZQQDZXGSFRLSWCWSJTBHAFSIASPRJAHKJRJUMV
GKMITZHFPDISPZLVLGWTFPLKKEBDPGCEBSHCTJRWXBAFS
PEZQNRWXCVYCGAONWDDKACKAWBBIKFTIOVKCGGHJVLNHI
FFSQESVYCLACNVRWBBIREPBBVFEXOSCDYGZWPFDTKFQIY
CWHJVLNHIQIBTKHJVNPIST

河防口长城