现代密码学

孙玉花
中国石油大学 理学院
sunyuhua_1@163.com
2020年2月

课程简介

- **课程名称**

  -现代密码学

- **课程性质、学时**

  -专业基础课；48学时（理论40学时，实验8学时）

- **课程目标**

  -知识：了解现代密码学基本理论，掌握现代密码基本技术（密码算法、密码协议），了解现代密码学发展方向。

  -能力：密码技术基本应用能力，自学能力。

- **预备知识及要求**

  -预备知识：离散数学，Matlab编程

- **成绩构成**

  -半开卷笔试（70%），实验成绩（20%），平时成绩（10%）

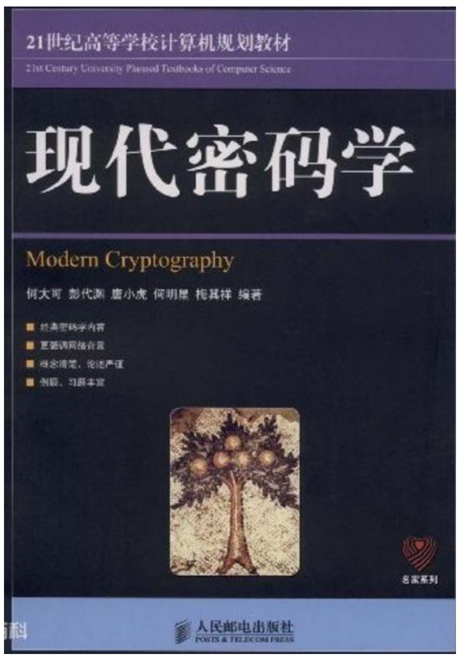
- **教材**

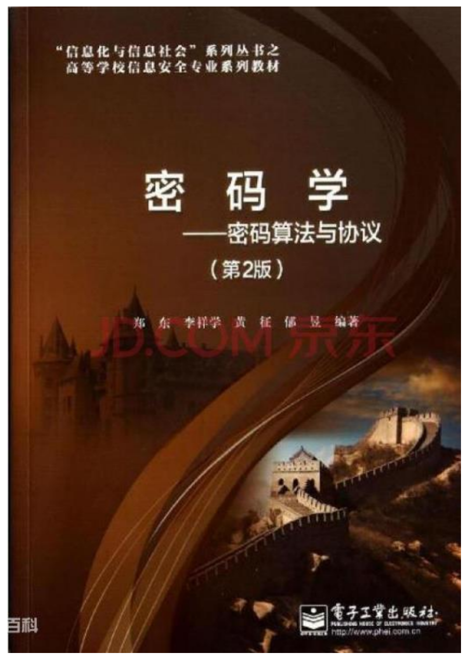
  -何大可，彭代渊，唐小虎等编著，现代密码学

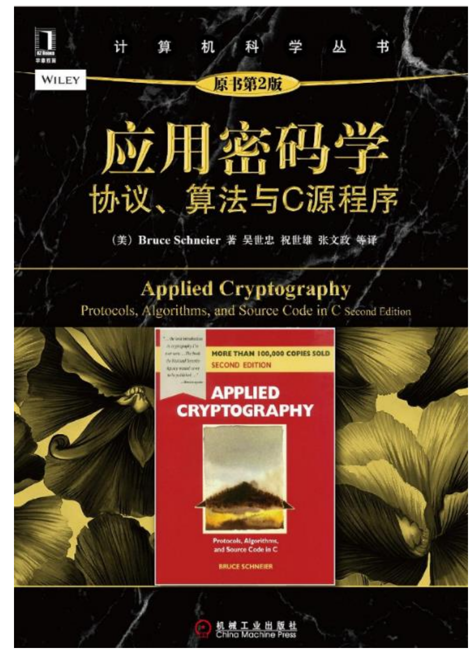
- **主要参考书**

  -郑东等人编著，密码学

  -Brusce Schneier 著，吴世忠，张文政 等译，应用密码学







- **授课时间**

  -按照教务处导出课表

- **答疑**

  -开学之前：QQ在线答疑

  -**开学之后：文理楼488**

主要内容

- 第1章 概论
- 第2章 流密码
- 第3章 分组密码
- 第4章 公钥密码
- 第5章 Hash函数与消息认证
- 第6章 数字签名
- 第7章* 密码协议
- 第8章* 密钥管理

第1章 概论

- 1.1 信息系统安全与密码技术
- 1.2 密码系统模型和密码体制
- 1.3 几种简单密码体制
- 1.4 初等密码分析
- 1.5* 密码学的信息论基础
- 1.6 密码学的复杂性理论基础

1.1 信息系统安全与密码技术

- **信息时代**
  - **农业革命⇒工业革命⇒信息革命**
  - **20世纪80年代美国Toffler A. 著《第三次浪潮》，预言：**

    计算机网络的建立与普及将彻底改变人类的生存和生活模式
- **信息、资源、能源是人类生存的三大支柱**

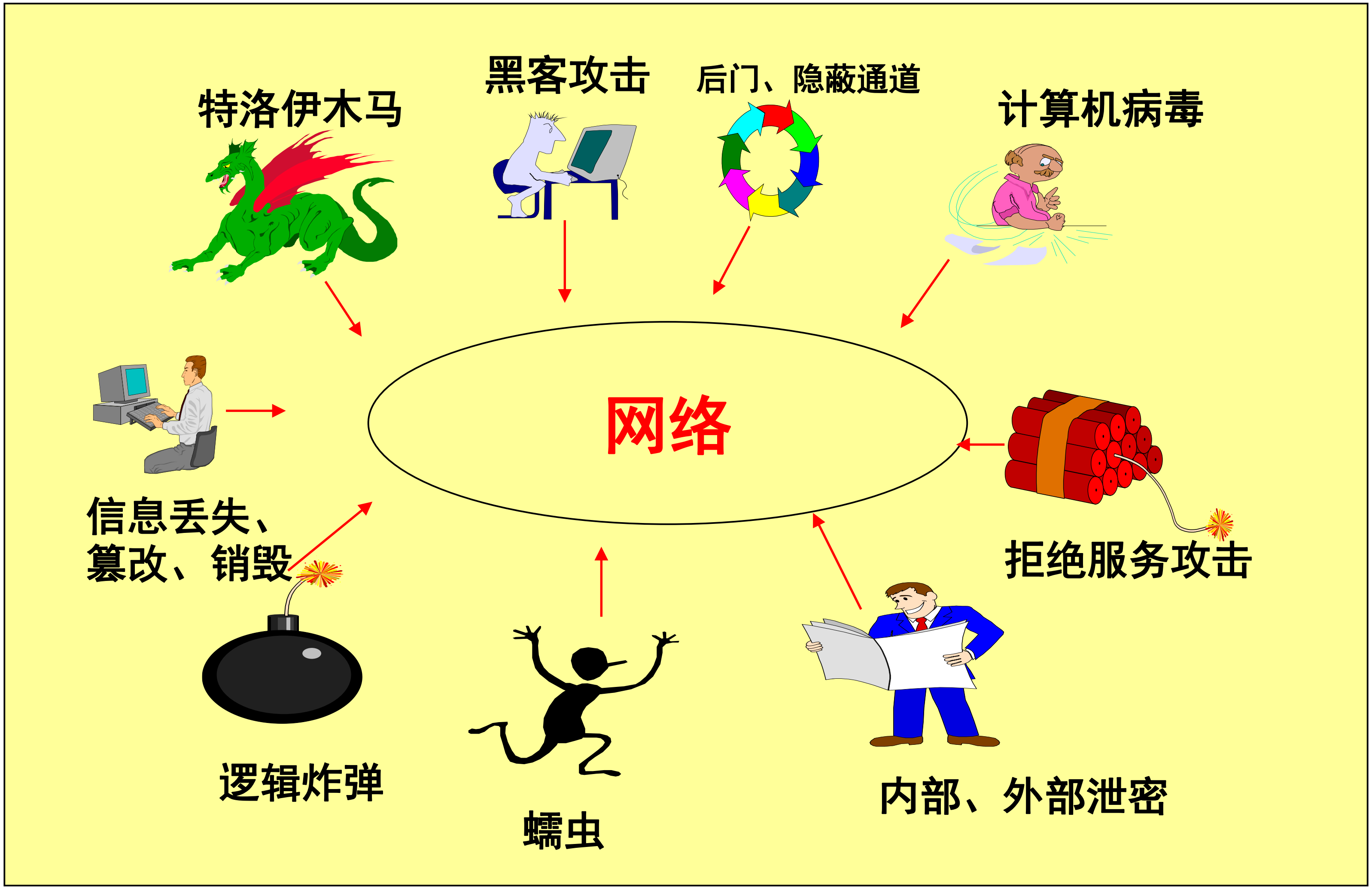
材料

1.1 信息系统安全与密码技术

- 民用
  - Internet普及
  - 电子政务
  - 电子商务
  - 电子金融
- 军事
  - 第三次军事革命
  - 信息战
  - 网络战

1.1 信息系统安全与密码技术

- 信息时代的安全威胁



1.1 信息系统安全与密码技术

- **安全威胁特点**
  - **网络遭受攻击数量与日俱增**
  - **网络病毒在全球范围内高速扩散**
  - **网络垃圾邮件成为新的焦点**
  - **网络犯罪（经济、政治）触目惊心**

1.1 信息系统安全与密码技术

- **安全威胁的危害**
  - **使用Internet的困扰**
  - **经济巨大损失**
  - **国家安全受到威胁**

- **信息安全技术的落后严重阻碍了社会的发展！**

1.1 信息系统安全与密码技术

- 各国政府的对策
  - 2002年11月27日美国总统签署《网络安全法案》。在接下来的几年中政府为大学拨款9亿美元，用于成立计算机安全中心，招收研究生进行安全研究。
  - 我国
    - 党的十五届五中全会明确指出，大力推进国民经济和社会信息化是覆盖现代化建设全局的战略举措。要以信息化带动工业化。
    - 在《科技教育发展“十五”重点专项规划（高技术产业发展规划）》中明确提出攻克信息保护、隐患发现、安全反应等关键技术，为国家信息基础建设提供技术支撑。
    - 1999年国务院颁布商用密码管理条例，对密码的管理使用进行了具体规定。

1.1 信息系统安全与密码技术

- **信息安全的基本属性**

  信息安全（information Security）、数据安全（data Security）

  - 机密性（confidentiality）
  - 完整性（integrity）
  - 认证性（nonrepudiation）——也称“不可否认性”或“抗抵赖”
  - 可用性（availability）
  - 公平性（fairness）
  - 可控性（controllability）

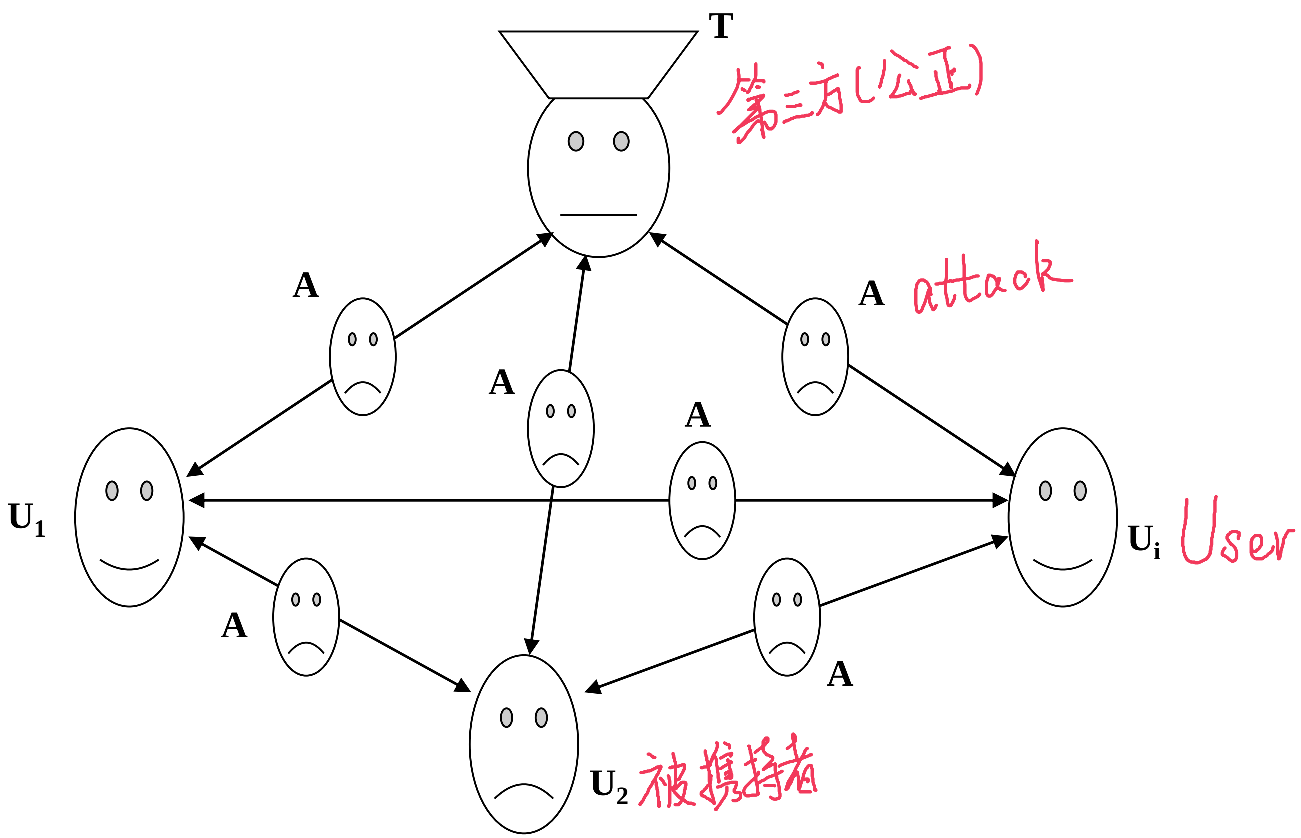
1.1 信息系统安全与密码技术

- 对信息安全的威胁或攻击: 对信息属性的侵害
  - 属于人为故意的威胁或攻击中，窃取、破译是对机密性的侵害
  - 篡改是对完整性的侵害
  - 伪造、重放是对认证性的侵害
  - 干扰、占用、资源耗尽以至摧毁信息处理器或载体是对可用性的侵害
  - 在电子媒体商品的网上交易中，获得商品后不按时付款或者收取货款后不按时提供商品，是对公平性的侵害

侵害对象	威胁或攻击手段	案例
机密性	入侵系统取得高级授权	猜测口令，系统漏洞，安置木马
	破译密码	穷举法搜索DES密钥，对密码器件的边信道攻击（side-channel attack）
完整性	插入、删除、篡改	用原消息m的MD5碰撞m′取代m
可用性	信道干扰	无线干扰
	摧毁系统硬件	微波炸弹，处理器内潜藏破坏性指令，嗜晶片微生物
	扰乱以至摧毁系统软件	计算机病毒
	用户恶意占用	内部用户资源占用、资源耗尽
	业务拒绝	“轰炸”端口
认证性	发送方身份假冒，接收抵赖	中间人攻击
	破坏收发审计记录	删除或篡改设备运行日志
公平性	非对等的密钥协商	单方面控制生成密钥参数
	利用非公平交易协议获取利益	中途终止不满足“公平性”的交易协议
可控性	破坏“密钥托管”，阻止“匿名撤消”	攻击相关协议，被收买或不诚实的托管人
	抗内容检测过滤的“穿透”	采用“特征字”变异技术
	阻止司法取证	破坏记录设备或介质，删除设备运行日志

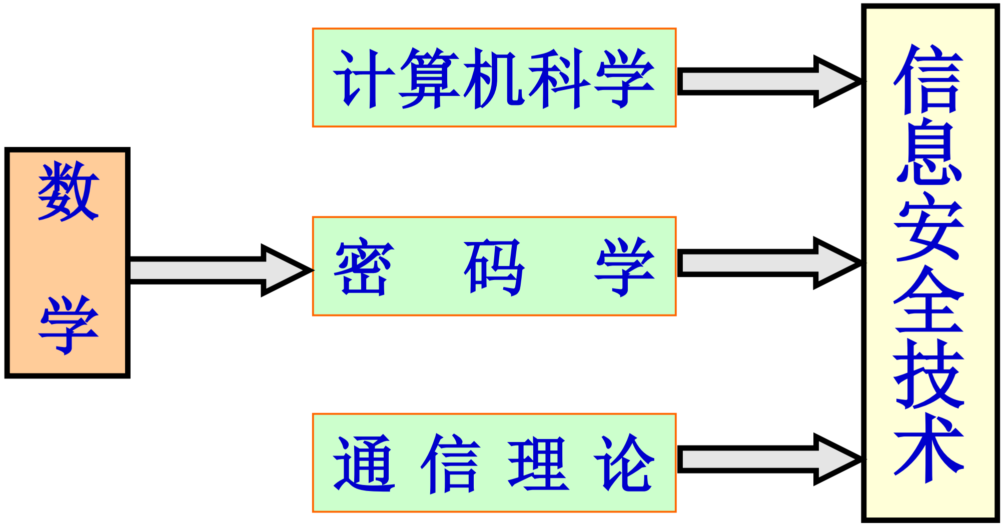
1.1 信息系统安全与密码技术

- **网络通信系统安全性分析构架**



1.1 信息系统安全与密码技术

- **信息安全关键技术 ——密码学理论与技术**

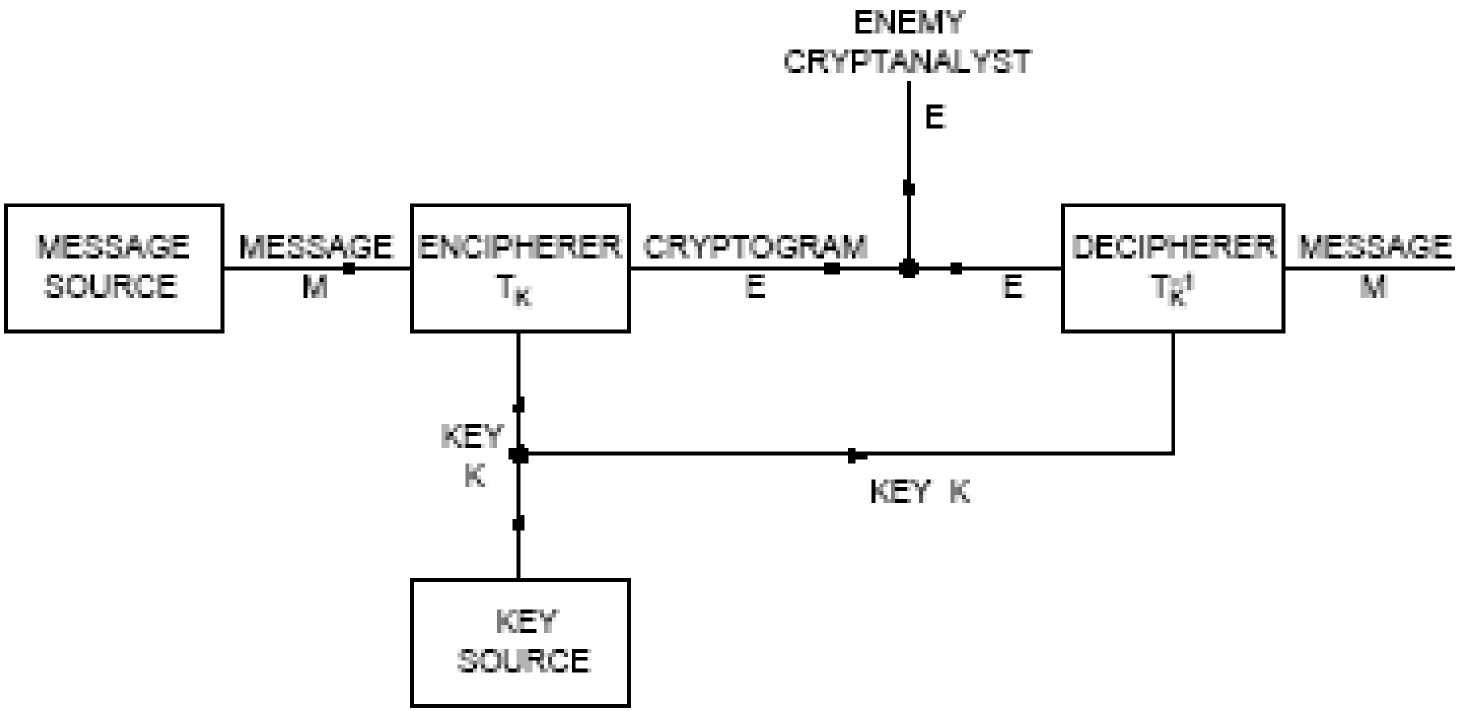


第1章 概论

- 1.1 信息系统安全与密码技术
- 1.2 密码系统模型和密码体制
- 1.3 几种简单密码体制
- 1.4 初等密码分析
- 1.5 密码学的信息论基础
- 1.6密码学的复杂性理论基础

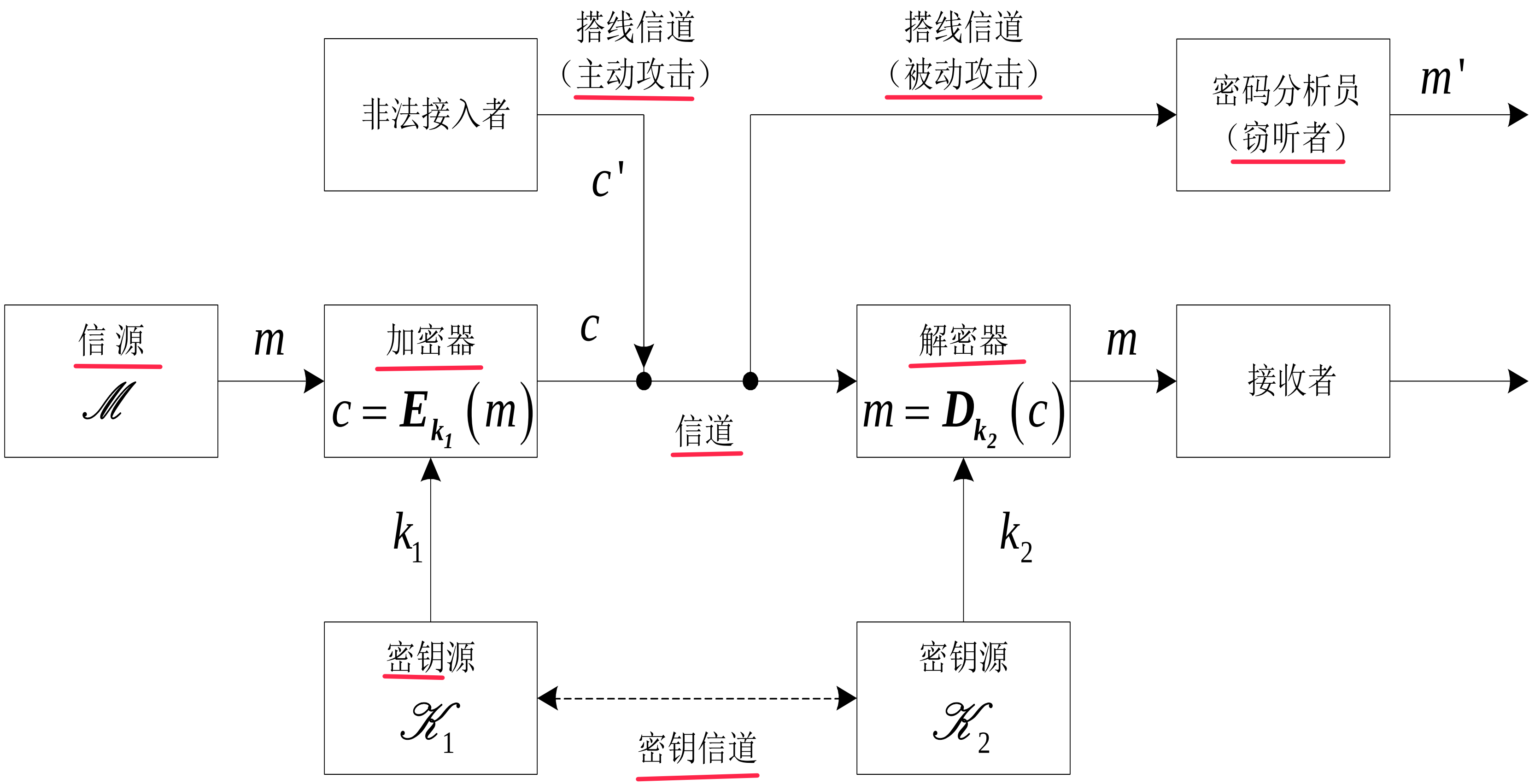
1.2 密码系统模型和密码体制

- **Shannon的保密系统模型**



1.2 密码系统模型和密码体制

- **现代密码系统模型**



1.2 密码系统模型和密码体制

- **密码体制(Cryptosystem)**

  —— **六元组** $(\mathscr{M}, \mathscr{C}, \mathscr{K}_1, \mathscr{K}_2, \mathscr{E}, \mathscr{D})$

  - **明文空间**: $\mathscr{M}$，$m\in\mathscr{M}$**称为明文(plaintext)**
  - **密文空间**: $\mathscr{C}$，$c\in\mathscr{C}$**称为密文(ciphertext)**
  - **加密密钥空间**: $\mathscr{K}_1$，

    $k_1\in\mathscr{K}_1$**称为加密密钥(encryption key)**
  - **解密密钥空间**: $\mathscr{K}_2$

    $k_2\in\mathscr{K}_2$**称为解密密钥(decryption key)**

  密钥(key): $k=(k_1, k_2)$

1.2 密码系统模型和密码体制

- **密码体制(Cryptosystem)**

  —— 六元组 $(\mathscr{M}, \mathscr{C}, \mathscr{K}_1, \mathscr{K}_2, \mathscr{E}, \mathscr{D})$

  - **加密变换簇: $\mathscr{E}$**

    $$\forall k_1 \in \mathscr{K}_1, \exists E_{k_1} \in \mathscr{E},$$

    加密变换（映射、函数、算法）$E_{k_1} : \mathscr{M} \to \mathscr{C}$

    (encryption map, function, algorithm)

  - **解密变换簇: $\mathscr{D}$**

    $$\forall k_2 \in \mathscr{K}_2, \exists D_{k_2} \in \mathscr{D},$$

    解密变换（映射、函数、算法）$D_{k_2} : \mathscr{C} \to \mathscr{M}$

    (decryption map, function, algorithm)

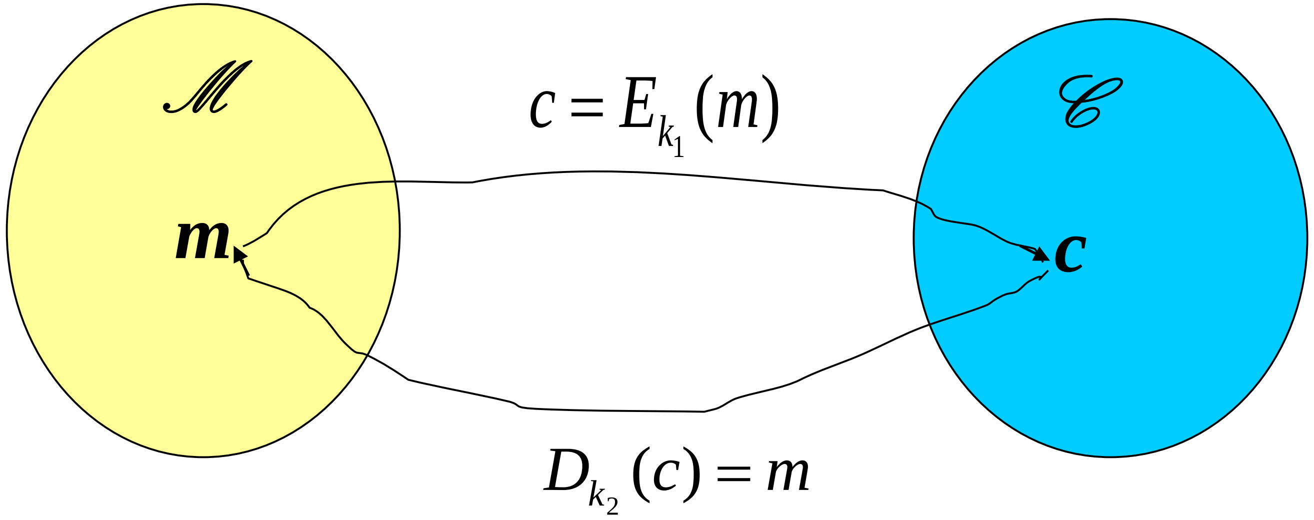
1.2 密码系统模型和密码体制

- **密码体制(Cryptosystem)**
  - **加密变换与解密变换的关系**

$\forall k=(k_1,k_2)\in \mathscr{K}_1\times\mathscr{K}_2, \exists E_{k_1}\in\mathscr{E}, D_{k_2}\in\mathscr{D}$,满足

$\forall m\in\mathscr{M}$ 有： $D_{k_2}(E_{k_1}(m))=m.$

D_{k_2}称为E_{k_1}的左逆变换.

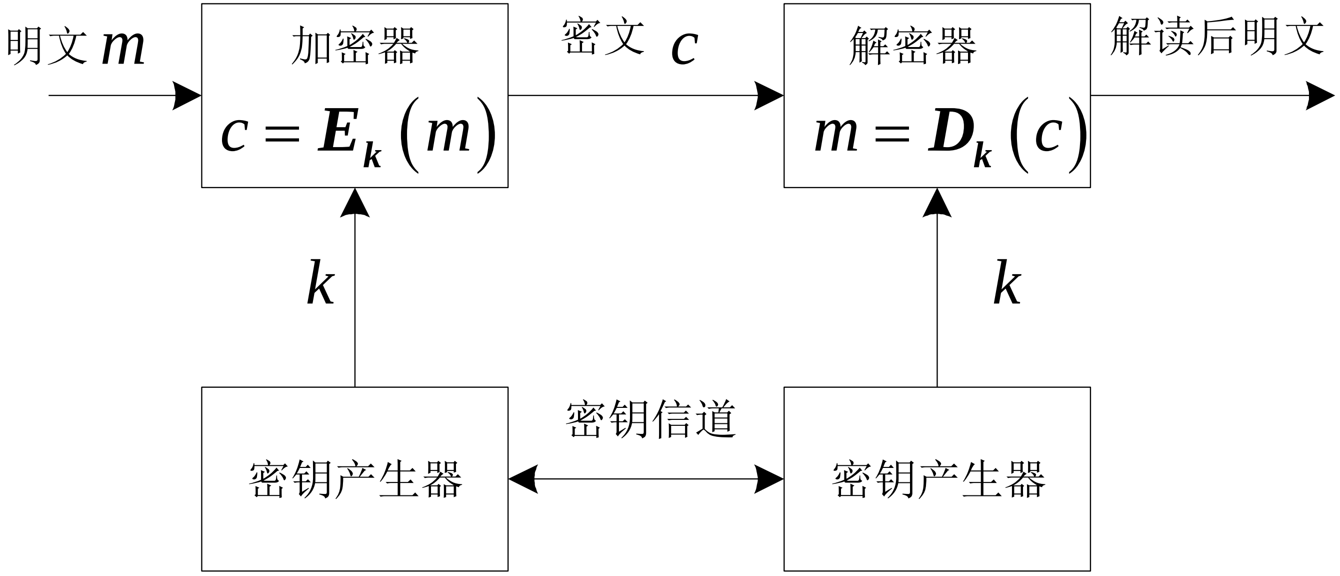


1.2 密码系统模型和密码体制

- **密码体制系统的分类**
  - **对称密码体制**(symmetric cryptosystem)

    $k=k_1=k_2$ **或** $k_1 \Leftrightarrow k_2$

    单钥、私钥 (one-key, private key)**密码体制**



1.2 密码系统模型和密码体制

- **密码体制系统的分类**
  - **对称密码体制(symmetric cryptosystem)**
    - **分组密码(block cipher)**

      将明文消息分为包含若干个符号的组，在选定密钥后使用固定的加密变换对明文分组逐组地进行加密。

      例如， DES(1977), AES(2001)
    - **流密码(stream cipher)**

      明文： $m=m_1m_2m_3.....$

      密钥： $k=k_1k_2k_3....$

      加密： $c_1=E_{k1}(m_1), c_2=E_{k2}(m_2), c_3=E_{k3}(m_3),....$

      密文： $c=c_1c_2c_3$ ....

      例如，GSM移动台（手机）到基站BS之间无线传输 中使用的加密算法A5/1是一种流密码

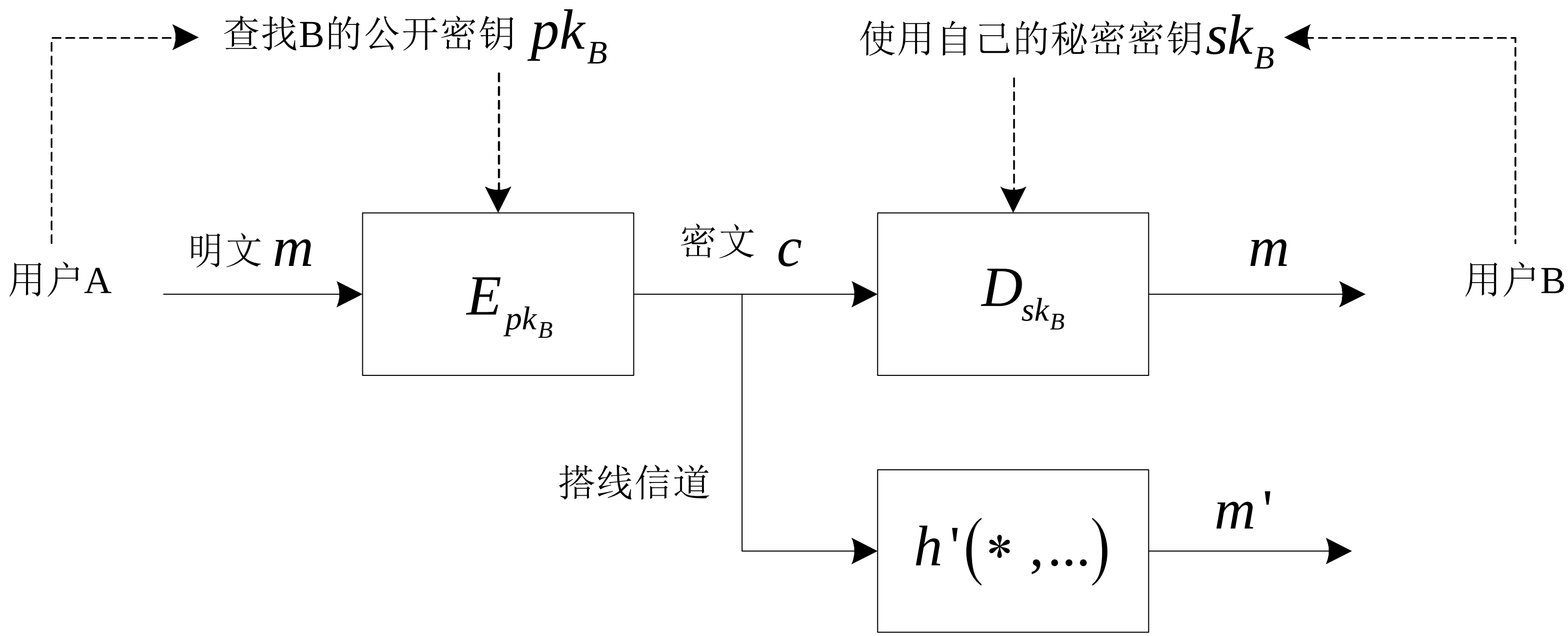
1.2 密码系统模型和密码体制

- **非对称密码体制(asymmetric cryptosystem)**

双钥、公钥 (two-key, public key)密码体制

$k_1 \neq k_2$ **或** k_1**不能**$\Leftrightarrow k_2$

公钥: $k_1=pk$；　**私钥**: $k_2=sk$



1.2 密码系统模型和密码体制

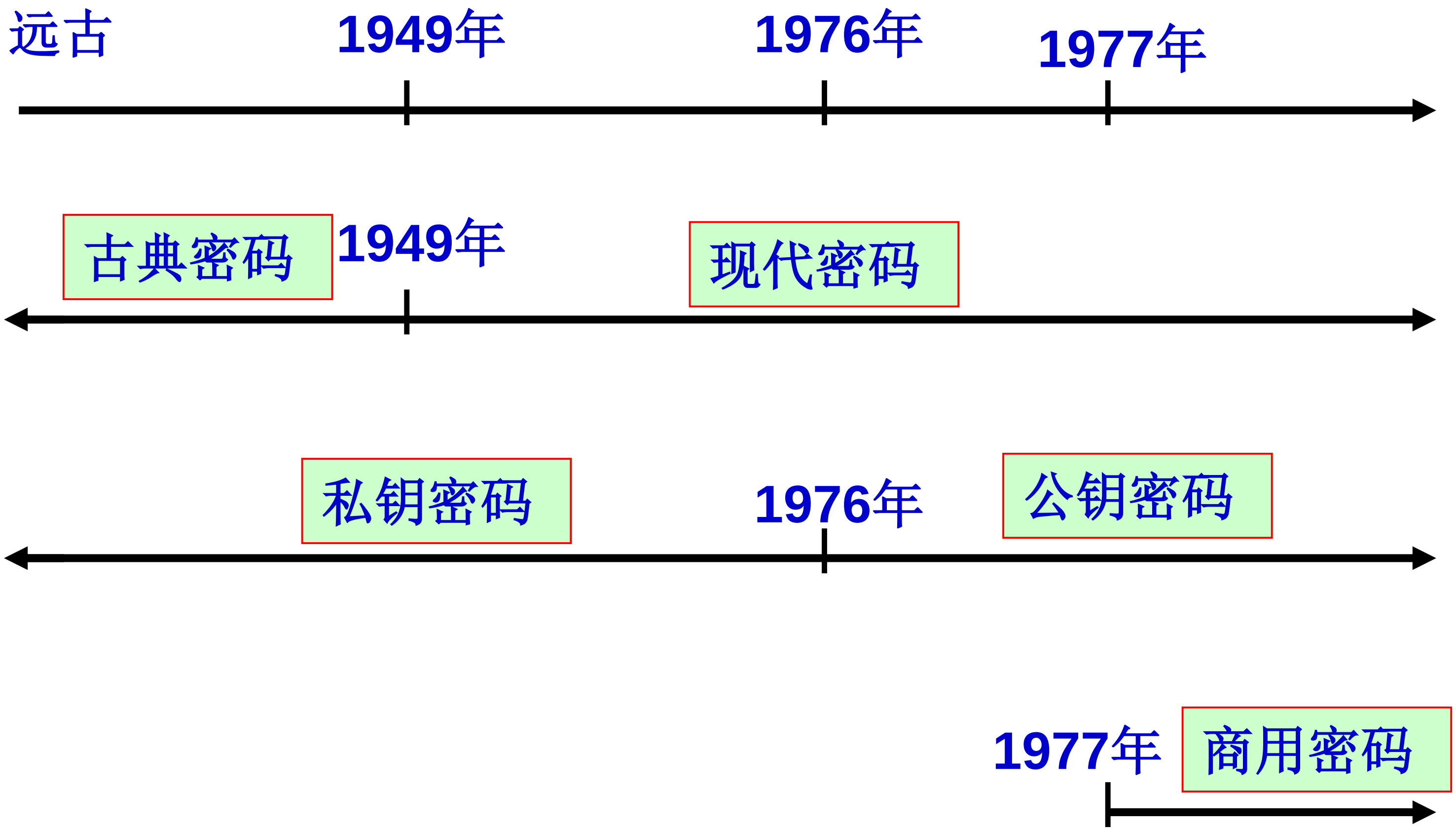
- **密码系统的设计原则**

设计加密函数与解密函数的学科称为密码编码学 (Cryptography)、密码学或保密学。

- **具有某种安全性**
  - **理论上不可破**
  - **实际上不可破**
- **Kerckhoff假设：系统的保密性不依赖于对加密体制或加（解）密算法的保密，而仅依赖于密钥的保密。**
- **加密和解密算法适用于密钥空间的全部元素**
- **系统便于实现和使用方便**

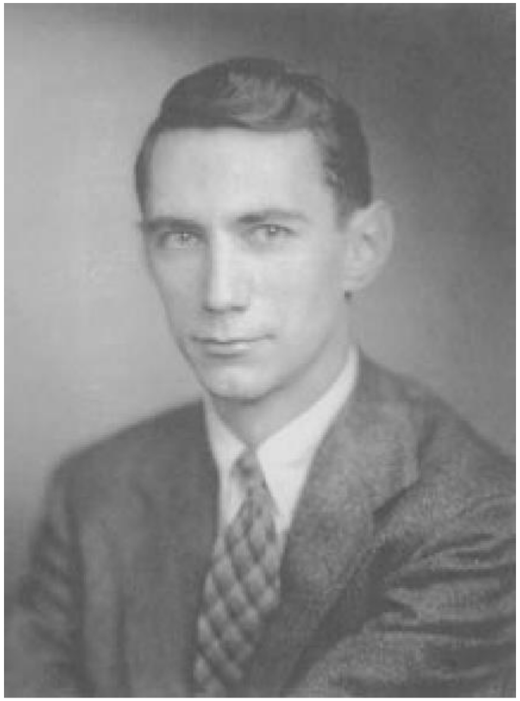
1.2 密码系统模型和密码体制

- **密码学发展简史**
  - **密码学发展史简图**



1.2 密码系统模型和密码体制

- 密码学发展简史
  - 古典密码时期(—1949)
    - 特定应用领域：军事、政治、外交
    - 神秘性
    - 艺术性
  - 现代密码学(1949—)：
    密码技术成为一门学科



著名论文:

Communication theory of secrecy systems,
Bell Syst. Tech. J., Volume 28, 656-715, 1949.

仙农(C. D. Shannon: 1916-2001)

1.2 密码系统模型和密码体制

- **密码学发展简史**
  - **公钥密码学(1976—)**
    - 计算机网络环境中的应用
    - W. Diffie和M. E. Hellman提出公钥密码的思想(1976)
    - 著名论文：W. Diffie and M. E. Hellman, New direction in cryptography, IEEE Tran. On Information Theory, IT-22, (6), 644-654,1976.
  - **密码学的商业应用 (1977—)**
    - 1977：美国国家标准局(National Bureau of Standards)颁布数据加密标准DES (Data Encryption Standard)
    - 1994：美国政府颁布数字签名标准DSS (Data Signature Standard)
    - 2001：美国政府颁布高级加密标准AES (Advanced Encryption Standard)

第1章 概论

- 1.1 信息系统安全与密码技术
- 1.2 密码系统模型和密码体制
- 1.3 几种简单密码体制
- 1.4 初等密码分析
- 1.5 密码学的信息论基础
- 1.6密码学的复杂性理论基础

1.3 几种简单密码体制

- 代换密码(substitution cipher) 的形象化实现过程
  - 明文空间: $\mathscr{M} = Z_q^L$ 长度为L
    明文字母表: $Z_q = \{0,1,\ldots,q-1\}$
    明文组(L-报文) : $m=(m_0, m_1,\ldots,m_{L-1}) \in Z_q^L$
  - 密文空间: $\mathscr{C} = Z_{q'}^{L'}$,
    密文字母表: $Z_{q'} = \{0,1,\ldots,q'-1\}$
    密文组 : $c=(c_0, c_1,\ldots,c_{L'-1}) \in Z_{q'}^{L'}$
  - 密钥空间: $\mathscr{K}$，$k \in \mathscr{K}$，$k=(k_e, k_d)$，包括加密和解密密钥
  - 加密变换: $f_{k_e} : Z_q^L \rightarrow Z_{q'}^{L'}$
    任给$m \in Z_q^L$，记$f_{k_e}(m) = f(k_e, m) = c$.
    设f是单射，则f的逆就是解密变换：$D_{k_d}(c) = f^{-1}(c) = m$

1.3 几种简单密码体制

- **代换密码(substitution cipher) 的细化分类**
  - $q=q'$，$L=L'=1$
    - ① **单表代换密码→古典密码**
      对所有的明文字母，都用一种固定的代换进行加密,即
      $$c_i = f_{ke}(m_i),\ i=0,1,\cdots$$
    - ② **多表代换密码→现代的流密码**
      对所有的明文字母，用一个以上的代换表进行加密，即
      $$[c_0, c_1, \cdots, c_{L-1}] = [f_{ke_0}(m_0), f_{ke_1}(m_1), \cdots, f_{ke_{L-1}}(m_{L-1})]$$ 多个一元函数
  - $q=q'$，$L, L' >= 2$，
    - ③ $L=L'$: f **可为** 1-1**映射；多字母代换密码→现代的分组密码**
      $$[c_0, c_1, \cdots, c_{L-1}] = f_{ke}[m_0, m_1, \cdots, m_{L-1}]$$ 多元函数
    - ④ $L<L'$: f **可为 1对多映射；某些抗量子密码的本质**
    - ⑤ $L>L'$: f **不可逆，不能作加密映射；现代的Hash函数**
      数字验证,防篡改

1.3 几种简单密码体制

一点说明：$Z_q=\{0,1,\cdots,q-1\}$ 是指 $\boxed{\text{模}q\text{剩余类环}}$。

这里的 Z_q 就是指离散数学中 N_q，也即模 q 剩余类环就是指离散数学中 $\langle N_q, +_q, \times_q \rangle$，下面将一直使用 Z_q，也一直默认 Z_q 是指模 q 剩余类环。

回顾：在离散数学中的代数系统 $\langle N_k, +_k, \times_k \rangle$ 是环，其中 $N_k=\{0,1,\cdots,k-1\}$，$+_k$ 和 $\times_k$ 是模 k 加法与乘法运算。

例如 $k=5$ 时，$N_k=\{0,1,2,3,4\}$，无零因子环。

$k=6$ 时，$N_6=\{0,1,2,3,4,5\}$，有零因子环

因为 $2\times_6 3=(2\cdot 3) \bmod 6=0$.

由于无零因子与消去律等价，$k=6$ 时有零因子说明集合 $N_6^*=\{1,2,3,4,5\}$ 关于 $\times_6$ 不能构成乘法群。

1.3 几种简单密码体制

现在考虑一个问题：Z_q 在满足什么条件时无零因子？

结论1.当q是素数(即质数)时 Z_q 是无零因子环，此时因为 Z_q 有单位元1，乘法满足交换律，因而 Z_q 是整环；进一步，由于有限的整环是域，所以当q是素数时，Z_q 是一个有限域！因而 Z_q 的所有非零元素组成的集合关于模q乘法构成群！

将 Z_q 的所有非零元构成的集合记为 Z_q^*，即 $Z_q^*=\{1,2,\cdots,q-1\}$，

即 $\langle Z_q^*, \times\rangle$ 是群。

例如，q=5时，所有非零元集合为 $\{1,2,3,4\}$，

由于 $\langle Z_5,+\rangle$ 是交换群，而 $\langle Z_5,+,\times\rangle$ 是交换环，单位元是1

只需要验证每个非零元关于乘法都有逆元，则可证明

$\{1,2,3,4\}$ 关于乘法构成群，从而 $\langle Z_5,+,\times\rangle$ 是有限域：

$1\times1=1$，$2\times3=6\equiv1 \bmod 5$，$4\times4=16\equiv1 \bmod 5$，

即每个元素可逆

1.3 几种简单密码体制

再如，q=7时，非零元集合为{1, 2, 3, 4, 5, 6}，

$1\times1=1$，$2\times4=8\equiv1 \bmod 7$，$3\times5=15\equiv1 \bmod 7$，$6\times6=36\equiv1 \bmod 7$

又如，q=11，非零元集合为{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}

$1\times1=1$，$2\times6=12\equiv1 \bmod 11$，$3\times4=12\equiv1 \bmod 11$，

$5\times9=45\equiv1 \bmod 11$，$7\times8=56\equiv1 \bmod 11$，$10\times10=100\equiv1 \bmod 11$

另例，q=13，非零元集合{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

$1\times1=1$，$2\times7=14\equiv1 \bmod 13$，$3\times9=27\equiv1 \bmod 13$，$4\times10=40\equiv1 \bmod 13$，

$5\times8=40\equiv1 \bmod 13$，$6\times11=66\equiv1 \bmod 13$，$12\times12=144\equiv1 \bmod 13$

1.3 几种简单密码体制

结论2. 当 $q>2$ 且 q 是合数时 Z_q 是有零因子环。

不妨设 q 的真因子分解可写为 $q=q_1q_2$，其中 $1<q_1<q$，$1<q_2<q$，则在 Z_q 中，q_1 与 q_2 均为 Z_q 上的零因子；进一步，在 Z_q 中，每一个满足条件 ① $a\in Z_q$ 且 $1<a<q$；② $\gcd(a,q)>1$ 的元素 a 均为 Z_q 的零因子。

例如，在 Z_6 中，2,3,4,都是零因子，而1,5不是零因子。

例如，$q=4$，与4不互素的元素集合为 $\{2\}$，$2\times2=4\equiv0 \bmod 4$.

又如，$q=6$，与6不互素的非零元集合为 $\{2,3,4\}$，

$2\times3=6\equiv0 \bmod 6$，$3\times4=12\equiv0 \bmod 6$

再如，$q=8$，与8不互素的非零元集合为 $\{2,4,6\}$，

$2\times4=8\equiv0 \bmod 8$，$4\times6=24\equiv0 \bmod 8$.

另外，$q=9$，与9互素的非零元集合为 $\{3,6\}$，$3\times6=18\equiv0 \bmod 9$.

1.3 几种简单密码体制

结论3. 当$q>2$且q是合数时，将Z_q中所有与q互素的元素组成的集合记为Z_q^*，则Z_q^*中的每个元素都不是零因子，且Z_q^*关于模q乘法构成乘法群，即Z_q^*中的每个元素都有乘法逆！

例如，$q=4$，与4互素的元素集合为$\{1,3\}$，满足：

①乘法封闭；②满足结合律；③ 1是单位元；④ $3^{-1}=3$，$1^{-1}=1$；

又如，$q=6$，$Z_6^*=\{1,5\}$，满足群的4条件性质。

再如，$q=8$，$Z_8^*=\{1,3,5,7\}$，下面验证群的4条性质：

① $1\times1=1$，$1\times3=3$，$1\times5=5$，$1\times7=7$，$3\times3=9\equiv1 \bmod 8$，$3\times5=15\equiv7 \bmod 8$，

$3\times7=21\equiv5 \bmod 8$，$5\times5=25\equiv1 \bmod 8$，$5\times7=35\equiv3$，$7\times7=49\equiv1 \bmod 8$.

②结合律是环$\langle Z_8,+,\times\rangle$的性质，成立.

③ 1是单位元；④由第①可知，每个元素的乘法逆是自己。

总结论：设$q>2$是正整数，Z_q^*是所有小于q且与q互素的正整数的集合，则$\langle Z_q^*,\times\rangle$构成乘法群。

1.3 几种简单密码体制

定义1. 对于一个大于1的正整数q，称集合Z_q^*的元素个数为q的欧拉函数值，即为$\varphi(q)$，也就是：

$\varphi(q)=|Z_q^*|$.

欧拉函数有两个很重要的性质：

(1) 设q是素数，m是正整数，则$\varphi(q^m)=q^{m-1}(q-1)$；

(2) 设m, n是两个互素的正整数，则$\varphi(m\cdot n)=\varphi(m)\cdot\varphi(n)$

由以上两个性质，可以通过因数分解的方式得到所有正整数的欧拉函数值：

设q的因式分解为$q=q_1^{m_1}q_2^{m_2}\cdots q_t^{m_t}$，其中$q_1,\cdots,q_t$是两两不同的素数，则由上面两个性质可知

$$\begin{aligned}\varphi(q)&=\varphi(q_1^{m_1})\cdot\varphi(q_2^{m_2})\cdots\varphi(q_t^{m_t})\\&=q_1^{m_1-1}(q_1-1)q_2^{m_2-1}(q_2-1)\cdots q_t^{m_t-1}(q_t-1)\end{aligned}$$

1.3 几种简单密码体制

刚才提到一个问题，即 Z_q^* 中的每个元素都存在乘法逆元，如何求逆元呢？这是现代密码学中经常遇到的一个问题，因为本章中的乘法逆元可以通过依依尝试的办法能很快确定，因此暂时不补充一般求解方法，但后面会补充非常有用的欧几里德算法，又叫辗转相除法。

现在，给出几类简单密码体制。

1.3 几种简单密码体制

- **单表代换案例1——仿射密码（affine cipher）**
  - ◆ $\mathscr{M}=\mathscr{C}=Z_q$
  - ◆ $\mathscr{K}=\{(a,b)\ |(a,b)\in Z_q\times Z_q,\ \gcd(b,q)=1\}$
  - ◆ **加密变换**

    $C = E_k(m)=a+bm \mod q \ \ (m\in Z_q)$
  - ◆ **解密变换**

    $m = D_k(c)=(c-a)b^{-1} \mod q \ \ (c\in Z_q)$
  - ◆ **密钥量太小，不安全！**

    设$q=26$，$|\mathscr{K}|=26\times\varphi(26)=26\times12=312.$

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

1.3 几种简单密码体制

- $k=(0, 1)$时，$E_k(m)=a+bm=m$ 为恒等变换
- $a=0$时

$E_k(m)=bm$ 为乘数密码（multiplicative cipher）

- $b=1$时

$E_k(m)=a+m$ 为移位代换密码（shift substitution cipher）或加法密码（additive cipher）

  - $E_k(m)=3+m$ 为凯撒密码（Caesar Cipher）
  - 全部26个加密变换的代换表合在一起构成一个形如26阶对称矩阵的表(该矩阵的第一行、第一列是从A到Z顺序排列的字母), 称为维吉尼亚表(Vigenere Table)

1.3 几种简单密码体制

◆例1.3.1 设密钥k=(3, 7)，则加密变换

$$E_k(m)=3+7m \mod 26 \quad (m\in Z_{26}),$$

明文字母m_0若为英文I，而I对应数字8，

故加密变换 $E_k(8)=3+7\times 8=59=7 \mod 26$，即密文$c_0=H$

解密变换： $D_k(c)=(c-a)b^{-1}=(c-3)\times 7^{-1} \mod 26.$

因为$7\times 15=105=1 \mod 26$, 所以$7^{-1}=15 \mod 26$,

$D_k(c)=(c-3)\times 15=15c-45=15c+7 \mod 26.$

$D_k(7)=15\times 7+7=285+7=25+7=8 \mod 26,$

即恢复明文$m_0=I.$

注：实际上，还可以用多项式函数代替上述的一次函数来作为加密变换，不再赘述！

1.3 几种简单密码体制

- **单表代换案例2——密钥短语密码（key phrase cipher）**
  - ◆选择一个英文短语作为密钥字（key word）或称密钥短语（key phrase），如HAPPY NEW YEAR，去掉重复的字母得到HAPYNEWR。将它依次写在明文字母表的下面，而后再将字母表中未在短语中出现过的字母依次写在这个短语后面，就可以构造一个代换表，如下所示：

$\mathscr{A}$:	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
$\mathscr{A}'$:	H	A	P	Y	N	E	W	R	B	C	D	F	G	I	J	K	L	M	O	Q	S	T	U	V	X	Z

  - ◆**是26个英文字母的一个置换！**

例如：明文是句子"the crypto system is secure."

则密文是字符串"QRN PMXKQJ OXOQNG BO ONPSMN"

1.3 几种简单密码体制

- **多表代换密码(polyalphabetic substitute) 思想**
  - 以两个或两个以上代换表依次对明文消息的字母进行代换的加密方法

    设明文字母表: Z_q ,

    代换序列: $\pi=(\pi_0, \pi_1,\ldots)$

    明文序列: $m=(m_0, m_1,\ldots)$

    密文序列: $c=E_k(m)=E_\pi(m)= \pi_0(m_0)\pi_1(m_1)\ldots$
  - 若π是非周期的无限序列，则称为非周期多表代换密码，或一次一密密码（one-time pad cipher）。
  - 周期多表代换密码：

    代换序列: $\pi=(\pi_0, \pi_1,\ldots,\pi_{d-1}, \pi_0, \pi_1,\ldots,\pi_{d-1},\ldots)$

    密文序列:

    $c=\pi_0(m_0)\pi_1(m_1)\ldots\pi_{d-1}(m_{d-1})\ \pi_0(m_d)\pi_1(m_{d+1})\ldots\pi_{d-1}(m_{2d-1})\ldots$

1.3 几种简单密码体制

- **多表代换案例1——维吉尼亚密码（Vigenere Cipher,1858）**

  设密钥$k=(k_0, k_1,\ldots,k_{d-1})\in Z_q^d$,

  代换序列: $\pi=(\pi_0, \pi_1,\ldots,\pi_{d-1})$,

  $$\pi_i(x)=x+k_i \mod q.$$

  明文序列: $m=(m_0, m_1,\ldots)$

  密文序列: $c=(m_0+k_0)(m_1+k_1)\ldots(m_{d-1}+k_{d-1})$

  $$(m_d+k_0)(m_{d+1}+k_1)\ldots\ldots$$

1.3 几种简单密码体制

◆**例1.3.2:** **设**q=26, d=6, k=CIPHER=(2,8,15,7,4,17)

明文: m=this cryptosystem is not secure

密文: c=VPXZGI AXIVWP UBTTMJ PWIZIT WZT

明文	t	h	i	s	c	r	y	p	t	o	s	y
	19	7	8	18	2	17	24	15	19	14	18	24
密钥	2	8	15	7	4	17	2	8	15	7	4	17
秘文	21	15	23	25	6	8	0	23	8	21	22	15
	V	P	X	Z	G	I	A	X	I	V	W	P

s	t	e	m	i	s	n	o	t	s	e	c	u	r	e
18	19	4	12	8	18	13	14	19	18	4	2	20	17	4
2	8	15	7	4	17	2	8	15	7	4	17	2	8	15
20	1	19	19	12	9	15	22	8	25	8	19	22	25	19
U	B	T	T	M	J	P	W	I	Z	I	T	W	Z	T

1.3 几种简单密码体制

◆ **多表代换案例2——博福特密码（Beaufort Cipher）**

设密钥$k=(k_0, k_1,\ldots,k_{d-1})\in Z_q^{\,d}$,

代换序列: $\pi=(\pi_0, \pi_1,\ldots,\pi_{d-1})$, $\pi_i(x)= k_i-x \mod q$.

明文序列: $m=(m_0, m_1,\ldots)$

$$c=(k_0-m_0)(k_1-m_1)\ldots(k_{d-1}-m_{d-1})(k_0-m_d)(k_1-m_{d+1})\ldots\ldots$$

1.3 几种简单密码体制

- **多字母代换密码思想**

多字母代换密码是字母L维向量空间到自身的一个可逆映射

f：$Z_q^L \rightarrow Z_q^L$；**即**

$$f(m_0m_1 \ldots m_{L-1})=c_0c_1 \ldots c_{L-1}。$$

令明文$m=m_0m_1\ldots$，**则相应密文为**

$$c=c_0c_1\ldots$$

$$=f(m_0m_1\ldots m_{L-1})f(m_Lm_{L+1}\ldots m_{2L-1})\ldots$$

1.3 几种简单密码体制

- 多字母代换案例1——Hill密码

基于矩阵的线性变换:

Z_{26}为模26的同余类集合, K是一个L*L矩阵,在Z_{26}上可逆,

即存在K^{-1}使得: $KK^{-1} = I$ (在Z_{26}上)

注：明文与密文都是L维的向量

$$m = (m_1, m_2 \ldots, m_L); \ c = (c_1, c_2, \ldots, c_L);$$

加密：$c = mK \bmod 26$;

解密：$m = cK^{-1} \bmod 26$;

1.3 几种简单密码体制

练习：(1) 假设Hill密码加密使用密钥 $K=\begin{bmatrix}4 & 9\\ 3 & 7\end{bmatrix}$

试对明文**best**加密。

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

密文：QLZJ

(2) 假设**Hill**密码加密同样使用上面的密钥，试对密文

HTVNSB解密。 $K^{-1}=\begin{bmatrix}7 & 17\\ 23 & 4\end{bmatrix}$ 明文：pretty

1.3 几种简单密码体制

- **综合案例1——换位密码（transposition cipher）**
  - ◆**明文：** ***m*= the simplest possible transposition ciphers**

    分成长度为5的组：

    ***m*= thesi | mples | tposs | iblet | ransp | ositi | oncip | hersx**
  - ◆**加密变换：将各组内字符按位置下标号（0~4）实施下述置换（permutation）**

$$E_k = \begin{pmatrix} 01234 \\ 30421 \end{pmatrix}$$

单表 多字母

  - ◆**密文：**

    ***c*=STIEH EMSLP STSOP EITLB SRPNA TOIIS**

    IOPCN SHXRE

换位密码

- **综合案例1——换位密码（transposition cipher）**
  - ◆**E_k的逆置换：**

$$E_k = \begin{pmatrix} 01234 \\ 30421 \end{pmatrix} \Longrightarrow D_k = \begin{pmatrix} 01234 \\ 14302 \end{pmatrix}$$

◆**解密：**

密文： c=**STIEH EMSLP STSOP EITLB SRPNA TOIIS IOPCN SHXRE**

明文： m= **thesi mples tposs iblet ransp ositi oncip hers**<u>**x**</u>

◆**密钥量：** $|\mathscr{K}|=L!$

第1章 概论

- 1.1 信息系统安全与密码技术
- 1.2 密码系统模型和密码体制
- 1.3 几种简单密码体制
- 1.4 初等密码分析
- 1.5 密码学的信息论基础
- 1.6密码学的复杂性理论基础

1.4 初等密码分析

- **利用密文推断出明文或解密密钥的学科称为密码分析学(Cryptanalysis)**

分析(analysis)=破译(break)=攻击(attacks)

- **攻击方式**
  - **主动攻击(active attack)**

    窜改通信中的数据流，或在通信中产生虚假数据流
  - **被动攻击(passive attack)**

    窃听或监视通信过程，从中获得信息

1.4 初等密码分析

- **攻击方法**

攻击方法
- 传统破译法
  - 穷举破译法
  - 数学分析法
    - 确定性分析法
    - 统计分析法
- 物理破译法
  - 时序分析
  - 能量分析

攻击方法

- **穷举破译法(exhaustive attack method)**
  - 方法：对截获的密文依次用各种可能的密钥试译，直到获得有意义的明文；或者利用对手已注入密钥的加密机（比如缴获得到），对所有可能的明文依次加密直到得出与截获的密文一致的密文。
  - 对策：将密钥空间和明文空间设计得足够大。
- **确定性分析法**
  - 方法：利用密文或者明文—密文对等已知量以数学关系式表示出所求未知量（如密钥等），然后计算出未知量。
  - 对策：设计具有坚实数学基础和足够复杂的加密函数
- **统计分析法**
  - 方法：密码破译者对截获的密文进行统计分析，找出其统计规律或特征，并与明文空间的统计特征进行对照比较，从中提取出密文与明文间的对应关系，最终确定密钥或明文。
  - 对策：扰乱密文的语言统计规律

攻击方法

- **物理破译方法(Kocher, 1996)**

  利用加密执行时的物理现象来确定密钥的密码分析方法，也被称为“边信道攻击”（side-channel attack）。

  所利用的物理现象有密码算法执行器件（加密芯片）的功耗，各算法步执行时间度量，甚至主机执行加密任务时主板上电容器发出的声音等等。

- **攻击类型**
  - **唯密文攻击(ciphertext-only attack)**

    密码分析者仅知道有限数量用同一个密钥加密的密文
  - **已知明文攻击(known plaintext attack)**

    密码分析者除了拥有有限数量的密文外，还有数量限定的一些已知“明文—密文”对
  - **选择明文攻击(chosen plaintext attack)**

    密码分析者除了拥有有限数量的密文外，还有机会使用注入了未知密钥的加密机，通过自由选择明文来获取所希望的“明文—密文”对
  - **选择密文攻击(chosen ciphertext attack)**

    密码分析者除了拥有有限数量的密文外，还有机会使用注入了未知密钥的解密机，通过自由选择密文来获取所希望的“密文—明文”对

1.4 初等密码分析

- **语言的统计规律性**
  - C.E.Shannon 1949年第一次透彻地阐明了密码分析的真谛，指出密码能够被破译的最根本原因是由于明文空间非均匀的统计特性
  - **英文字母频率表**

字母	概率	字母	概率	字母	概率	字母	概率
A	0.08167	H	0.06094	O	0.075	V	0.010
B	0.01492	I	0.06966	P	0.019	W	0.023
C	0.02782	J	*0.00153*	Q	*0.001*	X	*0.001*
D	0.04253	K	0.008	R	0.060	Y	0.020
E	0.12702	L	0.040	S	0.063	Z	*0.001*
F	0.02228	M	0.024	T	0.091		
G	0.02015	N	0.067	U	0.028		

1.4 初等密码分析

- **单表代换密码分析**

给定密文： EJM H AFJPD PBKRNM QJ AN ONPSMN BQO AFJPD NIW QRG GSOQ AN FHMWN NIJSWR QJ YNQNM OQHQBOQB PHF HIH FXOBOQRN AFJPD FNIWQR JE YNO GHX AN QRN ORJMQNOQJIN

$$\frac{17}{11+16+22+24+30+8+73}$$

◆ **首先求出密文字母的频度分布表**

密文字母	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
频数	6	5	0	3	2	8	3	7	5	10	1	0	6	17	10	6	14	7	3	0	0	0	4	2	2	0

◆ **先猜测**E_k: e→N,，**再利用英文高频度的双字母知识，比对密文中出现的双字母（词）及次高频度字母集** {t,a,o,i,n,s,h,r}，**猜测**E_k: t→Q, o→J, **得到部分试译结果：**

EoM H bFoPD PBKReM to be OePSMe BtO bFoPD FeIWtR G GSOt be FHMWe eIoSWR to YeteM OtHtBOtBPHF HIHFXOBO tRe bFoPD FeIWtR oE YeO GHX be tRe ORoMteOt oIe

单表代换密码分析(2)

- 利用三字母组合和元音辅音拼写知识，猜测单词tRe为the，即 E_k: h→R.
- 在余下的次高频度字母集{a,i,n,s,r}中选择适当的辅音字母作为密文O的原像, 即E_k: s→O.
- 利用构词分析从shoMtest猜测E_k: r→M,从oIe猜测E_k: n→I. 试译：

  Eor H bFoPD PBKher to be sePSre Bts bFoPD FenWth G GSst be FHrWe enoSWh to Yeter stHtBstBPHF HnHFXsBs the bFoPD FenWth oE Yes GHX be the shortest one

- 对余下的次高频度字母集{a,i}, 猜测 E_k: a→H. 试译：

  Eor a bFoPD PBKher to be sePSre Bts bFoPD FenWth G GSst be FarWe enoSWh to Yeter statBstBPaF anaFXsBs the bFoPD FenWth oE Yes GaX be the shortest one

单表代换密码分析(3)

- 对余下的中高频度字母F, 其原像应当是辅音, 故在中频度字母集{d,l}中猜测E_k: l→F. 利用构词分析从lenWth猜测E_k:g→W.试译：

  Eor a bloPD PBKher to be sePSre Bts bloPD length G GSst be large enoSgh to Yeter statBstBPal analXsBs the bloPD length oE Yes GaX be the shortest one

- 再猜测余下的映射

  E_k: f→E, c→P, k→D, m→G, u→S, i→B, p→K, d→Y, y→X.

  最后得明文：

  for a block cipher to be secure its block length m must be large enough to deter statistical analysis the block length of des may be the shortest one

单表代换密码分析(4)

◆**破译的代换表为：**

$\mathscr{A}$ ：　abcdefghijklmnopqrstuvwxyz

$\mathscr{A}'$：　HAPYNEWRB□DFGIJK□MOQS□□□X□

◆**由于密文中未出现C、L、T、U、V、Z，所以代换表尚不完整。如果有更多的密文供分析使用，就很有可能得到正确的代换表：**

$\mathscr{A}$ ：　abcdefghijklmnopqrstuvwxyz

$\mathscr{A}'$：　HAPYNEWRBCDFGIJKLMOQSTUVXZ

1.4 初等密码分析

- **对密码分析有用的英文统计特性**
  - **冠词the对统计特性影响极大，它使t、h、th、he和the在单、双和三字母统计中都为高频度元素。**
  - **英文中大约有一半的字以e、s、d和t作为字的结尾字母。**
  - **英文中大约有一半的字以t、a、s或w作为字的开头字母。**

第1章 概论

1.6 密码学的复杂性理论基础

- **1.6.1 问题与算法**
  - **问题（problem）**

    在计算机上求解的对象称为问题。

    问题的描述由两部分构成：（1）给定所有自由变量的一般性描述；（2）陈述“答案”或“解”必须满足的的性质。
  - **实例（instance）**

    如果给问题的所有自由变量都指定了具体的值，就得到该问题的一个实例。

1.6 密码学的复杂性理论基础

◆**例1.6.1 在二元域Z_2上求解下例布尔函数方程组**

$$\begin{cases} f_1(x_1,x_2,\ldots,x_n)=0 \\ f_2(x_1,x_2,\ldots,x_n)=0 \\ \qquad \cdots\cdots \\ f_m(x_1,x_2,\ldots,x_n)=0 \end{cases}$$

参数集合：$\{f_i(x_1,x_2,\ldots,x_n): i=1,2,\ldots m\}$

解的性质：$(u_1,u_2,\ldots,u_n)\in Z_2^n,\ f_i(u_1,u_2,\ldots,u_n)=0\ \ (i=1,2,\ldots m).$

[例] 取$n=m=3$时,

$$\begin{cases} f_1(x_1,x_2,x_3)=x_1+x_2x_3=0 \\ f_2(x_1,x_2,x_3)=1+x_1x_2=0 \\ f_3(x_1,x_2,\ldots,x_n)=1+x_1+x_2+x_3+x_1x_2x_3=0 \end{cases}$$

就是该问题的一个实例.

1.6 密码学的复杂性理论基础

- **1.6.1 问题与算法**
  - **算法（algorithm）**

    指求解某个问题的一系列具体步骤，并且要能在运行了有限时间（或运算步）之后给出“答案”，然后“停机”终止。

    如果算法*A*可以求解问题*Q*的任何一个实例，并且答案的正确率超过50%，那么就称算法*A*能解问题*Q*。通常按程序设计的习惯将给定的自由变量称为算法的“输入”，将“答案”或“解”称为算法的“输出”。
  - **问题的规模（size）**

    定义为求解该问题的算法所需输入数据的长度（比如bit数），通常用n表示.

1.6 密码学的复杂性理论基础

● 1.6.1 问题与算法

◆例1.6.2 欧几里德算法A（解“最大公约数”问题）

> **Step0.** 输入：正整数u, v，
>
> **Step1.** 若$v=0$，转Step3；
>
> **Step2.** $r \leftarrow u \bmod v$， $u \leftarrow v$, $v \leftarrow r$ 转Step1；
>
> **Step3.** 输出：u，停机。

□算法A停机后，变量单元u中的数值即最大公约数$\gcd(u, v)$。

假设初值满足$0<u, v \le N$，则“最大公约数”问题的规模（依惯例）为：

$$n=\lceil \log_2 N \rceil。$$

$\sqrt[l]{n}$

可以证明算法A所需执行的整数除法次数最多为：

$$\lfloor \log_R N \rfloor+2 \le 1.4405\, n+2，\quad R=(1+\sqrt{5})/2。$$

$(n-l)l$

1.6 密码学的复杂性理论基础

1.6.2 算法复杂性

- 一个算法的复杂性由该算法所要求的最大时间与存取空间确定
- 由于算法对于不同长度的输入数据所需要的执行时间T和存取空间S的大小往往不同，因此总是将算法的复杂度表示成长度n的函数
- 算法时间复杂度(Time Complexity) $T(n)$
  - 设A是一个可求解问题P的算法, 用A求解P的规模为n的实例所需用的最大时间称为算法A的<u>时间复杂度</u>.
  - 一般用求解算法实例的关键操作(基本操作)的次数作为计量单位
  - 平均时间复杂度: 用A求解P的规模为n的所有实例所需用时间的平均值称为算法A的<u>平均时间复杂度</u>，记为： $\overline{T}(n)$

1.6.2 算法复杂性

- **算法空间复杂度(Space Complexity) $S(n)$**
  - **设A是一个可求解问题P的算法，用A求解P的规模为 n的实例所占用的最大存储空间，$S(n)$称为算法A的<u>空间复杂度</u>.**
  - **一般对算法A的空间复杂性函数$S(n)$都有一定的限制.**
  - **平均空间复杂度：用A求解P的规模为n的所有实例所占用空间的平均值称为算法A 的<u>平均空间复杂度</u>，记为：$\overline{S}(n)$**

- **算法复杂度: $(T(n), S(n))$**

1.6.2 算法复杂性

◆**时间复杂度$T(n)$的表示: $T(n)=O(f(n))$**

求$T(n)$比较困难，取函数$T(n)$的 “主部”。

$$T(n)=O(f(n))\Leftrightarrow \exists k>0, n_0>0, \forall n>n_0, T(n)\le kf(n).$$

- **常数算法**

  $f(n)=C$, $T(n)=O(1)$

- **线性算法:**

  $f(n)$**是一次多项式,** $T(n)=O(n)$

- **多项式时间算法**

  $T(n)=O(p(n))$**，**$p(n)$**是**n**的次多项式.**

- **超多项式时间算法（或亚指数时间算法 ）**

$$T(n)=O(n^{\log_2 n}), O(e^{\sqrt{n\ln n}})$$

1.6.2 算法复杂性

- **指数时间算法(exponential time algorithm):**

  $T(n)=O(t^{p(n)})$,

  $t>1$为常数, $p(n)$为多项式.

1.6.2 算法复杂性

◆**例1.6.3 求计算*n*次多项式*p*(*n*)值的算法的时间复杂度. 设**

$$p(x)=a_nx^n+a_{n-1}x^{n-1}+\ldots+a_1x+a_0.$$

□**算法*A*:** $a_ix^i \to p(x)$.

计算a_ix^i需作$(i+1)$次乘法. 计算$p(x)$共需作

$$\sum_{i=0}^{n}(i+1)=\frac{(n+2)(n+1)}{2}=\frac{n^2}{2}+\frac{3n}{2}+1$$

次乘法, n次加法. 算法A的时间复杂度为:

$$T(n)=O(n^2).$$

1.6.2 算法复杂性

- **例1.6.3 求计算n次多项式$p(n)$值的算法的时间复杂度. 设**

$$p(x)=a_nx^n+a_{n-1}x^{n-1}+\dots+a_1x+a_0.$$

- **算法B: $T(n)=O(n)$.**

计算步骤	加法次数	乘法次数
$y_1=a_nx$	0	1
$y_2=(y_1+a_{n-1})x$	1	1
$y_3=(y_2+a_{n-2})x$	1	1
…	…	…
$y_n=(y_{n-1}+a_1)x$	1	1
$y_{n+1}=y_n+a_0$	1	0

1.6.2 算法复杂性

◆**例1.6.4 求计算指数函数x^n算法的时间复杂度.**

将n表示成二进制

$$n = a_r 2^r + a_{r-1} 2^{r-1} + \ldots + a_1 2 + a_0,$$

其中$r = \lfloor \log_2 n \rfloor$.

$$\begin{aligned} x^n &= x^{a_r 2^r + a_{r-1} 2^{r-1} + \ldots + a_1 2 + a_0} \\ &= x^{a_r 2^r} \cdot x^{a_{r-1} 2^{r-1}} \cdot \ldots \cdot x^{a_1 2} \cdot x^{a_0} \\ &= (x^{a_r 2^{r-1}} \cdot x^{a_{r-1} 2^{r-2}} \cdot \ldots \cdot x^{a_1})^2 \cdot x^{a_0} \\ &\qquad \ldots\ldots\ldots\ldots\ldots\ldots \\ &= ((\ldots(x^{a_r})^2 \cdot x^{a_{r-1}} \cdot \ldots \cdot x^{a_2})^2 x^{a_1})^2 \cdot x^{a_0}. \end{aligned}$$

计算x^n共需作r次平方运算,和至多r次乘法运算.

时间复杂度为: $T(n) = 2r = O(r) = O(\log_2 n)$.

1.6.2 算法复杂性

◆**不同时间复杂性算法的时间需求量级**

设机器每秒执行10^6条指令，$n=10^6$为输入规模

算法类型	复杂性	操作次数	实际时间
常数	$O(1)$	1	1微秒
线性	$O(n)$	10^6	1秒
二次	$O(n^2)$	10^{12}	11.6天
三次	$O(n^3)$	10^{18}	32000年
超多项式时间	$O(e^{\sqrt{n\ln n}})$	1.8×10^{1618}	6×10^{1600}年
指数时间	$O(2^n)$	10^{301030}	3×10^{301016}年

1.6.2 算法复杂性

- **计算机速度的提高对算法时间需求的影响**
  - **设现在的机器每秒执行10^6条基本操作**
  - **当计算机处理能力显著提高时，对超多项式阶算法或指数阶算法处理能力的提高影响甚微.**

1小时内的操作次数

计算设备条件 / 处理能力 / 时间复杂性	**用现在的计算机**	**用快100倍的计算机**	**用快10000倍的计算机**
$O(n)$	$N_1 \approx 3.6\times10^9$	$100N_1$	$10000N_1$
$O(n^2)$	$N_2 \approx 6\times10^4$	$10N_2$	$100N_2$
$O(n^3)$	$N_3 \approx 1.5\times10^3$	$4.64N_3$	$21.5N_3$
$O(e^{\sqrt{n \ln n}})$	$N_4 \approx 10^4$	$1.38N_4$	$1.79N_4$
$O(2^n)$	$N_5 \approx 32$	$N_5+6.64$	$N_5+13.29$

- **1.6.3 问题按复杂度分类**
  - **图灵机(Turing Machine)**

    具有无限读写能力的有限状态机, 它是一种理想的计算机模型.

    - **确定型图灵机(DTM: deterministic Turing machine)**

      每一步操作结果及下一步操作内容都是唯一确定的.

    - **非确定型图灵机(NDTM: non-deterministic Turing machine)**

      每一步操作结果及下一步操作内容可以有多种选择.

      非确定型图灵机解一个问题分为两个阶段: 猜测与验证.

1.6.3 问题按复杂度分类

- **问题复杂度（problem complexity）**
  - **设*P*是一个问题, *F*是能求解*P*的全体算法的集合. 对任意的算法$A \in F$, 用$T_A(n)$表示*A*的时间复杂度. <u>问题*P*的时间复杂度</u>定义为:**

$$T(n)=\min\{T_A(n) \mid A \in F\}.$$

  - **<u>问题复杂度</u>由在图灵机上解其最难实例所需的最小时间与空间确定.**

最好的算法

用

  - **问题复杂度可以理解为: 由解该问题的最有效的算法所需的时间与空间来度量.**

1.6.3 问题按复杂度分类

◆**P类问题**

□**定义1.6.1（P类问题）**

> **如果存在一个DTM，在多项式时间内能求解问题*Q*，则称问题*Q*是多项式时间可解的，简称为P问题。全体P问题构成的类记为P。**

□**P问题也称为易解的（tractable）**

□**如果问题*Q*不属于P，则称问题*Q*是难解的或难的（intractable or hard）**

□**不可判定问题（undecidable problem）**

能证明无法构造一个算法求解的（难）问题.

1.6.3 问题按复杂度分类

- **NP类问题**
  - **定义1.6.2（NP类问题）**

非确定图灵机

> 如果在多项式时间内在NDTM上能求解问题*Q*，即如果NDTM的“答案猜测器”能猜出问题*Q*的答案，则NDTM能在多项式时间内验证它，则称问题*Q*是非确定性多项式时间可解问题，简称为NP问题。全体NP问题构成的类记为NP。

1.6.3 问题按复杂度分类

- **例1.6.5 背包问题(knapsack problem) (子集和问题 : subset sum problem)K:**

  给定n个整数的集合$A=\{a_1, a_2,\ldots, a_n\}$, 和一个整数$S$, 确定是否存在$A$的子集$B$, 使得

  $$\sum_{x\in B} x = S.$$

  - **对于给定A的子集B, 容易验证$\sum_{x\in B} x=S$是否成立，即:**

    $$K\in NP.$$

  - **选择A的子集B, 共有2^n种结果. 即试验所有子集的时间复杂度为: $T(n)=O(2^n)$.**

1.6.3 问题按复杂度分类

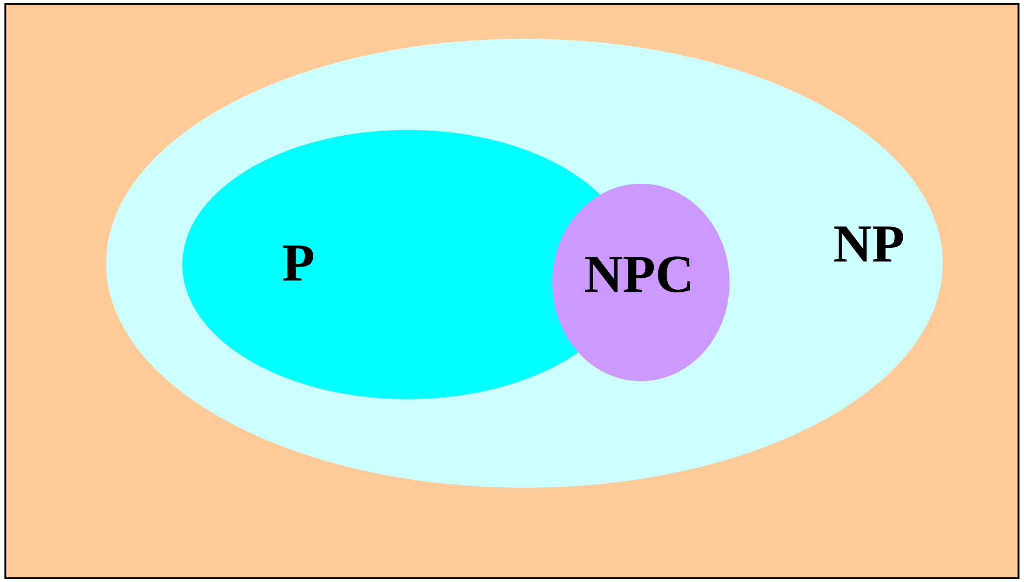
- **P类与NP类的关系**
  - $P \subseteq NP$
  - **世界难题(big problem): $P=NP$?**
- **NPC类问题**
  - **定义1.6.3（NPC类问题）**

> 设Q是一个NP问题, 如果NP的任何一个问题都可以通过多项式时间转化为该问题Q, 则称Q是 NP完全的, 称为NPC问题. 全体NP完全问题构成的类记为NPC.

- **性质: 如果NPC中有一个问题属于P，则NP问题都属于P. 即有P=NP.**
- **NPC是NP中“最难”的问题.**
- **对于NPC问题, 目前不存在有效的算法.**

1.6.3 问题按复杂度分类

◆**P类、NP类与NPC类问题的关系**



1.6.3 问题按复杂度分类

- **典型的NPC问题**
  - ◆ **背包问题** 子集和
  - ◆ **哈密尔顿(Hamilton)问题:**

    在有n个顶点的图中, 求一条经过每个顶点一次且仅一次的回路.
  - ◆ **平方剩余问题:** $x^2 = a + nb$

    已知正整数a, b, 求以下同余方程的解:

    $$x^2 \equiv a \bmod b.$$ 余数
  - ◆ **可满足性问题(SAT: satisfiability)**

    判断一个n元布尔(Boolean)函数$y=f(x_1, x_2,\ldots, x_n)$是否存在一组赋值$(t_1, t_2,\ldots, t_n)$, 使得 $f(t_1, t_2,\ldots, t_n)=1$.

第1章　　习　题

- **P33-34: 习题3, 4, 5, 6, 7, 8.**

第1章 概论

- 1.1 信息系统安全与密码技术
- 1.2 密码系统模型和密码体制
- 1.3 几种简单密码体制
- 1.4 初等密码分析
- 1.5 密码学的信息论基础
- 1.6密码学的复杂性理论基础

1.5* 密码学的信息论基础

- **1.5.1 密码系统的熵**

已知密码体制: $(\mathscr{M}, \mathscr{C}, \mathscr{K}, \mathscr{E}, \mathscr{D})$

◆**明文字母表：** $A=\{a_i, i=0,1,\ldots, q-1\}=Z_q$**的概率分布**

$$\begin{bmatrix} A \\ P(A) \end{bmatrix} = \begin{bmatrix} a_0, & a_1, & \ldots, & a_{q-1} \\ p(a_0), & p(a_1), & \ldots, & p(a_{q-1}) \end{bmatrix},$$

$$p(a_i) = p(M = a_i) \geq 0, \quad \sum_{i=0}^{q-1} p(a_i) = 1.$$

明文: $m=(m_0, m_1,\ldots, m_{L-1})\in A^L$, **如果信源无记忆，则**

$$p(m) = \prod_{i=0}^{L-1} p(m_i).$$

明文空间: $\mathscr{M}=A^L=Z_q^{\,L}$.

明文熵： $H(\mathscr{M})=H(A^L)=H(Z_q^{\,L})$.

1.5.1 密码系统的熵

◆**密钥字母表：** $B=\{b_i, i=0,1,\ldots,s\text{-}1\}=Z_s$**的概率分布**

$$\begin{bmatrix} B \\ P(B) \end{bmatrix} = \begin{bmatrix} b_0, & b_1, & \ldots, & b_{s-1} \\ p(b_0), & p(b_1), & \ldots, & p(b_{s-1}) \end{bmatrix},$$

$$p(b_i) = p(K = b_i) \geq 0, \quad \sum_{i=0}^{s-1} p(b_i) = 1.$$

密钥:$k=(k_0,k_1,\ldots,k_{r\text{-}1})\in B^r$**，密钥相互独立**

密钥空间:$\mathscr{K}\doteq B^r = Z_s^{\ r}$

密钥熵: $H(\mathscr{K})=H(B^r) = H(Z_s^{\ r})$

1.5.1 密码系统的熵

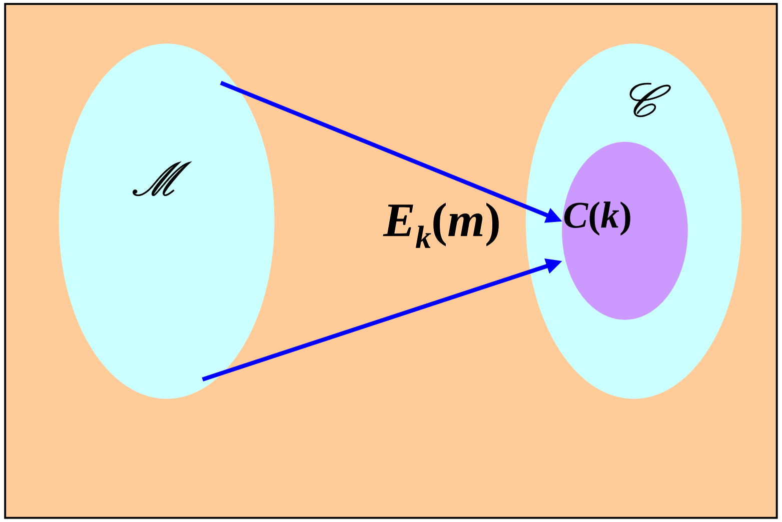
- **密文字母表：**

与明文字母表相同$W=\{a_i, i=0,1,\ldots,q-1\}=Z_q$，**其概率分布由明文和密钥的统计特性决定**

密文： $c=(c_0, c_1,\ldots,c_{L'-1})$

密文空间： $\mathscr{C}=W^{L'}=Z_q^{L'}$

密文熵： $H(\mathscr{C})=H(W^{L'})=H(Z_q^{L'})$



$\forall k \in \mathscr{K}$，令

$C(k)=\{E_k(m): m \in \mathscr{M}\}$.

$\mathscr{C}$的概率分布为：$\forall c \in \mathscr{C}$，有

$$p(\mathscr{C}=c) = \sum_{\{k: c\in C(k)\}} p[\mathscr{K}=k]\,p[\mathscr{M}=D_k(c)].$$

1.5.1 密码系统的熵

◆给定明文*m*, 关于密文*c*的条件概率

设 $m\in\mathscr{M}, c\in\mathscr{C}$, 有

$$p(C=c\mid M=m)=\sum_{\{k:m=D_k(c)\}}p[K=k].$$

◆给定密文*c*, 关于明文*m*的条件概率

设 $m\in\mathscr{M}, c\in\mathscr{C}$, 有

$$p[M=m\mid C=c]$$

$$=\frac{p[M=m]\times\sum_{\{k:m=D_k(c)\}}p[K=k]}{\sum_{\{k:c\in C(k)\}}p[K=k]p[M=D_k(c)]}.$$

1.5.1 密码系统的熵

◆**给定密文*c*,关于密钥*k*的条件概率**

设 $k\in\mathscr{K}$, $c\in\mathscr{C}$, 有

$$p[K=k \mid C=c]=\frac{p[K=k]\times p[C=c \mid K=k]}{p[C=c]}$$

$$=\frac{p[K=k]\times \sum\limits_{\{m:E_k(m)=c\}} p[M=m]}{\sum\limits_{\{k:c\in C(k)\}} p[K=k]p[M=D_k(c)]}.$$

1.5.1 密码系统的熵

- **明文熵：** $H(\mathscr{M})=H(M^L)$
- **密钥熵:** $H(\mathscr{K})=H(B^r)$
- **密文熵：** $H(\mathscr{C})=H(C^{L'})$
- **在已知密文条件下密钥的含糊度：** $H(\mathscr{K}|\mathscr{C})$
- **在已知密文条件下明文的含糊度：** $H(\mathscr{M}|\mathscr{C})$
- **密钥含糊度定理**

定理1.5.1　任给密码系统 $(\mathscr{M}, \mathscr{C}, \mathscr{K}, \mathscr{E}, \mathscr{D})$**，有**

$$H(\mathscr{K}|\mathscr{C})=H(\mathscr{M})+H(\mathscr{K})-H(\mathscr{C}).$$

1.5.2 完善保密性

- **对于唯密文破译，从密文提取有关密钥的信息，或者从密文提取有关明文的信息：**

$$I(\mathscr{K};\mathscr{C})=H(\mathscr{K})-H(\mathscr{K}|\mathscr{C})$$

$$I(\mathscr{M};\mathscr{C})=H(\mathscr{M})-H(\mathscr{M}|\mathscr{C})$$

- **对于合法接收者，因为**

$$H(\mathscr{M}|\mathscr{K}\mathscr{C})=0,$$

所以 $I(\mathscr{M};\mathscr{K}\mathscr{C})=H(\mathscr{M}).$

1.5.2 完善保密性

定理1.5.2　任给密码系统 $(\mathscr{M}, \mathscr{C}, \mathscr{K}, \mathscr{E}, \mathscr{D})$，**有**

$$I(\mathscr{M};\mathscr{C}) \geqslant H(\mathscr{M})-H(\mathscr{K}).$$

证明：因为$H(\mathscr{K}) \geqslant H(\mathscr{K}|\mathscr{C})$

$$=H(\mathscr{K}|\mathscr{C})+H(\mathscr{M}|\mathscr{K}\mathscr{C}) = H(\mathscr{M}\mathscr{K}|\mathscr{C})$$

$$=H(\mathscr{M}|\mathscr{C})+H(\mathscr{K}|\mathscr{M}\mathscr{C}) \geqslant H(\mathscr{M}|\mathscr{C}),$$

所以, $I(\mathscr{M};\mathscr{C})=H(\mathscr{M})-H(\mathscr{M}|\mathscr{C}) \geqslant H(\mathscr{M})-H(\mathscr{K}).$

- $H(\mathscr{K})$**越大, 密文中含有的关于明文的信息量就越小。密钥设计要求:**
  - **密钥相互独立**
  - **每个密钥等概率**
  - **密钥量足够大**

1.5.2 完善保密性

- **完善保密性(perfect secure), 或具有无条件安全性(unconditional security): 满足**

$$I(\mathscr{M};\ \mathscr{C})=0.$$

- **完善保密密码系统对唯密文破译而言是绝对安全的，但是并不能保证系统在更强的破译条件下（已知明文破译、选择明文破译、选择密文破译）也是安全的。**

定理1.5.3 密码系统 $(\mathscr{M}, \mathscr{C}, \mathscr{K}, \mathscr{E}, \mathscr{D})$具有完善保密性的必要条件是

$$H(\mathscr{K})\geqslant H(\mathscr{M}).$$

1.5.2 完善保密性

- **例1.5.3 构造一个完善保密的密码系统**

字母表： $A = B = W = Z_2$,

明文、密钥长度: $L = L' = r$,

明文空间: $\mathscr{M} = \{m = (m_1, m_2, \ldots, m_L), m_i = 0,1\}$,

密钥空间: $\mathscr{K} = \{k = (k_1, k_2, \ldots, k_L), k_i = 0,1\}$.

假设$\mathscr{M}$与$\mathscr{K}$相互独立，密钥为随机二元序列，

即任给$k \in \mathscr{K}$, 有

$$p(K=k) = 1/2^L,$$

因而，

$$H(\mathscr{K}) = L(\text{bit}).$$

1.5.2 完善保密性

- **例1.5.3 构造一个完善保密的密码系统**

采用弗纳姆体制，加密变换为:

$$c=E_k(m)=m\oplus k=(m_1\oplus k_1, m_2\oplus k_2,\ldots,m_L\oplus k_L),$$

密文空间: $\mathscr{C}=\{c=(c_1, c_2,\ldots,c_L), c_i=0, 1\}$.

由于$p(k_i=0)=p(k_i=1)=1/2$, **所以**

$$p(c_i=0)=p(c_i=1)=1/2.$$

因此，加密变换E_k: $m_i\to c_i$**等价于一个转移概率为1/2的二元对称信道(BSC).**

由于 BSC的信道容量$C_a=0$, **且**

$$I(M^L; C^L)\le LC_a=0.$$

所以, $\quad I(M^L; C^L)=0.$

即该密码系统是完善保密的.

1.5.2 完善保密性

- **例1.5.3 构造一个完善保密的密码系统**
- **该密码系统在唯密文破译下是安全的**
- **在已知明文攻击下是不安全的**

因为若知道了明文—密文对(m, c)，由

$$c=m\oplus k,$$

可求得

$$k=m\oplus c.$$

1.5.3 唯一解距离、理论保密性与实际保密性

- **唯一解距离**
  - **伪密钥(spurious key)**

    在唯密钥攻击条件下，密码分析者对截获的密文c，可以用所有的密钥k进行解密，从而找出有意义的明文。因此，密码分析者可能找出许多密钥，对c解密均有意义。但其中只有一个是正确的，其它是假的，称为伪密钥(spurious key)。

密码分析者希望伪密钥数为零！

定义1.5.1 长度为n的密文串的伪密钥的期望数记为：$\bar{s}_n$.

1.5.3 唯一解距离、理论保密性与实际保密性

- **唯一解距离**

  - **语言的冗余度（redundancy）**

定义1.5.2 设L是一种自然语言，A是L的字母集, M^n是A^n上的随机变量

语言L的熵（或速率）：$H_L = \lim\limits_{n\to\infty} \frac{H(M^n)}{n}$.

语言L的冗余度：$R_L = 1 - \frac{H_L}{\log_2 |A|}$.

H_L表示语言L中每个字母的熵，它是“有意义的”明文字母串中每个字母的平均信息的度量。

R_L是语言L中“多余字母”所占比例的度量。

1.5.3 唯一解距离、理论保密性与实际保密性

◆**例1.5.4　对于英语L, 有**

$H(M)=4.15$**比特/字母;**

$H(M^2)/2=3.62$**比特/字母;**

$H(M^3)/3=3.22$**比特/字母;**

……

$1.0 \le H_L \le 1.5$ **比特/字母.**

若取$H_L=1.25$，**则有**

$R_L=0.75$.

英文的冗余度为75%。

1.5.3 唯一解距离、理论保密性与实际保密性

- **唯一解距离**
  - **伪密钥的期望值**

> **定理1.5.4　设 $(\mathscr{M}, \mathscr{C}, \mathscr{K}, \mathscr{E}, \mathscr{D})$是一个密码系统，$A$是明文字母表，$B$是密文字母表，$|A|=|B|$，$R_L$是明文语言的冗余度。如果密钥的选取满足均匀分布，则对于任意一个长度为n的密文字母串，当n充分大时，伪密钥的期望值满足：**
>
> $$\bar{s}_n \geq \frac{2^{H(\mathscr{K})}}{|A|^{nR_L}} - 1.$$

1.5.3 唯一解距离、理论保密性与实际保密性

◆唯一密文破译下的唯一解距离（unicity distance）

> **定义1.5.3** 密码系统$(\mathscr{M}, \mathscr{C}, \mathscr{K}, \mathscr{E}, \mathscr{D})$的唯一解距离$n_0$定义为使得伪密钥的期望值为零的密钥长度$n$，即有
>
> $$\bar{s}_{n_0} = 0.$$

- **当截获的密文长度大于n_0时，原则上就可以唯一地确定系统所用的密钥（如果能够将密钥的可能取值限制在很小的范围内，通过试验也就不难破译了），也就是说，原则上可以破译该密码。**
- **若截获的密文符号数量少于n_0时，就存在有多种可能的密钥，密码分析者无法从中确定哪一个解是正确的。**

1.5.3 唯一解距离、理论保密性与实际保密性

◆ **唯密文破译下的唯一解距离（unicity distance）**

- **唯一解距离的近似值**

$$n_0 \approx \frac{H(\mathscr{K})}{R_L \log_2 |A|}.$$

- **例：对于单表代换密码体制，假设每个密钥被均匀选取,有**

$$|A| = 26, R_L = 0.75,$$

$$|\mathscr{K}| = 26!, H(\mathscr{K}) = \log_2 26 = 88.4,$$

$$n_0 \approx \frac{88.4}{0.75\log_2 26} = \frac{88.4}{0.75 \times 4.7} \approx 25.$$

1.5.3 唯一解距离、理论保密性与实际保密性

- **理论保密性**

  理论保密性是假定密码分析者有无限的时间、设备和资金条件下，研究唯密文攻击时密码系统的安全性。一个密码系统，如果对手有无限的资源可以利用，而在截获任意多的密文下仍不能被破译，则它在理论上是保密的。

- **实际保密性**

  一个密码系统的破译所需要的工作量，如果超过了对手的能力（时间、资源等），则认为该系统是实际保密的（practical secrecy）。

- **可证明安全的（provable secure）**

  破译密码的难度等价于（不低于）数学上的某个已知难题。