

密码学基础

范明钰

信息安全研究中心

主要内容

- 先修课程
- 课程的主要内容及安排
- 考核方式

- 密码学与信息安全
- 基本概念、历史
- 数学模型简介
- 破译理论概念

预备知识和先修课程

动脑：

- 获取知识的能力
- 综合分析和应用能力

动手：

- 综合分析：统计、数据挖掘
- 应用：操作系统，网络通信协议

基础知识：

- 概率论
- 信息论
- 复杂性理论
- 数论、组合数学
- 网络通信体系与协议

课程的主要内容划分及课时安排

- 基础部分 (16学时)
- 核心部分 (28--30学时)
- 应用部分 (2--3学时)

主要线索

- 基本理论：基础部分→基本方法
- 基本方法：基础+核心+应用
- 研究成果：算法，及其协议
- 面临的挑战：贯穿于上述各部分之中

参考资料

- 说明：教材仅供参考，以本讲稿为主
- 密码编码学与网络安全：原理与实践（第六版），2015年，电子工业出版社
- 范明钰，王光卫，密码理论与技术，清华大学出版社，2009年
- 王育民，何大可，保密学—基础与应用，西安电子科技大学出版社，1990年
- 赖溪松，韩亮，张真诚，计算机密码学及其应用，国防工业出版社，2001年7月
- 卢开澄，计算机密码学---计算机网络中的数据保密与安全，清华大学出版社，1998年7月

联系信息

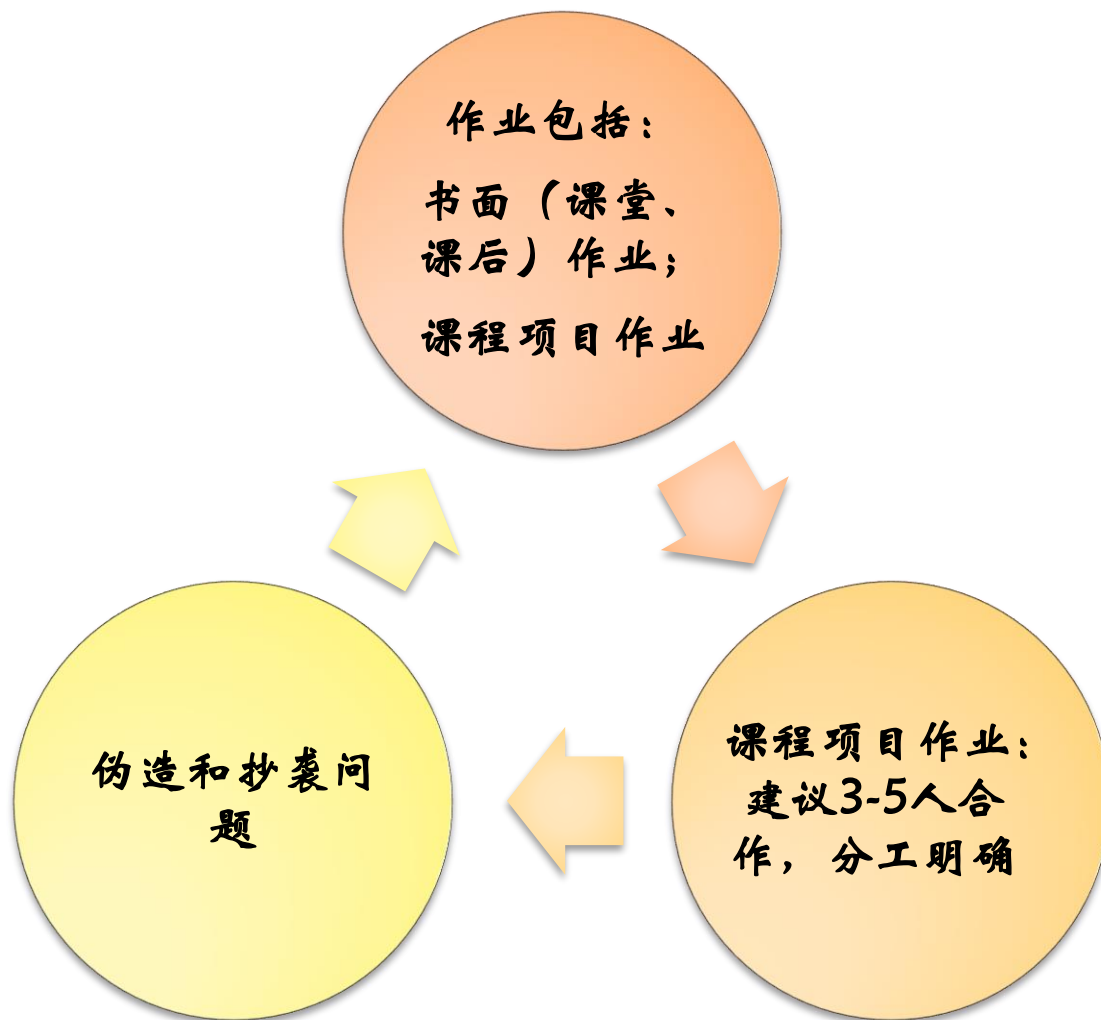
- 信息安全研究中心
- 主楼B-102
- 电话：1330 822 2580
- Email: ff98@163.com

- 学习群：211438617
- 资料下载密码：
- 助教：

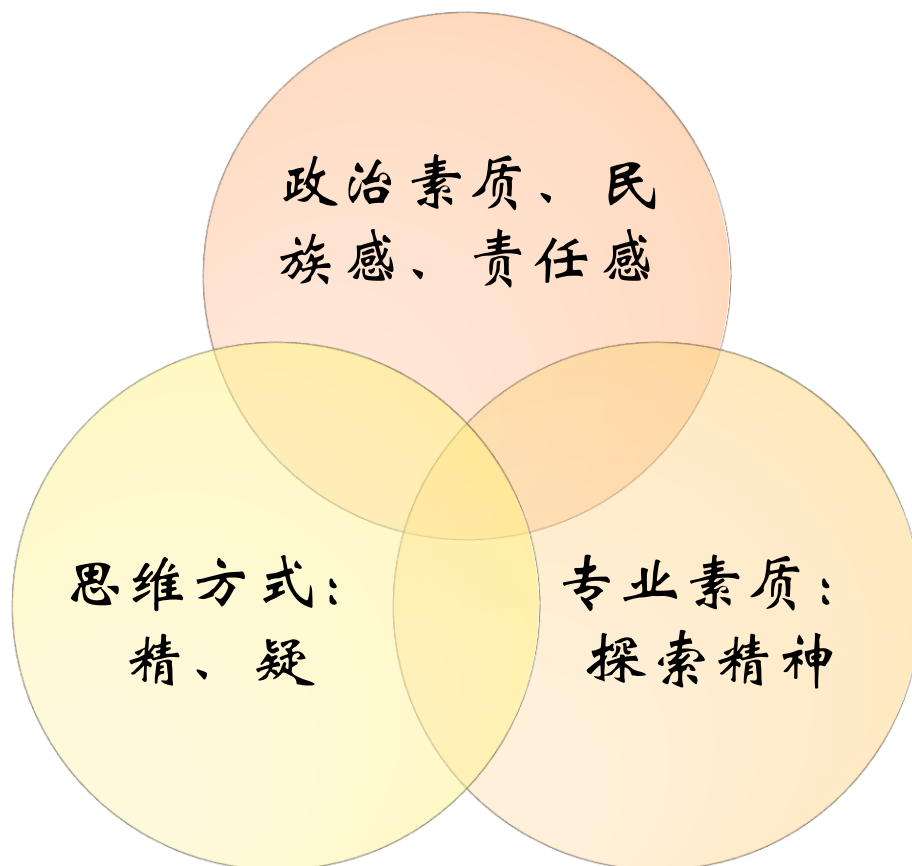
课程基本要求

- 课堂讲授 + 作业实践
- 了解和掌握密码学的基本原理、技术、及最新研究成果
- 具有密码理论基础和基本应用实践能力
- 可自主选择时间和投入，以达到基本或更高要求
- 平时作业(30%)+ 期末笔试(70%)

作业要求

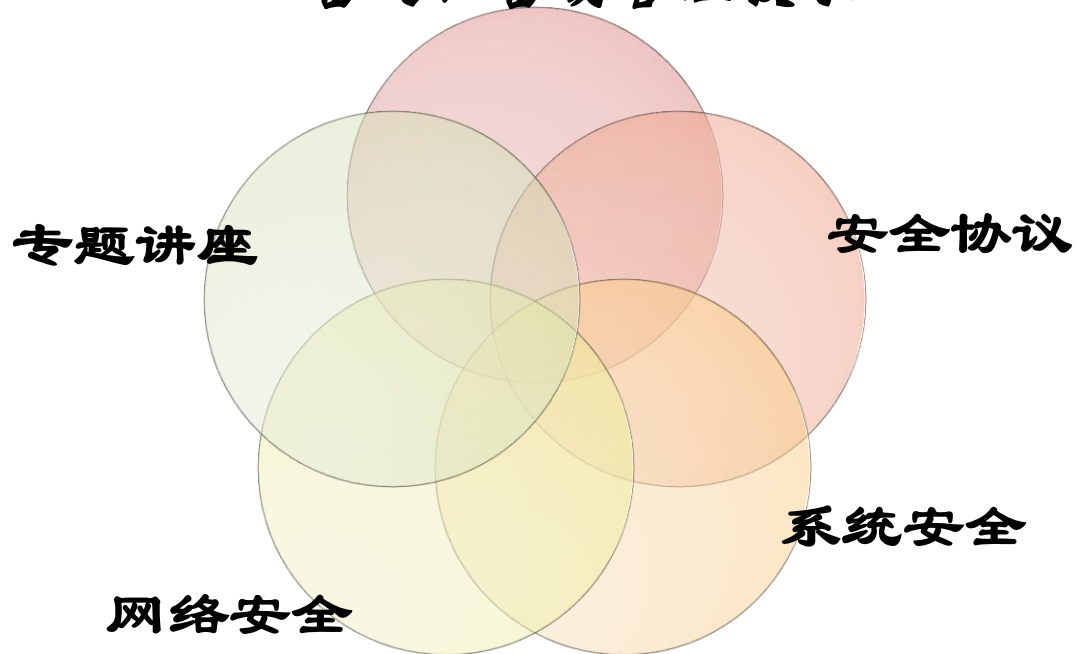


素质要求



课程体系

密码学基础：基本理论、经典密码、现代密码、密钥管理技术



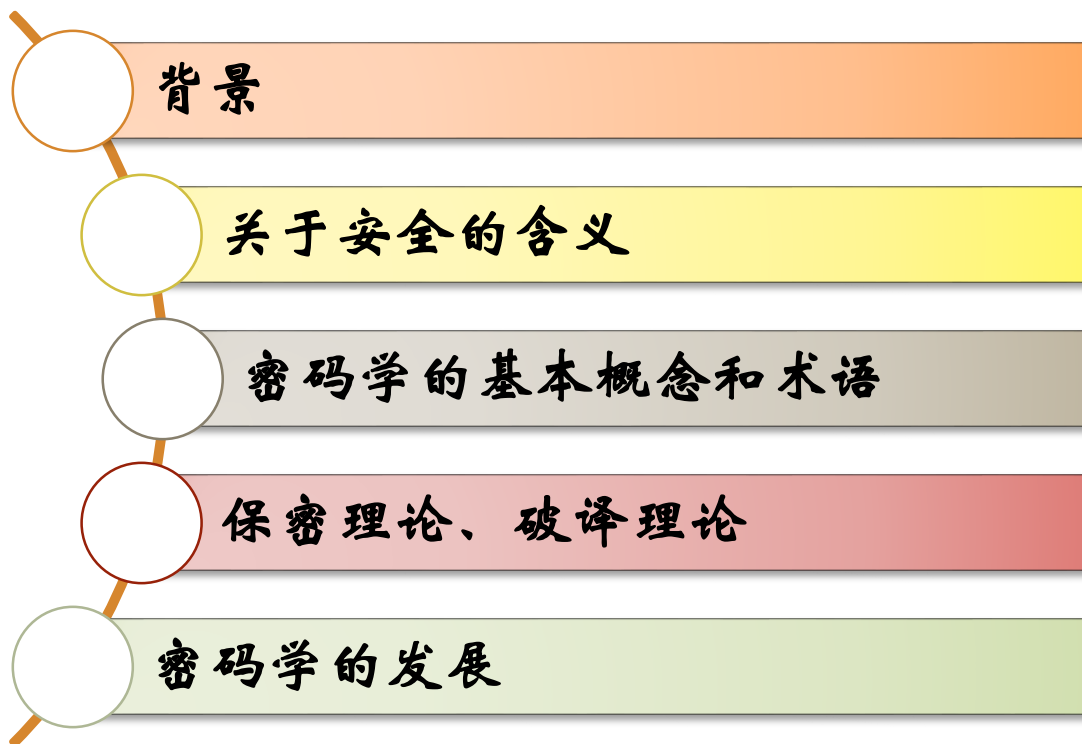
课堂作业-1

- 在你的概念中，密码学的主要研究内容有哪些？用简洁的语言，说说你所认识的密码学，以及你对课程的希望或要求。
- 现场提交



Questions ?

接下来主要内容



课程背景

- 现实应用中的需要：密码从军事走向生活
- 国家民族高度



背景

- 随着计算机的广泛使用，尤其是internet的出现与发展，**网络信息的安全**受到越来越多的关注；各国政府高度重视
- 发达国家和地区投入巨资，保护其信息基础建设的安全，增强其综合实力
- 在中国，则更多地依赖于国家政策，维护自身的信息安全
- 讨论：哪种方式靠谱？Why？

背景

- **全球一体化**的经济和科技的发展，使得国与国，人与人之间有了更加紧密的依靠和渗透
- **信息成为资源**：国际形势复杂多变，世界范围内黑客对计算机网络的攻击与破坏活动日趋猖獗，一些发达国家正在利用其信息技术、经济和军事优势来达到称霸世界的目的

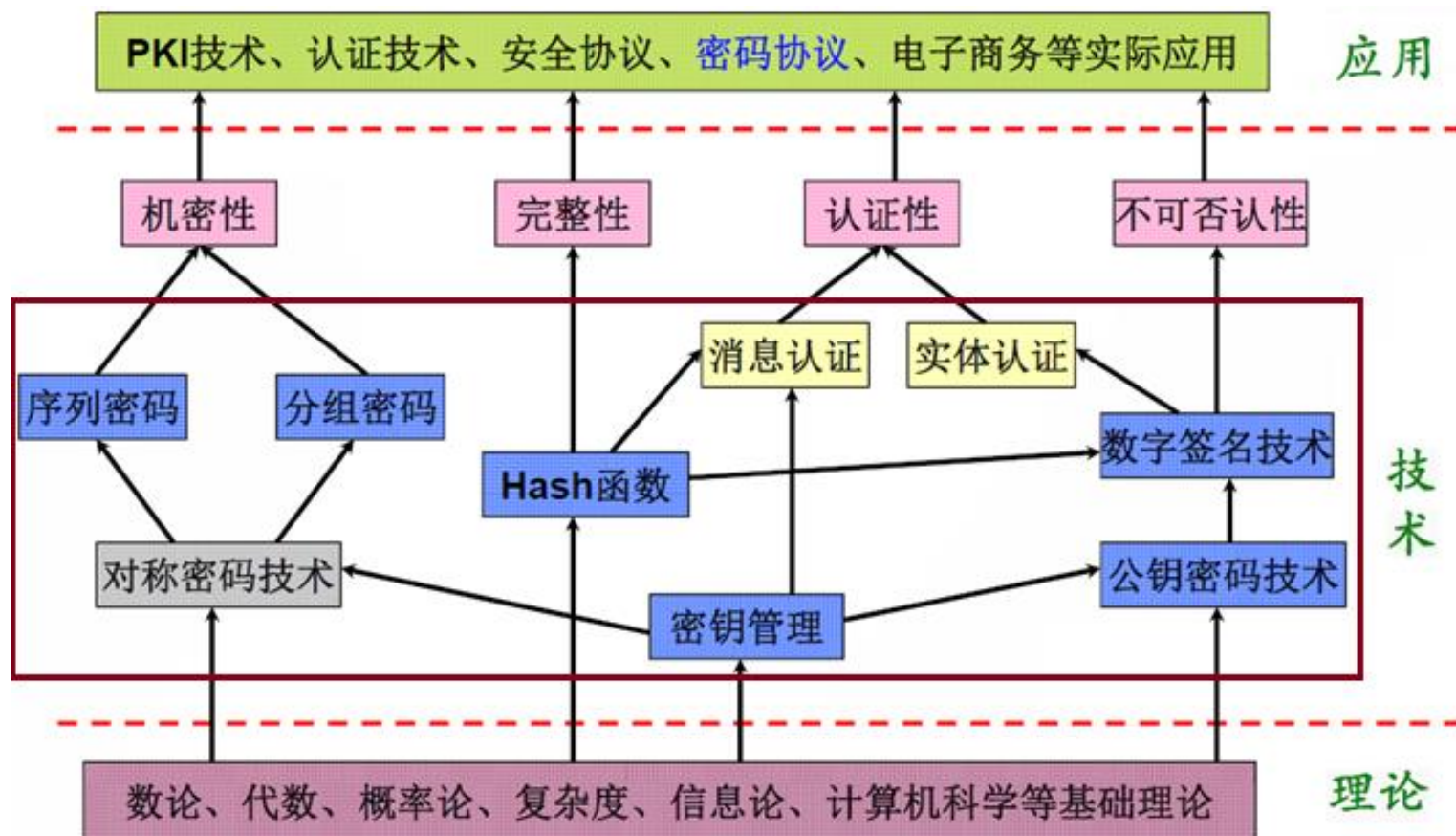
密码技术，是解决资源掠夺的关键技术之一

- 保护信息的主要手段：密码技术
- 其他手段：法律、取证技术
- 讨论：非密码技术的信息保护？

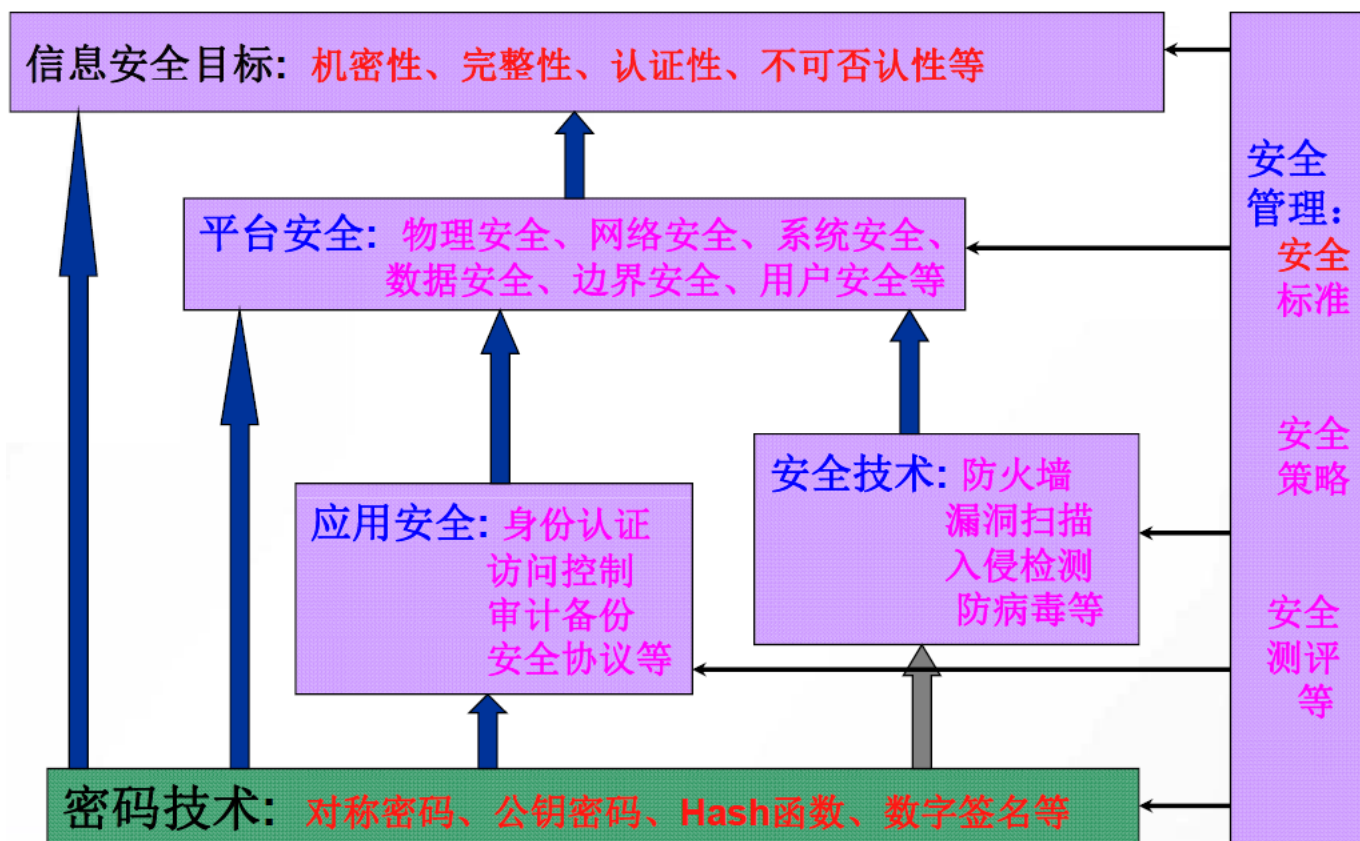
密码学的地位和作用

- 密码学在信息安全领域起着基本的、无可替代的作用。
- 事实上，密码是解决网络信息安全的关键技术，是现代数据安全的核心。
- 美国密码学会会长，戴维·卡恩说：“只有拥有核技术、太空技术和密码技术的国家，才称得上是世界强国。”

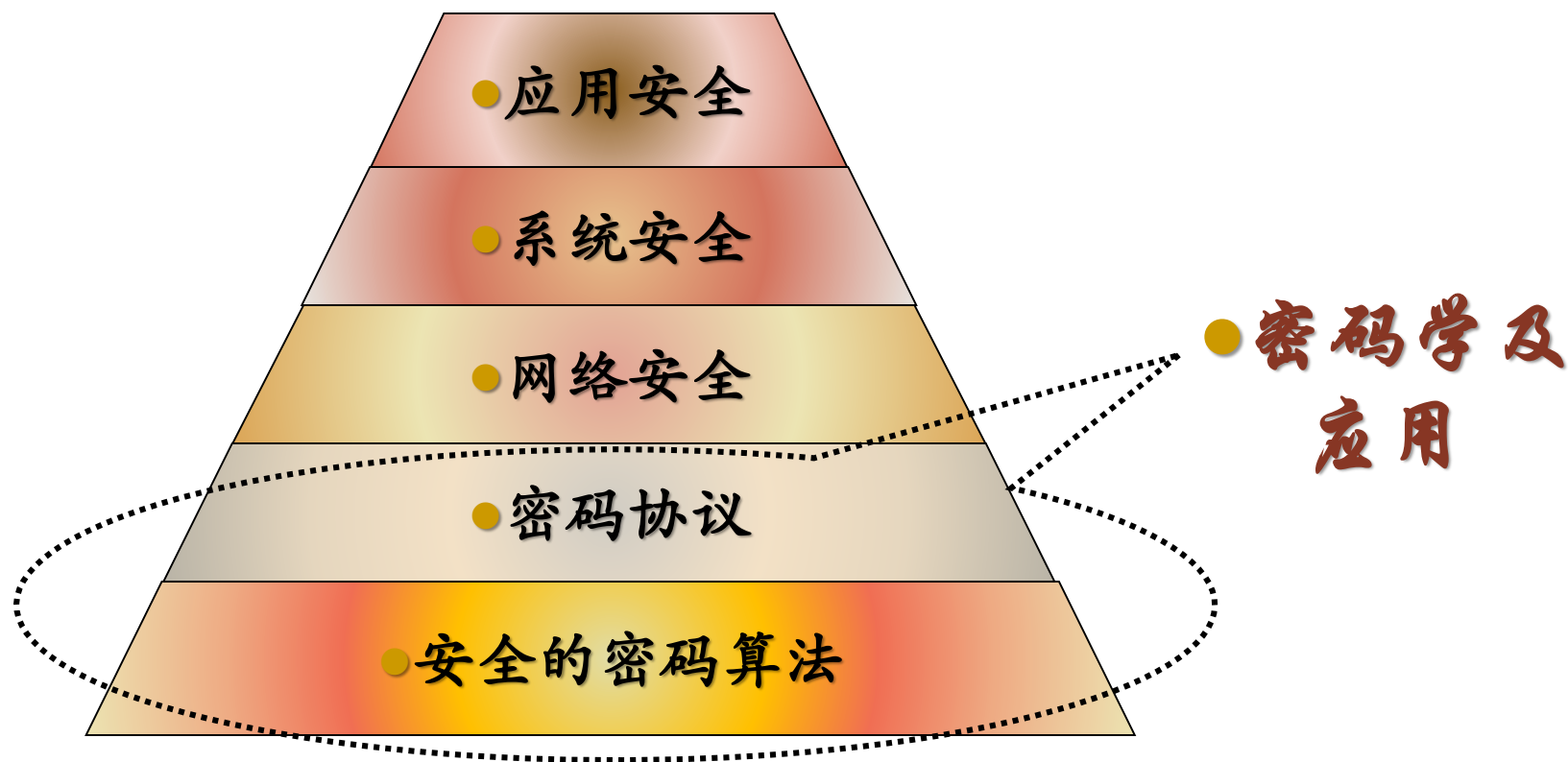
密码学的主要研究内容



密码技术与信息安全研究内容的关系



密码学的地位和作用—换一个角度看



密码技术应用范围的演变

- 点对点通信 → 专用网络 → 互联网络
- 数据保护 → 数据安全 → 网络安全 → 信息安全

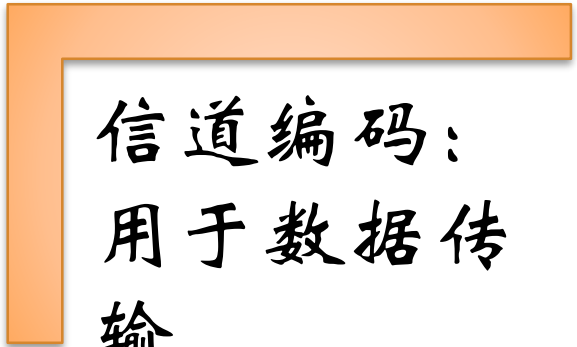
密码学涉及范畴及外延的演变

通信保密
(COMSEC): 60-70
年代, 信息保密

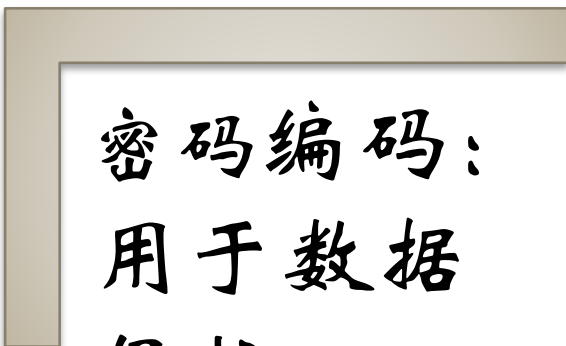

信息保障(IA): 90
年代-

信息安全
(INFOSEC): 80-90
年代, 机密性、完
整性、可用性、不
可否认性等

密码编码与信道编码



信道编码：
用于数据传
输



密码编码：
用于数据
保护

密码学的基本概念

密码编码学
(Cryptography)

- 算法设计

密码分析学
(Cryptanalytics)

- 算法安全性，以及破译

密钥管理学(Key
management)

- 主要研究密钥的产生、存储与分配方法，以及密钥的整个生命过程

密码的起源

- 可追溯到人类刚刚出现，并且尝试去学习如何通信的时候
- 随着文字的出现和使用，确保通信的机密性成为一种艺术，古代发明了不少加密信息和传达信息的方法
- 例如我国古代的烽火就是一种传递军情的方法，再如古代的兵符就是用来传达信息的密令
- 就连闯荡江湖的侠士，都有秘密的黑道行话，更何况是那些不堪忍受压迫义士在秘密起义前进行地下联络的暗语，这都促进了密码学的发展。

密码学的故事

- 希特勒时期，德国使用一种名为“Enigma”的密码机；英国完成了针对“Enigma”的绰号叫“炸弹”的密码破译机，使得同盟国几乎掌握了纳粹德国的绝大多数军事秘密和机密，而德军对此却一无所知
- 太平洋战争中，美军成功破译日本海军的密码机，在中途岛彻底击溃了日本海军，击毙山本五十六，导致太平洋战争的决定性转折
- 因此，我们可以说，密码学为战争的胜利立了大功。

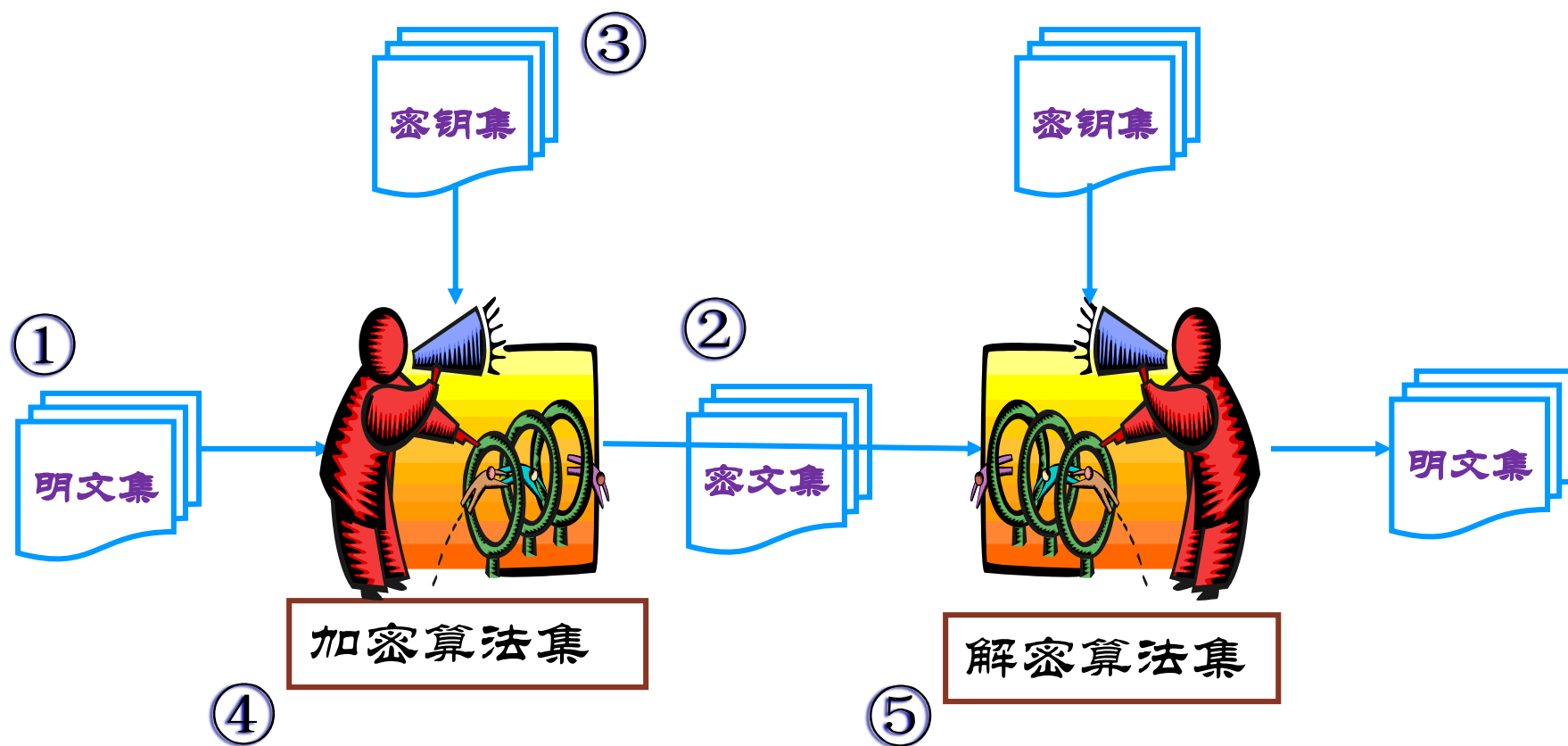
密码学基本术语

- 受保护的消息(信息)称为明文。对明文进行保护的过程称为加密；被加密的消息称为密文；把密文转变为明文的过程称为解密
- 密码算法：用于加密和解密的数学函数。对明文进行加密操作时所称作加密算法，对密文解密时称为解密算法
- 密钥：指示加密/解密变换的参数
- 上述进行明密变换的法则，称为密码的体制。

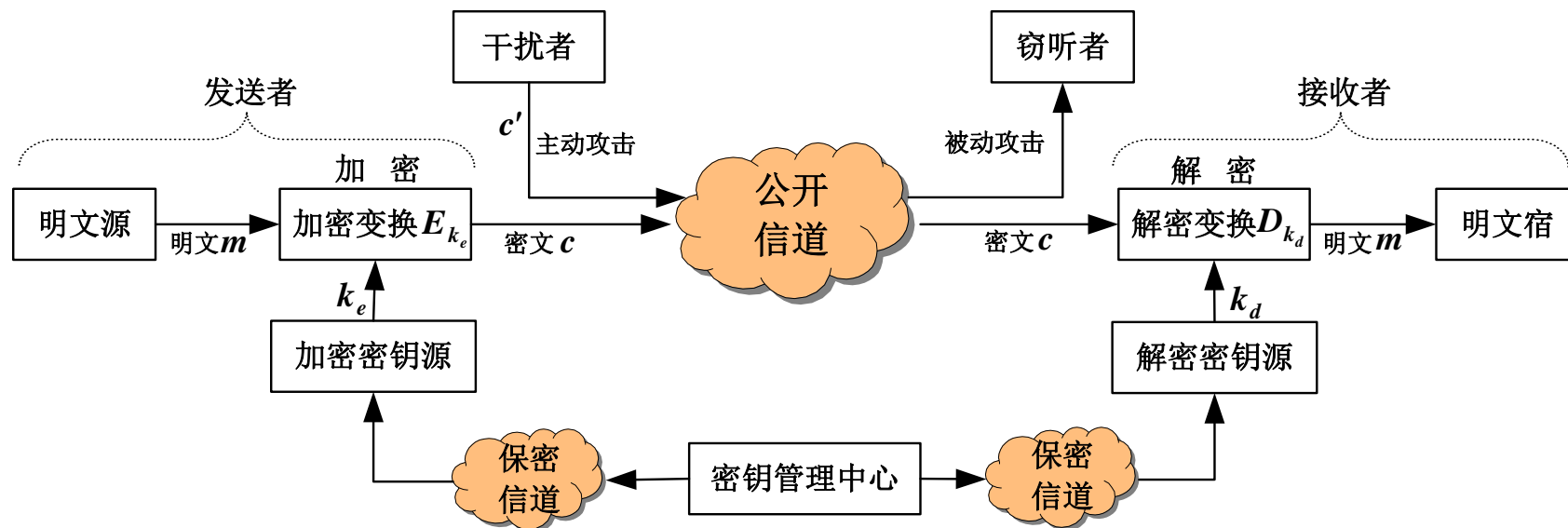
密码体制--是一个五元组 (P, C, K, E, D) ，满足：

- P 是可能明文的有限集：**明文空间**
- C 是可能密文的有限集：**密文空间**
- K 是可能密钥的有限集：**密钥空间**
- 对于任意 $k \in K$ ，有一个加密算法 $e_k \in E$ ，使得 $e_k: P \rightarrow C$ ，
即 $e_k(x)=y$ (这里 $x \in P$ ， $y \in C$)，称 E 为**加密变换族**(E_k)
- 相应的解密算法 $d_k \in D$ ， $d_k: C \rightarrow P$ ，满足 $d_k(e_k(x)) = x$ ，
称 D 为**解密变换族**(D_k)

密码体制——概念的图示



密码通信系统模型



密码系统的理论及数学模型

- 概率模型
- 密码算法设计的基本原则
- 密码分析的基本方法
- 对密钥的假定
- 密码系统的运算
- 信息量与熵、模糊度
- 理想安全与实际安全
- 强力破解
- 算法设计要点

概率模型

- ◆ **密码系统**是一族可逆映射，从消息空间映射到密文空间。其中各映射具有一定的使用概率，由实际密钥确定具体的映射
 - ✓ 可逆映射 E_1, E_2, \dots, E_s 的使用概率为 p_1, p_2, \dots, p_s
 - ✓ 解密映射 D_1, D_2, \dots, D_s 分别与加密映射 E_1, E_2, \dots, E_s 互逆
- ◆ **密码系统相同**是指映射集、消息空间、密文空间、密钥空间相同，且相应的先验概率相同

密码算法设计的基本原则

○ 柯克霍夫原则

- 公开加密映射族、解密映射族（加解密算法）；
- 公开明文的先验概率 q_i ，和密钥的先验概率 p_i ；
- 公开密文
- 仅保密实际消息和实际密钥

○ 利用先验概率，计算消息和密钥的后验概率

- 当某个消息或密钥的后验概率接近为1，其它的接近0时，则该密文被破译，否则密文安全

密码分析的基本方法

◆ 将消息视为一个整体，不考虑消息内部文字间的关系

- 消息可视为一个马尔科夫链随机过程，不同消息的概率由马尔科夫链决定
- 将消息简化，用一个符号 m_i 代替，并赋予一个概率

◆ 不考虑在明文中插入的无效内容，基本密码系统中不考虑多次加密

- 只增加了系统的复杂性，没有从根本上改变基本性质
- 复杂密码系统可以由多个基本密码系统构成

可能的密钥与实际的密钥同等重要

- 存在可能的密钥，将密文映射为与明文不同的有意义消息。正是这些可能的密钥提供了密码系统的安全性
- 合法解密者知道实际的密钥，可以确认明文；窃听者只知道可能密钥的先验概率，无法确认真实明文
- 例如，策略游戏（如象棋）中，可能的威胁与实际威胁同等重要

信息量

○ 什么是信息？

- 信息是消息的有效内容;信息蕴涵于事件的不确定性之中

◆ 举例：

○ 事件：

- 明年学校男生比女生多——几乎是必然的，信息量趋于零
- 明年学校女生比男生多——可能性很小，信息量极大

◆ 概率分布：明年学校男生和女生谁更多的信息量

- 入学前，有大于/小于/等于三种可能，存在不确定性
- 入学后，仅存一种结果，信息不确定性降为零
- 信息量=消息获得前的不确定性

信息量的度量——熵

- 信息量 $H(x)$ ，是事件概率 $p(x)$ 的函数，满足：
 - ◆ 是概率 $p(x)$ 的单调递减函数
 - ◆ 信息量 $H(x)$ 非负
 - ◆ 当概率 $p(x)=1$ 时，信息量 $H(x)=0$
 - ◆ 独立事件的信息量是事件信息量之和
- 定义：
 - 随机事件 x 的信息量 $H(x)=-\log p(x)$
 - 随机分布 X 的信息量 $H(X)=\sum p(x)H(x)=-\sum p(x)\log p(x)$
 - 又称为信源的信息量，信源的熵
 - 实验前，表示信源平均信息量，即平均不确定度
 - 实验后，表示平均获得的信息量，不确定度降为零

举例

○ 性别消息

- ◆ 消息：男，女；假设概率各1/2
- ◆ 信息量：“男”和“女”的信息量各 $\log_2 2 = 1$ bit
- ◆ 信源的熵 $1/2 \log_2 2 + 1/2 \log_2 2 = 1$ bit

- ◆ 消息：男，女；其中男的概率3/4，女的概率1/4
- ◆ 信息量：“男”的信息量 $\log_2 4/3 \approx 0.4$ bits
 - “女”的信息量 $\log_2 4 = 2$ bits
 - 信源的熵 $3/4 \log_2 4/3 + 1/4 \log_2 4 \approx 0.8$ bits

条件熵

- 举例：明年学校男生比女生多

◆ 信息的多少取决于哪个“学校”，即条件

- 条件概率

$$p_x(y) = \frac{p(x, y)}{p(x)} = \frac{p(x, y)}{\sum_y p(x, y)}$$

- Y的条件熵定义：在每个x条件下，Y的熵的平均值

$$\begin{aligned} H_X(Y) &= -\sum_x [p(x) H_x(Y)] = -\sum_x \left[p(x) \left(\sum_y p_x(y) \log p_x(y) \right) \right] \\ &= -\sum_{x,y} p(x, y) \log p_x(y) \end{aligned}$$

消息中的冗余：冗余度

- 冗余，为了描述信息而包含在消息中的多余部分
- 定义：长度为 N 比特的消息 M ，假设它的信息量为 $H(M)$ ，则其冗余度为：

$$D = 1 - \frac{H(M)}{N}$$

- 总冗余为

$$D_N = N - H(M)$$

- 当消息集中消息总数为 G 时，信源编码为 $\log G$ 比特。
总冗余为

$$D_N = \log G - H(M)$$

举例

冗余度

○ 据统计

- ◆ 英语冗余度的上限为80%，下限为67%，平均值为73%；
- ◆ 俄语的冗余度平均值约为70%；
- ◆ 现代汉语冗余度的上限为73%，下限为55%，平均值为63%，文言文的冗余度就更低了

○ 例：全班45人的成绩单，

- ◆ 信息量： $45\log_2 101 \approx 299.62$ 比特 < 38字节
- ◆ 以txt文本存储，约需134字节，冗余度0.72
- ◆ 将该文件用RAR压缩，约需91字节，冗余度0.58

○ 压缩就是减少冗余

○ 冗余度给出无损压缩比的极限 $1/(1-D)$

理论安全

- 密码系统的安全性
- 理论安全：密码分析员有无限的时间和人力，仍无法破解
- 实际安全：密码分析员无法在有限时间内，使用有限的人力破解
 - 破译的成本超过该信息的价值
 - 破译的时间超过该信息的有用生命周期

完美安全 PERFECT SECRECY

- 含义：窃听者截获的密文不能提供任何信息
- 消息（或密文）的后验概率等于其先验概率

$$P_c(m) = \frac{P(m)P_m(c)}{P(c)}$$

$$P_c(m) = P(m) \quad \text{or} \quad P_m(c) = P(c)$$

完美安全的充分必要条件

- 对所有的消息 m 和密文 c ，都有 $P_m(c) = P(c)$ ，即 $P_m(c)$ 独立于 m ：
 - ◆ 换言之，对任意 m_i ， m_j 和 c 都有：将 m_i 映射到 c 的总概率，等于将 m_j 映射到 c 的总概率
 - ◆ 考虑 $P_m(c) = P(c) \neq 0$ ，可以从每个 m 映射到每个 c 。
 - 消息数等于密文数、密钥数等于消息数
 - 应用可能性？

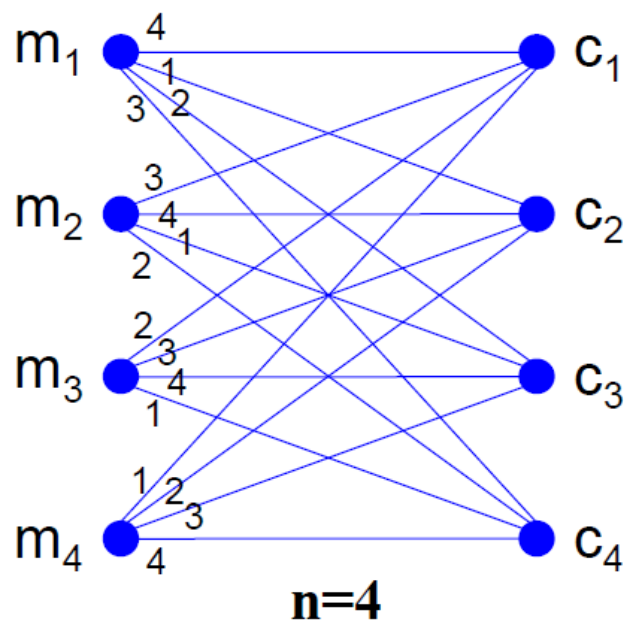
完美安全系统的构造例子

- 设有 n 个消息、 n 个密文和 n 个映射，标记为 m_i, c_i, T_i ，令：

$$T_i m_j = c_s \quad i + j = s \pmod{n}$$

- 例中

$$P(c) = P_c(m) = \frac{1}{n}$$



密钥量需求

- 信源的信息量为： $H(M) = -\sum P(m) \log P(m)$
- 信源包含 n 个等概消息时，一条消息的信息量至多为 $\log n$

- 密钥集的信息量或不确定度为：

$$H(K) = -\sum P(k) \log P(k)$$

- 为掩盖上述明文信息所需密钥的不确定度至少为 $\log n$ ，即至少有 n 个等概密钥。
- 当明文集为无限集合时，完美安全系统所需密钥集也必须为无限集合

完美安全系统：应用场合

- 一般用于加密最重要的信息，或者消息集很小的场合。
- 缺点：与信息等量密钥的产生与传递

模糊度

○ 密码分析的概率模型：

- 在截获消息前，可以给每个消息和密钥设定一个先验概率；
- 待截获长度为N的消息后，计算相应的后验概率；
- 通常，随着N的增加，多数消息的后验概率降低，少数增加，直至最后只剩下一个消息后验概率接近于1，其它接近0。
- 凯撒密码作用在英文文本上的实例及分析

Decipherments	N = 1	N = 2	N = 3	N = 4	N = 5
C R E A S	.028	.0377	.1111	.3673	1
D S F B T	.038	.0314			
E T G C U	.131	.0881			
F U H D V	.029	.0189			
G V I E W	.020				
H W J F X	.053	.0063			
I X K G Y	.063	.0126			
J Y L H Z	.001				
K Z M I A	.004				
L A N J B	.034	.1321	.2500		
M B O K C	.025		.0222		
N C P L D	.071	.1195			
O D Q M E	.080	.0377			
P E R N F	.020	.0818	.4389	.6327	
Q F S O G	.001				
R G T P H	.068	.0126			
S H U Q I	.061	.0881	.0056		
T I V R J	.105	.2830	.1667		
U J W S K	.025				
V K X T L	.009				
W L Y U M	.015		.0056		
X M Z V N	.002				
Y N A W O	.020				
Z O B X P	.001				
A P C Y Q	.082	.0503			
B Q D Z R	.014				
H(decimal digits)	1.2425	.9686	.6034	.285	0

模糊度

- 获得一定密文后，如何估计不确定度？
- 密码系统中的模糊度
- **消息模糊度**：
$$H_C(M, N) = -\sum_{c,m} P(c, m) \log P_c(m)$$
- 求和范围是所有长度为N的消息和密文
- **密钥模糊度**：
$$H_C(K, N) = -\sum_{c,k} P(c, k) \log P_c(k)$$
- 求和范围是所有密钥和所有长度为N的密文
- 都是N的函数，简写为 $H_C(M)$ 、 $H_C(K)$

唯一解距离

- 密钥的唯一解距离：唯密文攻击情况下，使 $H_c(K)$ 接近为零的最小密文长度 N_{UK} ：

$$H(K) - H_c(K) \leq D_N$$

$$H(K) = DN_{UK}$$

$$N_{UK} = \frac{H(K)}{D}$$

- 消息的唯一解距离：当 $H_c(M)$ 接近零的最短密文长度 N 称为消息的 **唯一解距离** N_{UM} 。

密码破译的确认

○ 冗余度的作用

- ◆ 消息中，字符间、上下文存在的一系列规则，产生冗余
- ◆ 密文中冗余被分散，但仍存在。累计足够多的密文将保证只有一对消息和密钥满足这些规则，此时破译成功

○ 唯一解距离是针对唯密文攻击所做的分析。实际操作中往往会采用更有效的方式

○ 唯密文攻击时，所需的密文长度通常远大于唯一解距离

○ 当宣称某种密码系统和密钥被破译时：

- ◆ 若使用的密文长度远大于唯一解距离，则可信；
- ◆ 若使用的密文长度相当于或小于唯一解距离，则很可疑。

人的参与

○ 方法及优点

- 剔除元音字母和其它不会导致歧义的字母，必要时修改少量字母，然后再加密
- 降低冗余（大约3或4倍到1倍），唯一解距离也会放大相应倍数
- 实际上是将解密者的语言能力也做为解密器的一部分
- 目前机器智能有限，难以模仿人脑智能，该措施可以有效地增加破译难度

○ 缺点

- 解密时间长，对合法解密者的语言能力要求较高，不利于普及
- 加密者对“不会导致歧义的字母”的判断未必正确，可能导致无法唯一解密

实际安全

- ◆ 工作特性
- ◆ 强力搜索攻击
- ◆ 统计攻击
- ◆ 可能词攻击
- ◆ 算法设计要点总结

工作特性

- 定义：工作特性是从 N 个字符密文确定密钥所需的平均（对所有消息和密钥）**工作量**（单位：工时），是实际安全的一个量化度量
- 实际安全系统，要求在它所希望传输的字符量范围内工作特性足够高。
- 具有有限资源的攻击者在合理的时间内不能破译系统

密码系统的设计要求

- 设一个好的密码系统，必须将最小工作量最大化
- 不仅要考虑标准密码分析方法，要确保没有任何捷径可以破译密码——很难！
- 如何确认一个非理想系统的唯一解距离足够大，用任何方法分析都需要极大的工作量？
- 研究密码分析员可能使用的每一种方法，总结出抵制规律，在设计密码系统时应用这些规律
- 设计系统，使得它的破解工作等价于某些复杂性问题
- 下面讨论密码分析员常用的分析方法

练习-2

1. 读书报告：论述密码编码的发展和密码破译的发展。
2. 下次上课时交。

下次内容

◆ 古典密码学