网络空间安全数学基础（1）

网络空间安全学院

高莹

2020年 9 月 23 日-30日

课程内容安排

- 第1章 保密系统的理论基础
- 第2章 椭圆曲线基础
- 第3章 格理论基础
- 第4章 图论基础
- 第5章 博弈论基础

参考教材

- 第1章 保密系统的信息论基础：《密码学原理与实践（第三版）》，2009年电子工业出版社，作者（加）斯廷森（Stinson，D.R）
- 第2章 椭圆曲线：An introduction to mathematical cryptography
- 第3章 格： An introduction to mathematical cryptography
- 第4章 图论基础：图论教材待选，任意一本可提前参考
- 第5章 博弈论基础：博弈论教程，奥斯本、鲁宾斯坦著，中国社会科学出版社，2000年

Ch1 保密系统的理论基础

- 1948年Shannon在贝尔实验室技术期刊上发表文章“通信的数学理论”，通过量化度量精确描述事件的信息、熵和时间集合的信息和熵，以及两个事件之间的条件信息、互信息等。通过这些量化的描述，给出了信源编码定理、信道容量定理和信道编码定理，从而奠定了信息论的理论基础。
- 1949年，Shannon又在该期刊发表了“保密系统的信息理论”，该论文奠定了现代密码学的系统研究的理论基础。
- 这两篇论文开辟了现代通信和保密通信的科学研究方向。

本章主要内容

- 1.概率论基础
- 2.熵及其基本性质
- 3.完善保密性
- 4.伪密钥与唯一解距离
- 5.扩展的欧几里得算法（整数和多项式）
- 6.乘积密码体制
- 7.AES中的数学原理

1.1 概率论基础

- 一个试验可能产生多个结果，每个结果称为一个简单事件，所有可能结果的集合称为样本空间。根据需要，我们只考虑有限多个可能结果的离散样本空间。
- 定义1：样本空间S上的一个离散随机变量X，用$P(X=x)$表示随机变量取 x时的概率，简记为$P(x)$，对于任意的$x\in S$，则有$0\le P(X=x)\le 1$，而且可以得到

$$\sum_{x\in S} P(x)=1$$

- 定义2：一个事件E是样本空间S的一个子集，事件发生的概率记为$P(E)$。特别的，当E是一个简单事件x时$P(E)=P(x)$。事件E发生的概率$P(E)$为：

$$P(E)=\sum_{x\in E}P(x)$$

- 定义3：假设X和Y分别是定义在样本空间S_1和S_2上的随机变量。联合概率$P(x,\ y)$是X取x且Y取y时的概率。条件概率$P(x/y)$表示当Y取y时X取x的概率。如果对于任意的$x\in S_1$和$y\in S_2$,有$P(x,\ y)=P(x)\ P(y)$,则称随机变量X和Y是统计独立的。而且对于联合概率和条件概率存在以下关系：

$$P(x,y)=P(x/y)P(y)\qquad P(x,y)=P(y/x)P(x)$$

- 贝叶斯定理（Bayes定理）：如果$P(y)>0$，那么

$$P(x/y)=\frac{P(y/x)P(x)}{P(y)}$$

- 由定义3的条件概率和联合概率的两个关系，直接可得贝叶斯定理（Bayes定理）。而且由定义3可得推论：
- 推论1：X和Y是相互独立的随机变量，当且仅当对所有的$x\in S_1$和$y\in S_2$，有$P(x/y)=P(x)$。
- 定义4：设S是一个样本空间，X是S上的一个随机变量，且X是一个从样本空间S到实数集R的函数；对于每一个简单事件$x\in S$，X分配一个实数$X(x)$。X的数学期望定义为 $E(X)=\sum_{x\in S}X(x)P(x)$

本章主要内容

- 1.概率论基础
- 2.熵及其基本性质
- 3.完善保密性
- 4.伪密钥与唯一解距离
- 5.扩展的欧几里得算法（整数和多项式）
- 6.乘积密码体制
- 7.AES中的数学原理

1.2 熵及其基本性质

1.2.1 信息论的相关概念

单符号离散信源：如果信源发出的消息是离散的、有限或无限可列的符号或数字，且一个符号代表一条完整的消息，则称这种信源为单符号离散信源。

信源空间：若信源的输出是随机事件x，其出现概率为$p(x)$，则它们所构成的集合，称为信源的概率空间或简称为信源空间。

自信息量：一个随机事件的自信息量定义为其出现概率对数的负值。即

$$I(x_i) = -\log p(x_i)$$

- 例：一个等概率的二进制随机序列，求任一码元的自信息量。

  解：因为二进制序列只有0和1，而且等概率$P(0)=P(1)=1/2$，所以有

$$I(0)=I(1)=-\log_2(1/2)=\log_2 2=1\ \text{bit}$$

- 例：对于n位的2进制数，假设每一符号的出现完全随机且概率相等，求任一符号的自信息量。

  解：因为对于一个n位的2进制数的每一位可以从0，1两个数字中任取一个，所以有2^n个等概率的可能组合。所以，

$$p(x_i)=1/2^n$$

$$I(x_i)=-\log_2 P(x_i)=-\log_2(1/2^n)=n\ \text{bit}$$

1.2.2 熵的定义

- 在数学中，事件的不确定性可用不确定度来描述，它同样是事件概率的函数，同时在数值和量纲上与自信息量相等。
- 根据日常知识，各个出现概率不同的随机事件所包含的不确定度是有差别的。一个出现概率接近于1的随机事件，发生的可能性很大．所以它包含的不确定度就很小。
- 反之，一个出现概率很小的随机事件，很难猜测在某个时刻它能合发生，所以它包含的不确定度就很大。

1.2.2 熵的定义

- 若是确定性事件，出现概率为1，则它包含的不确定度为0。显然随机事件的信息量和不确定度有很密切的联系，如果发生一个不确定度小的事件，则带来的信息量较小；反之，如果发生一个不确定度高的事件，它带来的信息量很大；如果是必然事件，则没有信息量。
- 虽然事件的自信息量和该事件的不确定度有如此密切的关系，但两者的含义却有本质的区别。
- 不确定度只与事件的概率有关，是一个统计量具有某种概率分布的随机事件不管发生与否，都存在不确定度，不确定度表证了该事件的特性。

1.2.2 熵的定义

- 而自信息量只有该随机事件出现时才给出，不出现时不给出，因此自信息量是在该事件发生后给予观察者的信息量。
- 联合自信息量：若有两个消息x_i，y_j同时出现，可用联合概率$p(x_i, y_j)$表示，这时的联合自信息量定义为$I(x_i, y_j) = -\log_2 p(x_i, y_j)$。
- 当x_i和y_j相互独立时，有$p(x_i, y_j) = p(x_i)\, p(y_j)$，那么根据对数运算的性质，有

$$I(x_i, y_j) = I(x_i) + I(y_j),$$

x_i、y_j所包含的不确定度在数值上也等于它们的自信息量。

1.2.2 熵的定义

- 若两个消息出现不是独立的，但他们是有相互联系的，这时可用条件概率$p(x_i/y_j)$来表示，即在事件y_j出现的条件下，随机事件x_i发生的条件概率，则它的**条件自信息量**定义为条件概率对数的负值：

$$I(x_i/y_j)=-\log_2 p(x_i/y_j)。$$

- 虽然信息函数$I(x_i)$使得信息度量成为可能，是信息度量的有力工具，但在信息度量方面仍然存在某些不足。

- 首先，信源产生符号x_i不是确定事件，是以$p(x_i)$为概率的随机事件，相应的自信息量$I(x_i)$也是一个以$p(x_i)$为概率的随机性的量。显然，用一个随机性的量来度量信息是不方便的。

1.2.2 熵的定义

- 其次，信息函数$I(x_i)$只能表示信源发某一特定的具体符号x_i所提供的信息量。不同的符号x_i，有不同的自信息量。所以它不足以作为整个信源的总体信息测度。
- 能作为信源总体信息测度的确定的量，应是信源x可能发出的各种不同符号x_i $(i=1, 2, ..., r)$含有的自信息量$I(x_i)$ $(i=1, 2, ..., r)$，在信源的概率空间$(p(x_1), p(x_2), p(x_3), ..., p(x_r))$中的统计平均值。

1.2.2 熵的定义

- 为了指明是信源x的信息测度，我们把这个统计平均值记为$H(x)$，定义信源的平均不确定度$H(x)$为信源中各个符号的不确定度的数学期望。即
- $$H(X)=E[I(X)]=\sum_i p(xi)I(xi)=-\sum_i p(xi)\log p(xi)$$
- 其单位为比特/符号或者比特/符号序列。
- **信息熵：**我们称$H(x)$是信源x的“信息熵”。它表示信源x每发一个符号（不论发什么符号）所提供的平均信息量

1.2.2 熵的定义

- 例：设信源符号集$X=\{x_1, x_2, x_3, x_4\}$每个符号发生的概率分别为$p(x_1)=1/2$，$p(x_2)=1/4$，$p(x_3)=1/8$，$p(x_4)=1/8$，求信源熵。

  解：$H(X)=1/2\log 2+1/4\log 4+1/8\log 8+1/8\log 8$

  $=1/2+1/2+3/8+3/8$

  $=1.75$ 比特/符号

- 例：二元信息源的符号只有两个0和1，求其信息熵。

  解：设输出符号0和1的概率分别为$p(0)$，$p(1)$，则$p(0)+p(1)=1$。

  即信源的概率空间为：$\begin{pmatrix} X \\ P \end{pmatrix}=\begin{pmatrix} 0 & 1 \\ p(0) & p(1) \end{pmatrix}$

1.2.2 熵的定义

- (1) 当$p(0)=p(1)=1/2$时，信源X的信息熵

$$H(X)=-p(0)\log p(0)-p(1)\log p(1)$$
$$=-1/2\log(1/2)-1/2\log(1/2)$$
$$=1/2+1/2=1 \text{ 比特/信源符号}$$

- (2) 当$p(0)=0$，$p(1)=1$时，信源X的信息熵

$$H(X)=-p(0)\log p(0)-p(1)\log p(1)$$
$$=-0\log 0-1\log 1$$
$$=0$$

注：规定$\log 0=0$

- (3) 当$p(0)=1$，$p(1)=0$时，同样可得信源X的信息熵

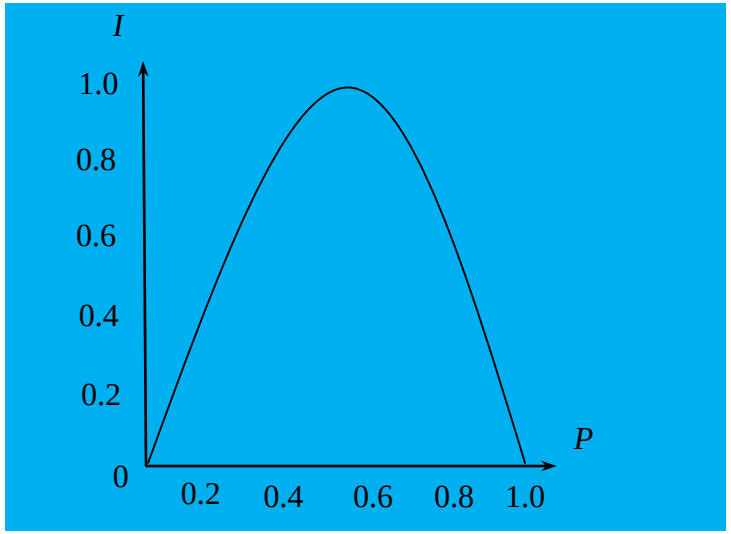
$$H(X)=-p(0)\log p(0)-p(1)\log p(1)$$
$$=-1\log 1-0\log 0$$
$$=0$$

- 如果以$p(0\le p\le 1)$表示p (0)，以q表示p (1)，则二元信源X的信息熵为：

$$H(X)=-p\log p-q\log q$$
$$=-p\log p-(1-p)\log(1-p)$$
$$=H(p)$$

1.2.2 熵的定义

- 在该例子中，信源信息熵$H(X)$是概率p的函数，通常用$H(p)$表示，p取值于$[0,1]$区间。$H(p)$的函数如图所示。



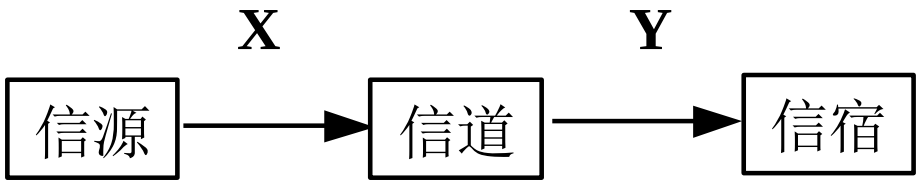
熵函数$H(p)$

- 从图中可看出，如果二元信息源的输出符号是确定的，即$p=1$或$q=1$，则该信源不提供任何信息。反之，当二元信息符号0和1以等概率发生时，信源熵达到极大值，等于1比特信息量。

1.2.2 熵的定义

- 熵是在平均意义上来表征信源的总体特性的。
- 正如不确定度与自信息量的关系那样，信源熵是表征信源的平均不确定度，平均自信息是消除信源不确定度时所需要的信息的量度，即收到一个信源符号，全部解除了这个符号的不确定度。或者说获得这样大的信息量后，信源不确定度就被消除了。两者在数值上相等，但含义不同。
- 某一信源，不管它是否输出符号，只要这些符号具有某些概率特性，必有信源的熵值；这熵值是在总体平均上才有意义，因而是一个确定值。
- 一般写成$H(X)$，X是指随机变量。而另一方面，信息量则只有当信源输出符号而被接收者收到才有意义，这就是给予接收者的信息度量 。

1.2.3 条件熵的定义



最简单的通信系统模型

- 在给定y_j条件下，x_i的条件自信息量为$I(x_i/y_j)$，对X集合的条件熵$H(X/y_j)$为：

$$H(X/y_j)=\sum_i p(x_i/y_j)I(x_i/y_j)$$

1.2.3 条件熵的定义

- 进一步，在给定Y（即各个y_j）条件下，X集合的条件熵$H(X|Y)$定义为

$$H(X/Y)=\sum_j p(y_j)H(X/y_j)=\sum_{i,j} p(y_j)p(x_i/y_j)I(x_i/y_j)$$

$$=\sum_{i,j} p(x_i,y_j)I(x_i/y_j)$$

- 即条件熵是在联合符号集合$X\times Y$上的条件自信息量的联合概率加权统计平均值。条件熵$H(X|Y)$表示已知Y后，X的不确定度。相应地．在给定X（即各个x_i）条件下，Y集合的条件熵$H(Y/X)$定义为：

$$H(Y/X)=\sum_{i,j} p(x_iy_j)I(y_j/x_i)=\sum_{i,j} p(x_i,y_j)\log p(y_j/x_i)$$

1.2.4 联合熵的定义

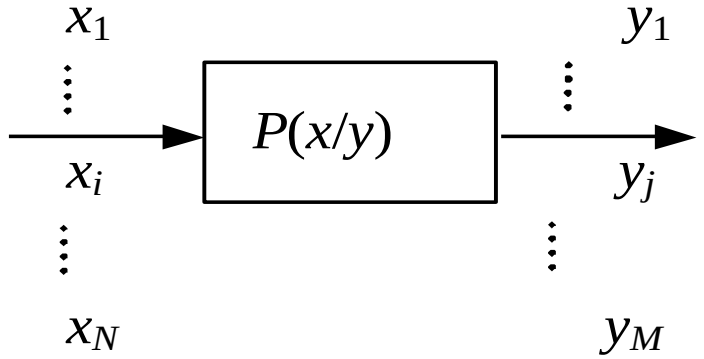
- 联台熵是联合符号集合$X\times Y$上的每个元素对$(x_i，y_j)$的自信息量的概率加权统计平均值，定义为：
- $H(X,Y)=\sum_{i,j}p(x_i,yj)I(xi,yj)=-\sum_{i,j}p(x_i,y_j)\log p(x_i,y_j)$
- 联合熵$H(X,Y)$表示X和Y同时发生的不确定度。

1.2.5 互信息的定义

- 若信道存在干扰，信宿收到从信道输出的某一符号y_j后，能够获取多少关于从信源发某一符号x_i的信息量?
- 当信道存在干扰时，信源发x_i，信宿收到的y_j可能是x_i的某种变型，亦即除了信源给出的信息外，还可能有纯粹是信道给出的“信息”。利用前述条件自信息量的概念可以得知，在收到y_j后，考虑从发端发x_i这一事件中获得的信息量，应该是

$$I(x_i;\ y_j) = I(x_i) - I(x_i / y_j)$$

$$I(x_i;\ y_j) = \log \frac{P(x_i / y_j)}{P(x_i)}$$



1.2.5 互信息的定义

- （1）对称性 $$I(x_i;\ y_j)=\log\frac{P(x_i/y_j)}{P(x_i)}=\log\frac{P(x_i/y_j)P(y_j)}{P(x_i)P(y_j)}$$

$$=\log\frac{P(x_iy_j)/P(x_i)}{P(y_j)}\ =\log\frac{P(y_j/x_i)}{P(y_j)}=I(y_j;\ x_i)$$

- （2）值域为实数
- 非平均互信息量的值可为正数、或者0或负数

1）$P(xi/y_j)=1$，$I(xi;\ y_j)=I(xi)$。

说明收到y_j后即可以完全消除对信源是否发x_i的不确定度，其物理含义是信宿获取了信源发出的全部信息量，这等效为信道没有干扰。

1.2.5 互信息的定义

2）$P(xi) < P(xi/y_j) < 1$，这时$I(xi) > I(xi/y_j)$，$I(xi;\ y_j) > 0$。

说明收到y_j后对信源是否发x_i所进行判断的正确程度，要大于x_i在信源集合中的概率，因此y_j获取了关于x_i的信息量。$I(xi;\ y_j)$越大，这种获取越多。

这正是实际通信时遇到的大多数情况，它对应着信道存在干扰但信宿仍能从信源中获取信息量的情况。

3）$P(xi/y_j) = P(xi)$，亦即$I(xi) = I(xi/y_j)$，$I(xi;\ y_j) = 0$

说明收到y_j后对信源是否发x_i所进行判断的正确程度，和x_i在信源集合中的概率是一样的，直观上不难理解这时在y_j中获取不到关于x_i的信息量。

事实上，假若x_i和y_j统计无关，即$P(xi，\ y_j)=P(xi)\,P(y_j)$，由贝叶斯公式容易推得$I(xi;\ y_j)=0$。这也说明这种情况实际上是事件$x_i$和事件$y_j$统计无关，或者说信道使得事件$x_i$和事件$y_j$变成了两码事，信宿得到的信息仅仅是由信道特性给出的，与信源实际发出什么符号无关，因此完全没有信息的流通。

4）$0<P(x_i/y_j)<P(x_i)$，亦即$I(x_i)<I(x_i/y_j)$，

$I(x_i;\ y_j)<0$

说明收到yj后对信源是否发xi所进行判断的正确程度，比xi在信源集合中的概率还要小，这时判断信源没有发xi似乎更合理些，但不能判断信源到底发了什么（特别是对应于信源有多个符号时）。这种情况事实上给出了信息量，但流通的不是关于xi的信息量，而是xi以外的事件的信息量。

综上所述，只有$P(x_i/y_j)=P(x_i)$，即$I(xi;\ yj)=0$时，才没有信息的流通。

（3）不大于其中任一事件的自信息量

由于$P(xi/y_j)\leq 1$，有

$$I(xi;\ y_j)\leq \log[1/P(xi)] = I(xi)$$

同理，由$P(xi/y_j)\leq 1$，有

$$I(y_j;\ x_i)\leq \log[1/P(y_j)] = I(y_j)$$

这一性质清楚地说明了非平均互信息量是描述信息流通特性的物理量，流通量的数值当然不能大于被流通量的数值。由于自信息量是为了确定某一事件出现所必须提供的信息量，因此，这一性质又说明某一事件的自信息量是任何其他事件所能提供的关于该事件的最大信息量。

- 非平均互信息量定量地描述了信息的流通问题，但它只描述了集合$\boldsymbol{X}$（信源）中某一具体符号x_i与集合$\boldsymbol{Y}$（信宿）中某一具体符号y_j通过某一媒质（信道）的信息流通情况，还不能作为信道上信息流通的整体测度。类似地，若从整体的角度且在平均意义上来度量信宿每接收到一个符号而从信源获取的信息量，就要引入平均互信息量的概念。
- 两个离散随机事件集合$\boldsymbol{X}$和$\boldsymbol{Y}$，若任意两个事件间的互信息量为$I(xi;\ y_j)$，则用其联合概率进行加权的统计平均值，称为两集合的**平均互信息量**，用$I(\boldsymbol{X};\ \boldsymbol{Y})$表示。

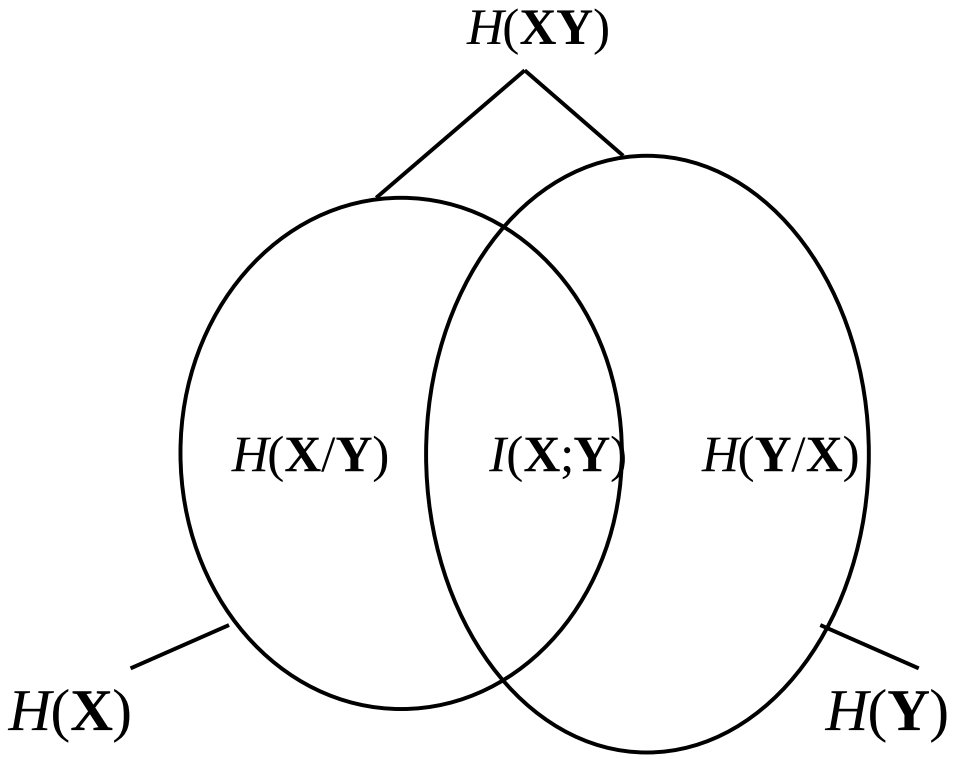
- 事实上，当信宿收到某一具体符号y_j后，从y_j中获取关于输入符号（不论是哪一个符号）的平均信息量，显然应该是在条件概率空间中的统计平均，可以用$I(\boldsymbol{X};\ y_j)$表示，有

$$I(\mathbf{X};\ y_j)=\sum_{i=1}^{N}P(x_i/y_j)I(x_i;\ y_j)=\sum_{\mathbf{X}}P(x/y_j)I(x;\ y_j)$$

再对其在集合$\mathbf{Y}$中取统计平均，得

$$\begin{aligned}I(\mathbf{X};\mathbf{Y})&=\sum_{j=1}^{M}P(y_j)I(\mathbf{X};y_j)\\&=\sum_{i=1}^{N}\sum_{j=1}^{M}P(y_j)P(x_i/y_j)\log\frac{P(x_i/y_j)}{P(x_i)}\\&=\sum_{\mathbf{XY}}P(xy)\log\frac{P(x/y)}{P(x)}=\sum_{\mathbf{XY}}P(xy)I(x;y)\end{aligned}$$

各种信息量之间的关系



1.2.6 熵的基本性质

◈ 性质1：非负性

$$H(X)=H(x_1, x_2, \ldots, x_n)\geq 0$$

其中等号只有在$x_i=1$时成立。因为$0\leq p(x_i)\leq 1$，则$\log p(x_i)$一定是一个负数，所以$H(X)$是非负的。

◈ 性质2：对称性

熵函数所有变量可以互换，而不影响函数值。即

$$H(x_1, x_2, \ldots, x_n)=H(x_2, x_1, \ldots, x_n)$$

1.2.6 熵的基本性质

- 因为熵函数只与随机变量的总体结构有关，例如下列信源的熵都是相等的：

$$\begin{pmatrix} X \\ P \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 \\ 1/3 & 1/6 & 1/2 \end{pmatrix} \begin{pmatrix} Y \\ P \end{pmatrix} = \begin{pmatrix} y_1 & y_2 & y_3 \\ 1/3 & 1/6 & 1/2 \end{pmatrix}$$

- 性质3：确定性

$$H(0,1) = H(1,0,0,\ldots,0) = 0$$

也就是说只要信源符号表中，有一个符号的出现概率为1，信源熵就等于零。

在概率空间中，如果有两个基本事件，其中一个是必然事件，另一个则是不可能事件，因此没有不肯定性，熵必为零。当然可以类推到n个基本事件构成的概率空间。

1.2.6 熵的基本性质

- 下面为了讨论熵的其它性质，首先给出关于凸函数的一个结论。
- 凸函数：一个实值函数f称为在区间I上是凸的，如果对于任意x，$y\in I$，$x\neq y$，都有

$$\frac{f(x)+f(y)}{2}\leq f\left(\frac{x+y}{2}\right)$$

- 函数f称为在区间I上是严格凸的，如果对任意x，$y\in I$，$x\neq y$，都有

$$\frac{f(x)+f(y)}{2}<f\left(\frac{x+y}{2}\right)$$

- Jensen不等式：设f是区间I上的一个连续的单调严格凸函数，

1.2.6 熵的基本性质

$$\sum_{i=1}^{n} a_i = 1 \qquad a_i > 0,$$

则

$$\sum_{i=1}^{n} a_i f(x_i) \le f\left(\sum_{i=1}^{n} a_i x_i\right)$$

其中$x_i \in I$，$1 \le i \le n$。当上式中的等号成立时，当且仅当$x_1 = x_2 = \ldots = x_n$。

- 容易证明，对数函数$f(x) = \log_2 x$在区间$I = (0, +\infty)$上是一个连续的严格凸函数。
- 现在我们继续给出关于熵的几个结论。下面的定理利用了对数函数$f(x) = \log_2 x$在区间$I = (0, +\infty)$上是严格凸的性质。

性质4：最大熵

- 假设X是一个随机变量，概率分布为p_1，p_2，……，p_n，其中$p_i>0$，$1\le i\le n$。那么$H(X)\le \log_2 n$，当且仅当$p_i=1/n$，$1\le i\le n$时等号成立。
- 证明：应用Jensen不等式，可得：

$$
\begin{aligned}
H(x) &= -\sum_{i=1}^{n} p_i \log_2 p_i \\
&= \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \\
&\le \log_2 \sum_{i=1}^{n} (p_i \times \frac{1}{p_i}) \\
&= \log_2 n
\end{aligned}
$$

- 当且仅当$p_i=1/n$，$1\le i\le n$，等号成立。

1.2.6 熵的基本性质

- 性质5： $H(X,Y) \le H(X)+H(Y)$，当且仅当 X 和 Y 统计独立时等号成立。
- 证明：假设 X 取值 x_i，$1 \le i \le n$，Y 取值 y_j，$1 \le j \le m$。可得：

$$p(x_i)=\sum_{j=1}^{m} p(x_i,y_j), 1 \le i \le n$$

$$p(y_j)=\sum_{i=1}^{n} p(x_i,y_j), 1 \le j \le m$$

所以

$$H(X)+H(Y)=-[\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)+\sum_{j=1}^{m} p(y_j)\log_2 p(y_j)$$

$$=-\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,y_j)\log_2 p(x_i)p(y_j)$$

1.2.6 熵的基本性质

- 由联合熵的定义和Jensen不等式可知：

$$H(X,Y)-H(X)-H(Y)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log_2\frac{1}{p(x_i,y_j)}+\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log_2 p(x_i)p(y_j)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log_2\frac{p(x_i)p(y_j)}{p(x_i,y_j)}$$

$$\le\log_2\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j)=\log_2 1=0$$

等号成立当且仅当对任意的$1\le i\le n$和$1\le j\le m$，

$$\frac{p(x_i)p(y_j)}{p(x_i,y_j)}=C,$$

1.2.6 熵的基本性质

C是一个常数。因为

$$\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i)p(y_j)=1$$

所以$C=1$，即对任意的$1\le i\le n$和$1\le j\le m$，

$$p(x_i,y_j)=p(x_i)p(y_j)$$

因此，当且仅当X与Y相互独立等号成立，命题得证。

- 性质6：$H(X,Y)=H(Y)+H(X/Y)=H(X)+H(Y/X)$
- 证明：

1.2.6 熵的基本性质

$$H(X,Y)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log_2 p(x_i,y_j)$$

$$=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log_2 p(y_j)p(x_i/y_j)$$

$$=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log_2 p(y_j)-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log_2 p(x_i/y_j)$$

$$=-\sum_{j=1}^{m}p(y_j)\log_2 p(y_j)-\sum_{i=1}^{n}\sum_{j=1}^{m}p(y_j)p(x_i/y_j)\log_2 p(x_i/y_j)$$

$$=H(X)+H(X/Y)$$

- 同理可证明$H(X,Y)=H(X)+H(Y/X)$。

1.2.6 熵的基本性质

- 性质7： $H(X/Y)\leq H(X)$，当且仅当X与Y相互独立等号成立。
- 证明：由性质5，性质6：

$$H(X,Y)\leq H(X)+H(Y)$$

$$H(X,Y)=H(Y)+H(X/Y)=H(X)+H(Y/X)$$

即可得$H(X/Y)\leq H(X)$

1.2.6 熵的基本性质

例：假设明文$P=\{a,b\}$满足$p(a)=1/4$，$p(b)=3/4$，设密钥$K=\{k_1,k_2,k_3\}$，$p(k_1)=1/2$，$p(k_2)=1/4$，$p(k_3)=1/4$，设密文$C=\{1,2,3,4\}$，且加密矩阵如下：

	a	b
k_1	1	2
k_2	2	3
k_3	3	4

- 设p是明文空间P上的随机变量，c是密文空间C上的随机变量，k是密钥空间K上的随机变量。计算熵$H(p)$，$H(k)$，$H(c)$。

解：$H(p)=p(a)\log_2 p(a)-p(b)\log_2 p(b)=-(1/4)\times(-2)-(3/4)\times(\log_2 3-2)$

$=2-(3/4)\log_2 3=0.81$

$H(k)=-p(k_1)\times\log_2 p(k_1)-p(k_2)\times\log_2 p(k_2)-p(k_3)\times\log_2 p(k_3)$

$=(1/2)\times\log_2(1/2)-(1/4)+\log_2(1/4)-(1/4)\times\log_2(1/4)=1.5$

1.2.6 熵的基本性质

- 由于密钥的选取不依赖与明文，所以可以假设明文和密钥是相互独立的。要求$H(c)$必须先求出$p(1)$，$p(2)$，$p(3)$，$p(4)$可得：

- $p(1)=p(a)\times p(k_1)$
  $=(1/4)\times(1/2)$
  $=1/8$
- $p(2)=p(a)\times p(k_2)+p(b)\times p(k_2)$
  $=(1/4)\times(1/4)+(3/4)\times(1/4)$
  $=7/16$
- $p(3)=p(a)\times p(k_3)+p(b)\times p(k_3)$
  $=(1/4)\times(1/4)+(3/4)\times(1/4)$
  $=1/4$
- $p(4)=p(b)\times p(k_3)$
  $=(3/4)\times(1/4)$
  $=3/16$

- $H(c)=-(1/8)\log_2(1/8)-(7/16)\log_2(7/16)-(1/4)\log_2(1/4)-(3/16)\log_2(3/16)$
  $=1.85$

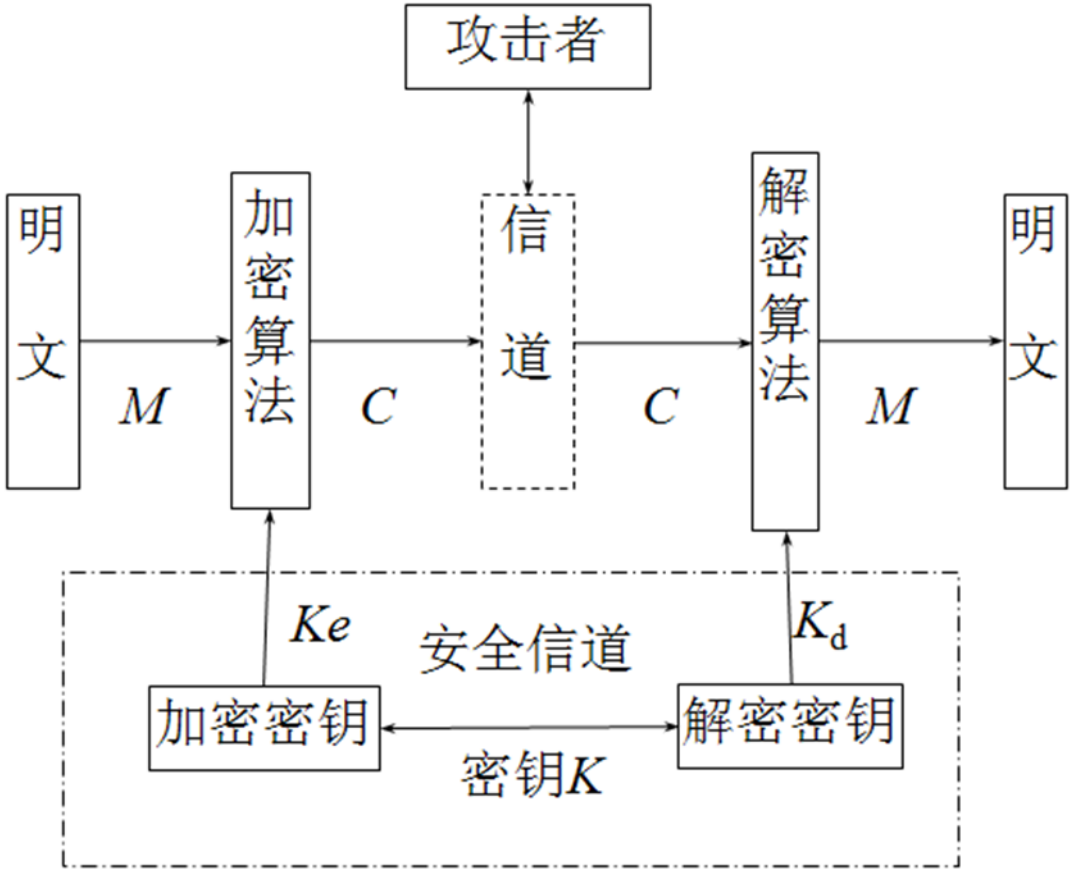
本章主要内容

- 1.概率论基础
- 2.熵及其基本性质
- 3.完善保密性
- 4.伪密钥与唯一解距离
- 5.乘积密码体制
- 6.AES中的数学基础

1.3 完善保密性

- 一个密码系统，通常简称为密码体制（Cryptosystem），由五部分组成：
- （1）明文空间M，它是全体明文的集合；
- （2）密文空间C，它是全体密文的集合；
- （3）密钥空间K，它是全体密钥的集合。其中每一个密钥K均由加密密钥K_e和解密密钥K_d组成，即$K=<K_e, K_d>$；
- （4）加密算法E，它是一组由M到C的加密变换；
- （5）解密算法D，它是一组由C到M的解密变换。

1.3 完善保密性



1.3 完善保密性

密码体制例子：移位密码

其加密变换为：

$E_k(m)=(m+k)\ mod\ q=c \quad 0\le m,\ c<q,$

密钥空间为$K=\{k \mid 0\le k<q\}$，元素个数为q，其中有一恒等变换，即$k=0$

解密变换为：$D_k(c)=(c-k)\ mod\ q=m$

1.3 完善保密性

- **完全保密性：**

  设$R=(P, C, K, E, D)$是一个密码体制。如果对任意$x\in P$和任意$y\in C$，都有$p(x|y)=p(x)$。则称密码体制R具有完善的保密性能，或者称为完全保密性。

- 假设(P, C, K, E, D)是一个特定的密码体制，密钥$k\in K$只用于一次加密。假设明文空间P存在一个概率分布。这样就为明文元素定义了一个随机变量，用X表示。

1.3 完善保密性

- $p(X=x)$表示明文x发生的先验概率。同时假设以固定的概率分布选取密钥（通常密钥选取是随机的，因此所有的密钥都是等概率的，在这里假设不是等概率的）所以密钥也定义了一个随机变量，用K表示；$p(K=k)$表示密钥k发生的概率。假设密钥和明文是统计独立的随机变量。
- 同样可以把密文看成随机变量，用Y表示。通过明文X和密钥K可以计算出密文Y的概率$p\,(Y=y)$。
- 对于密钥$k\in K$，定义$C\,(k\,)=\{\,E_{\mathrm{k}}(x),x\in P\,\}$，

1.3 完善保密性

- 也就是说$C(k)$代表密钥是k时所有可能的密文。对于任意的$y\in C$，我们有：

$$p(Y=y)=\sum_{k,y\in C(k)} p(K=k)p(x=d_k(y))$$

- 同样，对于任意的$y\in C$和$x\in P$，可如下计算条件概率$p(Y=y/X=x)$（即给定明文x，密文y的概率）：

$$p(Y=y/X=x)=\sum_{k,x=d_k(y)} p(K=k)$$

- 由贝叶斯公式计算$p(X=x/Y=y)$（也就是给定密文y，明文x的概率）得

$$p(X=x/Y=y)=\frac{p(X=x)p(Y=y/X=x)}{p(Y=y)}=\frac{p(X=x)\times\sum_{k,x=d_k(y)} p(K=k)}{\sum_{k,y\in c(k)} p(K=k)p(x=d_k(y))}$$

1.3 完善保密性

- 由上述公式可知道，只要知道了概率分布就可以求出在给出密文情况下明文的概率。
- 例：假设明文$P=\{a, b\}$满足$p(a)=1/4$，$p(b)=3/4$，设密钥$K=\{k_1, k_2, k_3\}$，$p(k_1)=1/2$，$p(k_2)=1/4$，$p(k_3)=1/4$，设密文$C=\{1, 2, 3, 4\}$，则加密矩阵如下：

	a	b
k_1	1	2
k_2	2	3
k_3	3	4

1.3 完善保密性

- 解：计算密文分布如下：

$$p(1)=(1/2)(1/4)=1/8$$
$$p(2)=(1/2)(3/4)+(1/4)(1/4)=7/16$$
$$p(3)=(1/4)(3/4)+(1/4)(1/4)=1/4$$
$$p(4)=(1/4)(3/4)=3/16$$

则明文空间上的条件概率分布为：

$$p(a|1)=1 \qquad p(b|1)=0$$
$$p(a|2)=1/7 \qquad p(b|2)=6/7$$
$$p(a|3)=1/4 \qquad p(b|3)=3/4$$
$$p(a|4)=0 \qquad p(b|4)=1$$

- 可以发现，只有密文$y=3$时，$p(a)=p(a|3)$，$p(b)=p(b|3)$。也就是说对于密文$y=3$满足完善保密性的定义，但是对于其他的密文不满足。

1.3 完善保密性

- 定理：假设移位密码中的26个密钥等概率使用，则对任意概率分布的明文，移位密码都具有完全保密性。
- 证明：这里令$P=C=K=Z_{26}$，对于$0\le K\le 25$，加密函数e_k定义为$e_k(x)=(x+k) \bmod 26 (x\in Z_{26})$。先计算$C$上的概率分布。假设$y\in Z_{26}$，则

$$p(Y=y)=\sum_{k\in z_{26}} p(K=k)p(x=d_k(y))$$

$$=\sum_{k\in z_{26}} \frac{1}{26}p(x=y-k)$$

$$=\frac{1}{26}\sum_{k\in z_{26}} p(x=y-k)$$

1.3 完善保密性

- 现在固定y，值$(y-k) \bmod 26$构成Z_{26}的一个置换。因此有：

$$\sum_{k \in z_{26}} p(x=y-k)=\sum_{k \in z_{26}} p(X=x)=1$$

- 得到对于任意的$y \in Z_{26}$，$p(y)=1/26$。

- 接下来，对于任意的x和y，我们有

$$p(y/x)=p(k=(y-x) \bmod 26)=1/26,$$

- （这是因为对于任意的x和y，满足$e_k(x)=y$的惟一的密钥$k=(y-x) \bmod 26$），现在应用贝叶斯公式，很容易计算出：

$$p(x/y)=\frac{p(x)p(y/x)}{p(y)}=\frac{p(x)\frac{1}{26}}{\frac{1}{26}}=p(x)$$

- 所以这个密码体制是完善保密的。

1.3 完善保密性

- 上面证明了只要每一个明文字符都用一个新的随机密钥加密，移位密码就是不可破的。下面让我们讨论一下更一般的情形。
- 从前面可知，利用贝叶斯定理，对于所有的$x\in P$和$y\in C$，$p(x/y)=p(x)$，同样也可得对于任意的$x\in P$和$y\in C$，$p(y/x)=p(y)$。若假设对于所有的$y\in C$，$p(y)>0$，固定任意的$x\in P$，对于任意的$y\in C$，则有$p(y/x)=p(y)>0$。
- 因此，对于任意$y\in C$，至少存在一个密钥k满足$e_k(x)=y$，这样就是有$|K|\geq|C|$。

1.3 完善保密性

在任意一个密码体制中，加密函数都是单射的，因此有$|C| \geq |P|$。

- 设(P, C, K, E, D)是一密码体制，满足$|P| = |C| = |K|$，该密码体制是完全保密当且仅当每一密钥被等概率的使用，且对任意明文x和密文y，存在唯一密钥k，将x加密成y。
- 证明：假设这个密码体制是完善保密的。由上面可知，对于任意的$x \in P$和$y \in C$，一定至少存在一个密钥k满足$e_k(x) = y$。因此有不等式：

$$|C| = |\{e_k(x), k \in K\}| \leq |K|$$

1.3 完善保密性

- 但是我们假设$|C|=|K|$，因此一定有：

$$\left|\{e_k(x), k \in K\}\right| = |K|$$

- 也就是说，不存在两个不同的密钥k_1和k_2使得

$$e_{k_1}(x) = e_{k_2}(x) = y$$

- 因此对于$x \in P$和$y \in C$，刚好存在一个密钥k使得$e_k(x)=y$。

- 记$n=|K|$，设$P=\{x_i, 1 \le i \le n\}$并且固定一个密文$y \in C$。设密钥为k_1，k_2，……，k_n，并且

$$e_{k_i}(x_i) = y \qquad 1 \le i \le n$$

使用Bayes定理，我们有

$$p(x_i / y) = \frac{p(y / x_i)p(x_i)}{p(y)} = \frac{P(K = k_i)P(x_i)}{P(y)}$$

1.3 完善保密性

- 考虑完善保密的条件$p(x_i|y)=p(x_i)$。在这里，我们有$p(k_i)=p(y)$，$1\le i\le n$。也就是说，所有的密钥都是等概率使用的。密钥的数目为K，我们得到对任意的$k\in K$，$p(k)=1/|K|$。若两个假设的条件都是成立的，可得到密码体制且完善保密的。
- **定理：**设(P, C, K, E, D)一次一密加密体制，则其具有完全的保密性。
- 证明：

  假设$n\ge 1$是正整数，$P=C=K=(Z_2)^n$。

  对于$k\in(Z_2)^n$，定义$e_k(x)$为k和x模2的和。

1.3 完善保密性

因此，如果$x=(x_1, x_2, x_3, ..., x_n)$并且$k=(k_1, k_2, ..., k_n)$，则：

$$e_k(x)=(x_1+k_1, ..., x_n+k_n) \bmod 2$$

- 解密和加密是相同的。如果$y=(y_1, y_2, ..., y_n)$，则

$$d_k(y)=(y_1+k_1, ..., y_n+k_n) \bmod 2$$

- 由上面给出的定理，容易看出一次一密密码体制提供了完善保密性。
- 对一次一密密码体制，密码分析者无法只从密文获得关于明文或者密钥的任何信息，即使获得了相应的密文，也只能得到这些密文对应的密钥，而不能获得其它密文对应的明文或密钥。
- 但一次一密密码体制要求每传送一个明文，都必须产生一个新的密钥并通过一个安全的信道传给对方，这样给密钥管理带来了很大的麻烦。

习题

- 1.考虑一个密码体制$M=\{a, b, c\}$，$K=\{k_1, k_2, k_3\}$和$C=\{1, 2, 3, 4\}$。假设加密矩阵为

	a	b	c
k_1	2	3	4
k_2	3	4	1
k_3	1	2	3

已知密钥概率分布为：$p(k_1)=1/2$，$p(k_2)=p(k_3)=1/4$，且明文概率分布为$p(a)=1/3$，$p(b)=8/15$，$p(c)=2/15$，计算$H(M)$，$H(K)$，$H(C)$，$H(M|C)$，$H(K|C)$。

- 2. 考虑一个密码系统$\{P, K, C\}$。

  a. 说明为什么$H(P, K)=H(C, P, K)=H(P)+H(K)$。

  b. 假设这个系统具有完全保密。证明$H(C, P)=H(C)+H(P)$和 $H(C)=H(K)$-$H(K|C, P)$。

  c. 假设这个系统有完全保密，并且对每一个明文密文对，最多只有一个相应的密钥　能够加密。证明$H(C)=H(K)$。

- 3. 假设在“一次一密”密码中，密文 y 和y'（两个2进制的n元数组）是使用同一个密钥K，分别加密明文x和x'得到的。证明$x+x'=y+y' \bmod 2$。