概率论在密码学中的应用

□ 郑 娟 朱金伟 曾文军

（华中师范大学数学与统计学学院 湖北·武汉 430079）

摘 要：本文通过介绍密码学中的保密通信系统模型，分析密码体制的安全性、可靠性来挖掘隐藏在各个机制中的概率统计规律，从而揭示出概率论在密码学理论中发挥的重要作用。

关键词：生日攻击 统计量 密码分析

中图分类号：TP3 **文献标识码**：A **文章编号**：1007-3973（2008）07-089-02

1 引言

21世纪是信息时代，信息已成为社会发展的重要战略资源.在信息化社会中，信息安全将扮演极为重要的角色，它直接关系到国家安全、企业经营和人们的日常生活.密码技术作为信息安全系统中保障数据安全的关键技术，其体制的完备性在于密码在计算机有限空间和时间条件下的不可破译性.而这种不可破译性又由密钥的随机安全性决定.在研究密码学过程中，概率论无疑起着相当重要的作用，它对于检测密钥的随机可靠性，分析攻击密码流，检测随机序列的伪随机性都具有十分重要的作用。

2 背景知识和概念

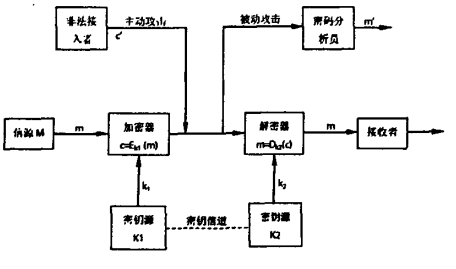
密码学：密码学是对与信息安全各方面（比如机密性、数据完整性、实体认证及数据源认证）有关数学技术的研究。

密码加密和解密：发送方将要发送的消息称为明文，明文被变换成看似无意义的随机消息，称为密文，这种变换过程称为加密.加密的逆过程称为解密.

密码分析：密码通信过程中，截收者通过分析可能从截获的密文推断出原来的明文或者密钥的过程称为密码分析.

3 密码学中的概率统计模型

密码学的研究是基于密码通信系统的.它包括了从信源到接收者的加密和解密过程，也包括在此过程中的受到的来自非法接入者的非法攻击.其模型的图解如下：



保密通信系统模型

其中 m 代表明文，c 代表密文，k_1 和 k_2 均代表密钥流.m' 为密码分析员解密获得的明文，c' 为攻击者获得的密文部分。

4 概率论在密码学中的应用

4.1 抗击主动攻击的生日攻击模型

密码加密的任务是为了防止通信双方的信息内容被第三方截取，即主要来自非法者的主动攻击（见保密通信系统模型）.采用消息认证，通过基于公开函数—杂凑（*Hash*）函数取得认证符就可以防止主动攻击。

确切的说，Hash函数（简称H）用于将任意长的消息M映射为较短的、固定长度的一个值 $H(M)$，即杂凑值、杂凑码或消息摘要.对于定义域 D 和值域 R，有 $H: D\to R$ 和 $|D|>|R|$.它应满足以下条件：

（*a*）已知 h，求使得 $H(x)=h$ 的 x 在计算上是不可行的，这一性质称为函数的单向性，称 $H(x)$ 为单向杂凑函数；

（*b*）已知 x，找出 $y(y\neq x)$ 使得 $H(y)=H(x)$ 在计算上是不可行的；

（*c*）找出任意两个不同的输入 x, y，使得 $H(y)=H(x)$ 在计算上是不可行的.

其中条件（*a*）保证攻击者截获 M 和 $C=H(S|M)$ 后，求不出 C 的逆 $S|M$，也就不可求出秘密值 S.条件（*b*）使得敌手无法在已知某个消息时，找到与该消息具有相同杂凑值的另一消息.条件（*c*）用于抵抗生日攻击。

由于该函数是多对一的，因此存在碰撞（具有同一输出的输入对）是不可避免的.实际上，限制 H 到一个 t 比特输入的域（$t>n$），在所有输出是等概率的意义下，假如 H 是“随机的”，则大约有 2^{t-n} 个输入对应同一个输出，且两个随机选择的输入产生同一输出的概率为 2^{-n}（与 t 无关）.抽象成数学中的生日攻击模型就是：

已知一杂凑函数 H 有 n 个可能的输出，$H(x)$ 是一个特定的输出，如果对 H 随机取 k 个输入，则至少有一个输入 y 使得 $H(y)=H(x)$ 的概率为0.5时，k 有多大？

解 因为 H 有 n 个可能的输出，所以输入 y 产生的输出 $H(y)$ 等于特定输出 $H(x)$ 的概率是 $1/n$，反过来说 $H(y)\neq H(x)$ 的概率是 $1-1/n$.y 取 k 个随机值而函数的 k 个输出中没有一个等于 $H(x)$，其概率为 $(1-1/n)^k$，所以 y 取 k 个随机值得到函数的 k 个输出中至少有一个等于 $H(x)$ 的概率为 $1-(1-1/n)^k$.由 $(1+x)^t\approx 1+tx$，其中 $|x|\ll 1$，可得：

$$1-(1-1/n)^k\approx 1-(1-k/n)=k/n,$$

若上述概率等于0.5，则 $k=n/2$.特别地，如果 H 的输出为 m 比特长，则可能的输出个数 $n=2^m$，则 $k=2^{m-1}$.

4.2 随机密钥构造过程中的测试统计量

密码学中最重要的并不是加密算法和解密算法，而是在加密过程和解密过程中所需要用到的加密密钥和解密密钥.加密密钥和解密密钥的生成又在于随机序列的获得.

事实上，不管利用硬件还是利用软件设计随机序列都

是很困难的，在多数情况下，我们需要通过已成熟的算法来获得近似的随机序列，即伪随机序列.判断序列是否具有伪随机性就需要用到概率统计测试，而这种测试对于被动攻击中的密码分析员（见密码通信系统）破解密码也同样具有重要意义：

4.2.1 频数检验

频数检验是最基本的检验.在进行随机性测试时，应该首先选择频数检验，频数检验通过后再选择其它检验，否则就不必选择其它检验了.有一些检验比如线性复杂度检验会耗费大量的时间，而频数检验是速度最快的一种检验方式，为了节省时间，我们有必要在进行其它检验以前，先选择频数检验.频数检验是用来检验一个位序列中0和1的个数是否近似相等，这正是随机序列所应具备的.

令 n_0 和 n_1 分别表示待测位序列 S 中0和1的个数.所使用的统计量为：

$$X_1=\frac{(n_0-n_1)^2}{n}$$

若 n 不小于10，则该统计量近似地服从自由度为1的 χ^2 分布.在实际使用中建议 $n\geq 10000$.

4.2.2 跟随测试（又称序列测试或双比特测试）

该测试的目的是判定位序列 S 的子序列00,01,10,11所出现的次数是否近似相等，这也是一个随机序列所应具备的特性.令 n_0 和 n_1 分别表示 S 中0和1的个数，且 $n_{00},n_{01},n_{10},n_{11}$ 分别表示 S 中子序列00,01,10,11出现的次数.注意 $n_{00}+n_{01}+n_{10}+n_{11}=(n-1)$，因为这些子序列允许相交.所使用的统计量为：

$$X_2=\frac{4}{n-1}(n_0^2+n_0^2+n_0^2+n_1^2)-\frac{2}{n}(n_0^2+n_1^2)+1$$

若 n 不小于21，则该统计量近似地服从自由度为2的 χ^2 分布.

4.2.3 游程检验

游程是序列的一个子串，由连续的0或者1组成，并且其前导和后继元素都与其本身的元素不同.游程检验主要检验待检序列 S 中游程总数是否符合随机性要求.游程测试可用来判定序列 S 中不同长度游程的个数是否与随机序列中所期待的一样，令 B_i,G_i 分别为 S 中长度为 i 的1游程和0游程的个数，所使用的统计量为：

$$X_3=\sum_{i=1}^{k}\frac{(B_i-e_i)^2}{e_i}+\sum_{i=1}^{k}\frac{(G_i-e_i)^2}{e_i}$$

其中 $e_i=(n-i+3)/2^{i+2}$，k 是满足 $e_i\geq 5$ 的最大的 i.该统计量近似地服从自由度为 $2k-2$ 的 χ^2 分布.

4.2.4 扑克测试

扑克测试用来确定每个长度为 m 的序列在 S 中出现的次数是否近似相等，令 m 是一个满足 $\left[\frac{n}{m}\right]\geq 5\cdot(2^m)$ 的正整数，令 $k=\left[\frac{n}{m}\right]$ 将序列拆分成 k 个不相交的部分，每部分的长度为 m，令 n_i 为第 i 种长度为 m 的序列所出现的次数.所使用的统计量为：

$$X_4=\frac{2^m}{k}\left(\sum_{i=1}^{2^m}n_i^2\right)-k$$

它近似地服从自由度为 2^m-1 的 χ^2 分布.$m=1$ 时即频率测试.

4.2.5 自相关测试

自相关测试用来检测序列 S 与其发生（非循环）移位后形成的序列之间的相关性.

令d为一个固定整数，$0<d\leq\left\lfloor\frac{n}{2}\right\rfloor$，序列 S 与 S 发生 d 移位后所形成的序列中的不同比特数为：$A(d)=\sum_{i=0}^{n-d-1}s_i\oplus s_{i+d}$，所使用的统计量为：

$$X_5=2\left(A(d)-\frac{n-d}{2}\right)/\sqrt{n-d}$$

若 $n-d>9$，则它近似地服从 $N(0,1)$（双边测试）.

4.3 其它概率模型

4.3.1 关于密码学中的周期随机序列模型

我们称有限域 F_q 上的序列 $a=(a_0,a_1,a_2,\cdots)$ 为周期序列，如果存在正整数 l，使对一切非负整数 k 都有 $a_{k+l}=a_k$，而称满足上式的最小正整数 l 为序列 a 的周期，且记为 $p(a)$.

Ruppel对周期随机序列的线性复杂度的均值进行研究发现周期为 2^n 的2值随机序列的线性复杂度的均值接近 2^n-1.

4.3.2 “秘密共享方案”中的概率模型[3]

“秘密共享方案”是关于子秘密的分配规则的说明.假设 Γ 是一个访问结构，$F=\bigcup_{\alpha\in\Omega}F_\alpha$ 是分配规则的集合，称 F 是一个实现访问结构 Γ 的完备秘密共享方案，如果它具有下列两个特性：

(1) 对任意一个参与者的授权子集 $B\subseteq M$，不存在两个分配规则 $f\in F_i$ 和 $g\in F_j$，$k,j\in K$，$k\neq j$，满足 $f_B=g_B$（即在任意一个授权子集中对参与者的任何一个秘密分配将确定密钥的值）；

(2) 对任何一个参与者的未授权子集 $B\subseteq M$，对任意的子秘密分配规则 $f_B\in S(B)$ 和每一个 $k\in K$，均有 $p(k\mid f_B)=p_K(k)$（即对 B 的子秘密分配没有关于密钥值的任何信息）.

上述讨论了概率论在密码通信系统中各个不同部分的应用，揭示了概率论在密码学中的重要作用，并且将一些概率论在密码中的应用分析得更为透彻.众多的密码学文献也向我们展示了这种重要性的实际应用效果.深入研究概率论的方法在密码中的应用对密码分析起着及其重要的作用.

（基金项目：国家创新性实验计划项目）

参考文献：

[1] 杨波.现代密码学[M].北京：清华大学出版社,2007：5~20,160~180.

[2] A.J. Menezes，P.C.Van Oorschot，S.A.Vanstone 著,胡磊，王鹏译.应用密码学手册[M].北京：电子工业出版社，2005:150~160.

[3] 李世取,黄晓英,刘文芳,张卫明,刘凤梅.密码学中的有关概率模型[M].电子工业出版社，2005:370~395.

概率论在密码学中的应用

作者： 郑娟， 朱金伟， 曾文军
作者单位： 华中师范大学数学与统计学学院,湖北·武汉,430079
刊名： 科协论坛（下半月）
英文刊名： SCIENCE & TECHNOLOGY ASSOCIATION FORUM
年，卷(期)： 2008(7)

参考文献(3条)

1. 李世取;黄晓英;刘文芳;张卫明 刘凤梅 密码学中的有关概率模型 2005

2. A.J.Menezes;P.C.Van Oorschot;S.A.Vanstone;胡磊， 王鹏 应用密码学手册 2005

3. 杨波 现代密码学 2007