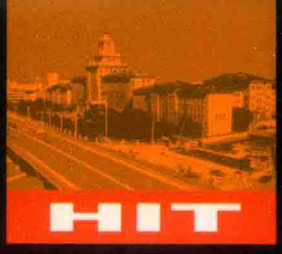
Outline of Modern Cryptographic Algorithms



数学·统计学系列

现代密码算法概论

曹正军　刘丽华　著

哈爾濱工業大學出版社
HARBIN INSTITUTE OF TECHNOLOGY PRESS

数学·统计学系列

现代密码算法概论

Outline of Modern Cryptographic Algorithms

曹正军　刘丽华

哈爾濱工業大學出版社
HARBIN INSTITUTE OF TECHNOLOGY PRESS

内 容 简 介

本书介绍了160个现代密码算法和协议，内容涉及现代密码算法的数学基础，现代密码算法在信息安全领域中的应用，以及与之相关的科学研究和社会实践. 涵盖了现代密码学的主要内容，包括哈希函数、非对称加密算法、公钥加密算法、密钥保管、数字签名、零知识证明、不经意传输、多方安全计算、外包计算、量子计算、电子货币等. 本书提供了大量的算法评析，既能激发读者的兴趣，又能引发读者的思考，有助于进一步夯实现代密码算法的理论基础，弥合理论研究与实践之间的缝隙.

本书可作为本科生、研究生选修教材或参考书. 部分章节可作为相关科技人员的培训教材.

图书在版编目（CIP）数据

现代密码算法概论 / 曹正军，刘丽华著. —哈尔滨：哈尔滨工业大学出版社，2019.5

ISBN 978-7-5603-8105-3

Ⅰ. ①现… Ⅱ. ①曹… ②刘… Ⅲ. ①密码算法－概论 Ⅳ. ①TN918.1

中国版本图书馆CIP数据核字(2019)第069415号

策划编辑 刘培杰 张永芹
责任编辑 张永芹 穆 青
封面设计 孙茵艾
出版发行 哈尔滨工业大学出版社
社 址 哈尔滨市南岗区复华四道街10号 邮编 150006
传 真 0451－86414749
网 址 http://hitpress.hit.edu.cn
印 刷 哈尔滨市工大节能印刷厂
开 本 787mm × 1092mm 1/16 印张 21.5 字数 535千字
版 次 2019年5月第1版 2019年5月第1次印刷
书 号 ISBN 978-7-5603-8105-3
定 价 98.00元

序

现代密码算法的研究内容相当丰富，体系较为复杂. 把散落的算法分成若干个类别，再遴选出一些代表性的结果进行解析，这项工作本身极具挑战性，国际上尚无这类著作. 本书作者研读了近千个密码算法，并结合他们自己近十年的研究成果撰写了这本专著，这是一项十分令人欣慰的工作.

本书选材难易适中，涵盖了产业界流行的公钥密码算法RSA, DSA，哈希算法MD5, SHA，对称分组加密算法DES, AES，和新近的身份基加密算法、属性基加密算法、无证书密码算法、全同态加密算法、量子算法、外包算法等，同时还论述了P和NP问题、比特币、区块链、PGP电子邮件系统、传输层安全协议TLS1.3. 作者的论述侧重于密码算法的实用性，淡化其理论性，这种方法对于一门应用学科来说无疑是值得提倡的.

书中很多观点和论述是非常独特的，有些论据还有待商榷，但瑕不掩瑜. 我相信这本专著会引领读者去探索现代密码算法的真谛，激发读者的思索.

曹珍富

2019年1月

于华东师范大学

前　言

密码学为信息安全提供了理论基础和技术支持, 是一门介于数学、计算机科学、通信科学之间的交叉学科. 现代密码学发展已逾四十年, 一些经典的密码算法也已经使用了三十多年, 没有这些密码算法的庇护, 就没有现在方便快捷的信息交流和网络支付.

设计安全、高效的密码算法, 开发符合实际需求的密码协议, 是密码学理论研究的基本目标. 近二十年来的密码学研究论文总量已超十万, 但能够付诸于实践的研究成果数目极少. 本书汇集了笔者近年来的研究经验和结果, 以实践为依托, 对那些新潮的密码模型、时髦的密码算法、功能丰富的密码协议进行评析.

本书分为三个部分, 基础篇、理论篇及专题篇. 基础篇包括密码学概述、现代密码算法的数学基础、提取数字指纹、实体和消息的认证、协商或分发密钥、对称分组加密算法与两个应用实例, 共6章. 理论篇包括密钥协商与管理、公钥加密算法、身份基加密算法、无证书密码算法、属性基加密算法、同态加密、广播加密与代理重加密、数字签名、代理签名与盲签名、群签名、承诺与零知识证明、不经意传输、安全多方计算与洗牌算法、外包计算, 共14 章. 专题篇包括 AKS 算法——兼谈 P 和 NP 问题、量子密码协议与算法、Bell 不等式和 Shor 算法、比特币和区块链, 共4章. 本书还收录了几十位专家的简介, 内容来自他们的个人主页或维基百科.

作者衷心感谢中国科学院数学与系统科学研究院刘木兰研究员、上海交通大学沈灏教授、华东师范大学曹珍富教授多年来在学术上给予的鼓励与支持. 感谢国家自然科学基金(60873227, 61303200, 61411146001)、教育部留学回国人员科研基金、上海市科委科技创新基金, 以及上海大学、上海海事大学基金会先后对书中部分研究内容的资助. 感谢哈尔滨工业大学出版社的刘培杰副社长和张永芹、穆青、杜莹雪等编辑, 以及为本书出版给予热心帮助的朋友们.

由于作者水平所限, 书中疏漏之处在所难免, 望广大读者给予批评指正, 谢谢!

曹正军, 刘丽华
caozhj@shu.edu.cn
liulh@shmtu.edu.cn

目　录

第1章　密码学概述

密码学是一门研究如何使敌手“看不懂、骗不了”的学问. “看不懂”是指敌手窃取到所有通信信号后, 无法恢复出信号中蕴藏的信息, 也就是通常所说的保密性或机密性. “骗不了”是指敌手无法假冒身份或者利用篡改后的信号来欺骗用户, 也就是通常所说的认证, 包括身份认证和消息认证.

本章内容包括信息安全与信号安全、信息安全等级、分析敌手的行为、加密通信的基本模型、口令与密钥、建立信任关系、公钥密码体制、哈希函数、量子计算机, 共9个部分. 阐述了信息安全与信号安全的差异, 界定了敌手的攻击行为, 辨明了口令与密钥之间的关系, 指明了信任关系在密码学中的重要地位, 普及了公钥密码体制的一些基本概念, 纠正了哈希函数就是加密算法的错误认识.

1.1　信息安全与信号安全

通常所说的信息是由符号、文字、图像、语音等构成的, 这些信息在实际数字通信中都表示成由0, 1构成的比特串. 比如中文字符“汉”的Unicode编码转换成二进制后得到的是11100110 10110001 10001001, 身在北京的张三怎样把这个比特串发给上海的李四呢? 需要利用通信信号来解决这个问题.

通信信号是指那些能够用来传递信息的物质, 比如无线电波、电信号、磁信号、光信号等. 电路中电压的大小可以用来表示0和1, 张三利用电压调制电路把11100110 10110001 10001001调制成相应的电信号, 这些电信号再通过光电转换器转换成不同强度或频率的光信号, 然后利用光纤传送出去. 在传送过程中, 光信号会衰退, 需要利用中继服务器来增强信号, 直至传递到上海李四端的接收设备. 接收设备把光信号转换为电信号, 然后转换成11100110 10110001 10001001, 再用对应的编码规则转换成“汉”.

信号在传送过程中会受到干扰, 比如环境噪声、通信设备的突发故障等. 为了保证接收端能恢复出正确的信号, 发送端要采取一些措施把原来的比特串变换成更长的比特串, 利用各个比特位之间的关联性实现检错、纠错功能. 这个过程就是通常所说的信道纠错编码.

信息安全包括很多内容, 最主要的是机密性和认证. 机密性是指非授权的用户无法读取信号中蕴藏的信息. 从形式上看, 非授权用户得到的只是由0, 1构成的比特串, 他不知道采用什么样的变换把获得的比特串转换成原来的信息. 认证是指用户能够确认对方的身份或者信息的来源.

光电信号的经典性态是容易调制和测量的, 敌手可以通过监听线路获得信号. 密码学总是假设敌手已经窃取了所有信号, 在这种情形下, 研究如何阻止敌手读取信号中蕴藏的信息或者篡改信号欺骗用户. 因为敌手在窃听的时候基本上没有干扰原来的信号, 所以用户能够恢复出发送端

发送的信号, 通信双方无法得知有没有敌手在窃听, 即传统的密码学不能发现窃听. 就机密性而言, 传统密码学的目的是阻止敌手获得蕴藏在信号中的信息, 是一种智力手段. 其方法是, 研究什么样的数学变换是不可逆的, 如何在变换中嵌入陷门(trapdoor).

1984年, Bennett 和 Brassard 发表了一篇论文[22], 宣称量子密码学能够发现窃听, 是绝对安全的, 其理论基础是量子力学的测不准原理. 传统的通信信号性态是指电压值、光的强度与频率、电磁波的频率等. 与这些性态不一样, 量子通信利用的信号性态是量子态, 比如光子的偏振方向. 因为一个未知量子态是无法复制的, 一旦敌手试图窃听量子信号, 将有一半的机会改变信道上的量子态, 所以接收方就无法恢复出原信号. 量子信号发送完成后, 双方通过一个可认证的传统信道(用于确认对方的身份)进行公开比对, 如果发现双方在采用相同的测量方案时测得的量子态是不一致的, 就可以断言量子信道上有窃听. 量子密码学的目的是阻止敌手获得信号, 是一种物理手段. 其方法是, 研究量子信号的制备与检测技术, 如何提升量子信号的传送距离.

敌手无法获得信号, 自然就无法获得蕴藏在信号中的信息. 因此, 一个通信系统是信号安全的, 也必然是信息安全的. 得不到信号自然就得不到信息, 这就是量子通信绝对安全的由来. 但在有敌手介入的情形下, 一个通信系统在阻止敌手获得信号的同时也必然无法保证目标用户获得正确的信号, 也就是说该系统是不稳定的.

通信的首要目的是稳定性, 要保证用户能够恢复出正确的信号[84]. 密码学总是假定敌手是一直存在的, 无论他在窃听信号还是在篡改信号. 密码学总是假定敌手所具备的物理技术手段比接收方更强. 因此, 一个通信系统如果从物理上剥夺了敌手窃取信号的能力, 那么也必然无法保证接收方获得正确的信号, 也就是说, 通信系统的稳定性与信号安全是不兼容的.

有些人宣称, 利用量子通信信号可以发现窃听, 然后定位敌手的地理位置, 再通过暴力手段清除敌手. 很显然, 这种辅以暴力的通信模式根本就不是密码学的研究目标, 也没几个人相信这种诱敌歼灭计在现实中真的行得通.

一个信息安全系统虽然不能从物理上削弱敌手截获信号的能力, 但是能够从智力上保证敌手无法获得蕴藏在信号中的信息, 也就是说, 通信系统的稳定性与信息安全是兼容的. 关于信息安全的这个特征(兼容性), 现有的文献都没有述及, 导致有些人认为牺牲信号的稳定性换取信息的安全性也是值得的. 这种认识是错误的! 正确的信号都收不到, 信息又怎么能传递出去呢?

1.2 信息安全等级

用与明文等长的密钥和明文进行异或运算得到密文, 这种加密方法称为一次一密. 密码学认为一次一密是绝对安全的. 但是这种加密方式与密码学的一个内在特征——保护机密的有效性是有冲突的, 因而在实践中是行不通的.

保护机密的有效性是指用来保护文本的密钥长度要远远小于文本自身的长度. 可以从以下两种情形来理解这个特征.

- 本地保密: 比如用户原来需要保密的明文 m 大小是100K, 用来做一次一密的密钥 s 的大小

也是100K, 虽然加密后得到的密文c不需要保护, 但密钥本身却是不能外泄的. 保密的对象从明文m换成了密钥s, 所需的存储量仍然是100K.

- 异地保密: 比如用户 Alice 想把100K的明文m发送给异地的 Bob, Alice 用来做一次一密的密钥s的大小也是100K, 那么 Bob 如何安全地获得密钥s呢? 如果有安全方式使得Bob能够获得密钥s, 还不如直接让 Bob 获得明文m.

关于这个特征(保护机密的有效性), 现有的文献都没有述及, 导致一些密码爱好者很难认识到一次一密在实践中是不可行的. 一些关于密码知识的科普作品四处滥用绝对安全这个概念也就不足为奇啦.

密码学所说的信息安全等级是用敌手破解一个密码系统所需要的计算量来衡量的. 计算量取决于算法运行参数的大小[245], 而不是设计参数的大小.

密钥长度是关系信息安全等级的一个决定性因素. 密码学经常会用"大数"这样的字眼来描述参数, 表1-1有助于读者增强对大数的认识[350].

表1-1

物理模拟量	大数
行星在银河系中的寿命	2^{64}分之一
每天被闪电杀死的可能性	2^{33}分之一
每年被淹死的可能性	2^{16}分之一

1.3　分析敌手的行为

评估一个通信系统的安全性首先得分析敌手的行为. 敌手假冒身份、篡改、窃听或者恶意破坏信号, 直接关系到通信的成败. 敌手的攻击可分为三种[84,358]:

- 被动攻击: 敌手只窃取信号, 接收端得到的信号与发送端发出的信号是相同的.
- 主动攻击: 敌手假冒身份或者篡改信号, 试图欺骗接收端. 接收端得到的信号与发送端发出的信号是不同的.
- 恶意攻击: 敌手蓄意破坏信号, 使得用户无法完成正常的通信.

窃听公共通信网络上的信号往往是容易的, 成本比较低(包括技术成本、社会成本), 也就是说, 被动攻击几乎无处不在. 篡改公共通信网络上的信号需要较高的成本. 阻断公共通信网络则需要付出很高的代价, 譬如切断光缆、屏蔽无线电波.

考虑到攻击的收益、成本的大小, 可以假设敌手攻击一个通信时有下面三种选择:

- 上策是假冒身份或者篡改信号, 设法欺骗用户.
- 中策是窃听信号, 设法获取通信内容.
- 下策是破坏信号, 让用户无法达成正常的通信.

考虑到敌手的攻击行为, 数据通信的安全目标主要包括三个方面:

- 通信实体的真实性——通信一方能够确认另一方的身份.
- 数据形式的完整性——接收端能够检测数据是否被篡改.
- 数据内容的机密性——只有合法用户才能够恢复数据的内容.

1.4 加密通信的基本模型

假设 Alice 与 Bob 在不安全的信道上进行通信, 敌手 Eve 试图发动攻击以便实现自己的目标. 比如 Alice 和 Bob 利用移动通信设备进行通话, 此时信道就是移动通信网, Eve 企图偷听他们的交谈内容; 或者 Alice 使用信用卡通过互联网向 Bob 购买商品, 信道就是互联网, Eve 企图获得 Alice 的信用卡信息或者冒充 Alice 与 Bob 进行交易.

两个实体利用信道进行通信的基本目标是, 确保对方能够正确地接收到自己发送过去的物理信号, 即必须确保接收端得到的物理信号 S' 与发送端发出的物理信号 S 是一致的(图1-1). 因为信道常常是公开的, 不是专有的, 所以敌手 Eve 能够获得信道上的物理信号 S'. 信号解码规则 T' 也是公开的, 从而 Eve 能够获得密文 C'.

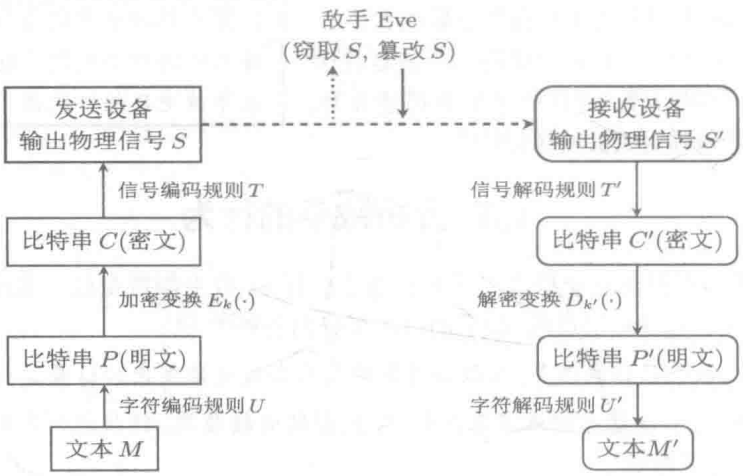


图1-1

通常所说的密码是指由加密变换与解密变换构成的算法, 算法的操作过程不需要保密, 算法的输入参数分为两种, 一种是可以公开的参数, 另一种是需要保密的私钥. 如果加密变换与解密变换是相同的, 或者几乎是相同的, 那么这样的加密系统就称为对称密码系统.

现代密码系统的安全性完全取决于私钥, 算法细节是公开的, 这个特点有利于实现算法的标准化, 推进算法普及工作[155,298,368].

1.5 口令与密钥

人们常常分不清口令和密钥的关系, 实质上, 口令是用户访问设备或算法的凭据, 密钥是算法调用的参数. 伪随机数缺少记忆规律, 是不适合作口令的. 绝大部分用户会选择自己熟悉的一些数字或字母作为口令, 比如生日、电话号码、一个特殊的日期等.

调用密码算法需要口令, 运行密码算法需要调用密钥. 各种密码算法或设备一旦置入参数

后, 用户就无须去记忆各个参数了, 包括自己的密钥. 比如用户只知道启用银行U盾的口令, 不知道U盾里的算法调用了哪些参数. 因此, 针对密码系统的暴力攻击有两种方式: 一是猜测用户启动密码系统或设备的口令, 二是猜测密码系统内部调用的密钥. 实践中, 可以对调用密码算法设置一些限制措施. 比如输入口令连续出错5次, 系统就自动关闭访问通道, 或者输入口令的间隔时间少于80 s就拒绝接受口令. 网络上有一些声称能破解密码的软件, 它们的破解方法就是利用预先编制的口令表(彩虹表)进行口令测试, 获取调用密码算法的权限.

密码系统的安全取决于密钥, 密钥的安全取决于它的随机性. 实践中多用伪随机数发生器来产生密钥, 这些发生器是基于数学方法设计的, 产生的序列是有周期的. 线性同余发生器的运算规则是 $X_n = aX_{n-1} + b \pmod{m}$, 其中 X_n 是序列的第 n 个数, X_{n-1} 是序列的第 $n-1$ 个数, a, b, m 是常数, 种子(初始值)是 X_0. 该发生器产生的序列周期不会超过 m. 线性同余发生器虽然适合软件实现, 但速度较慢. 反馈移位寄存器(图1-2)是一种非常适合硬件实现的更快速的伪随机数发生器[146,309].

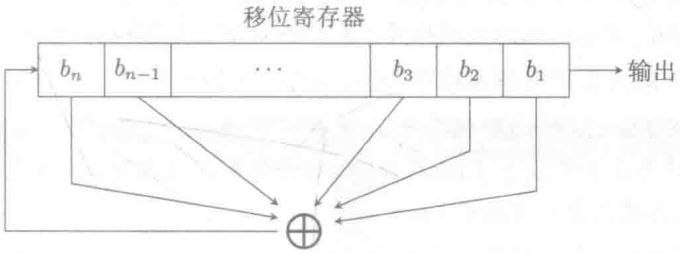


图1-2

据说, 可以用物理方法产生真随机数. 比如, 利用环境噪声的随机性来提取随机数, 利用入射角是45°的单个光子在界面处发生折射或反射的情况来提取随机数(量子随机数发生器). 利用物理方法产生随机数的弊端是, 设备复杂、效率低、很难与软件系统集成起来.

1.6　建立信任关系

信任是社会交往的基础. 有了信任人们才有可能发展各种协议交流信息, 提供各种各样的服务. 没有信任几乎什么事都办不成.

直接用身份证号作为网络用户的凭证, 再利用身份证号对数据进行加密与认证, 这种做法在实践中很难行得通, 因为身份证号是根据简单的编制规则产生的. 为此, 必须为网络用户创建新的"身份信息", 使得不同的用户拥有不同的"身份信息"(电子数据). 这些"身份信息"必须由权威机构确认后向社会公布. 基于这一想法, 人们引入了公钥基础设施PKI.

公钥基础设施是指, 每个用户需向公钥证书管理中心CA提交公钥申请, CA确认用户的身份信息 ID 和用户选择的公钥 PK 后向用户颁发公钥证书, 并向社会公开发布. 公钥基础设施确保了公钥的真实性, 确保了公钥所有者身份的真实性. 用户公钥的发布、暂停以及撤销均需通过CA来实施. 实质上, CA充当了可信第三方的角色.