

加密技术：从理论到实践

title page

For them

I owe much inspiration to both my parents. My mother Jannie constantly demonstrated me that computer graphics will never improve nature. She also converted one of my first METAPOST graphics into a patchwork that will remind me forever that handcraft is more vivid than computer artwork. My father Hein has spent a great deal of his life teaching math, and I m sure he would have loved METAPOST. I inherited his love for books. I therefore dedicate this document to them.

Colofon

This manual is typeset with CON_TE_XT MKIV. No special tricks are used and everything you see in here, is available for CON_TE_XT users. The text is typeset in Palatino and Computer Modern Typewriter. We used LUAT_EX as T_EX processing engine. Since this document is meant to be printed in color, some examples will look sub-optimal when printed in black and white.

Graphics

The artist impression of one of Hasselts canals at page ?? is made by Johan Jonker. The CDROM production process graphic at page ?? is a scan of a graphic made by Hester de Weert.

Copyright

Hans Hagen, PRAGMA Advanced Document Engineering, Hasselt NL
copyright: 1999-2020 / version 4: October 5, 2020

Publisher

publisher: Boekplan, NL
isbn-ean: 978-94-90688-02-8
website: www.boekplan.nl

Info

internet: www.pragma-ade.com
support: ntg-context@ntg.nl
context: www.contextgarden.net

序言

TBD

荣亮

2020年October月于苏州

目录

1	环游密码世界	5	2.4.6	Enigma的破译	24
1.1	密码学中的基本概念	5	2.5	本章小结	25
1.2	对称加密算法和非对称加密算法	6	3	对称加密算法	27
1.3	其他密码技术	7	3.1	流密码和分组密码	27
1.3.1	单向散列函数	7	3.2	RC4流式加密算法	28
1.3.2	消息认证码	7	3.2.1	初始化算法	28
1.3.3	数字签名	7	3.2.2	伪随机密码生成算法	28
1.3.4	伪随机数生成器	8	3.2.3	使用OpenSSL进行RC4加密	29
1.4	信息安全所面临的威胁及对策	8	3.2.4	针对RC4的破解	30
1.5	密码与信息安全常识	9	3.3	DES加密算法	30
1.5.1	任何时候不要尝试发明新的加密算法	9	3.3.1	DES的总体结构	31
1.5.2	不要使用低强度的密码	9	3.3.2	初始置换和最终置换	32
1.5.3	信息安全也是一门社会性课题	9	3.3.3	Feistel网络	32
1.6	本章小结	10	3.3.4	密钥调度	35
2	密码的历史典故	11	3.4	3DES加密算法	35
2.1	中国古代人民怎么加密?	11	3.4.1	3DES的总体结构	35
2.2	凯撒密码	12	3.4.2	密钥选项	36
2.3	简单替换密码	13	3.4.3	3DES密码现状	37
2.3.1	什么是简单替换密码	13	3.5	AES加密算法	37
2.3.2	用频率分析的方法破解简单替换密码	14	3.5.1	Rijndael算法	37
2.4	复式替换密码: Enigma	20	3.5.2	Rijndael的加密	37
2.4.1	Enigma的构造	20	3.5.3	Rijndael密钥扩展算法	41
2.4.2	Enigma的加密过程	22	3.5.4	Rijndael的解密	43
2.4.3	Enigma的解密过程	23	3.5.5	Rijndael的安全性	44
2.4.4	每日密码和通信密码	24	3.6	Chacha20加密算法	44
2.4.5	Enigma的弱点	24	3.6.1	ARX运算和操作函数	44
			3.6.2	ChaCha20分组函数	46
			3.6.3	ChaCha20加解密过程	48
			3.6.4	ChaCha20的应用	48
			3.7	本章小结	48

第1章 环游密码世界

随着物联网和智能家居的兴起，网络信息安全已经渗透到我們日常生活的方方面面。密码技术作为信息安全的基石，为网络通信提供了安全和可靠的技术手段。本章，我们先整体了解一下密码世界，看看各种密码技术如何为我们的信息安全保驾护航。

1.1 密码学中的基本概念

信息在人与人、人与机器、机器与机器之间交互的过程中存在被第三方（人或计算机）窃取的风险，密码技术提供了信息在通信双方交互的过程中免遭第三方窃取并破解，以及确保通信任何一方不被欺骗的一系列算法。

首先，我们来了解一下与密码技术有关的角色：

- 发送者 (sender)：消息的发送方。
- 接收者 (receiver)：消息的接收方。
- 窃听者 (eavesdropper)：监听在消息传送的通道上，窃取消息的恶意攻击方。
- 破译者 (cryptanalyst)：为研究密码强度而工作的密码破译人员或密码学研究者，要注意和窃听者的本质区别。

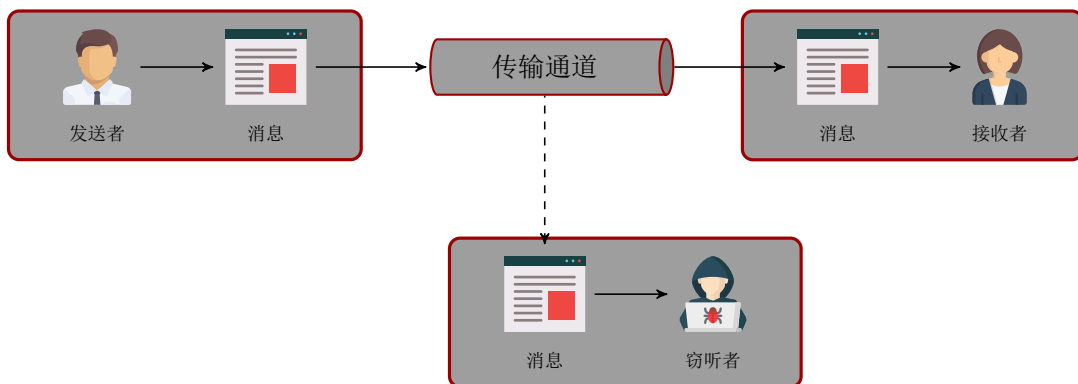
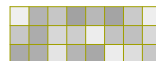


图1.1 消息的发送、接收和窃听

如图 1.1所示，如果发送者不对要发送出去的消息进行任何处理，很容易被窃听者窃取并获知消息的内容。为确保消息的机密性 (confidentiality)，发送者在发送消息之前需要对其进行加密 (encryption)。消息可以是任何类型的数据，例如，邮件、文档和交易等。通常，我们把加密前的消息称之为明文 (plaintext)，加密之后的消息称之为密文 (ciphertext)。接收者收到密文后将其恢复回明文的过程称为解密 (decryption)。

因为加密和解密需要相应的密钥才能完成，当窃听者窃取到加密后的密文之后，因为没有密钥，也就无法还原出原始的明文。这就相当于，我们在把重要的机密文件传给接收人之前，先把机密文件锁在保险柜里面，然后把保险柜交给物流公司帮忙交付给接收人，接收人收到之后用保险柜的钥匙打开取出该机密文件。在物流运输的过程中，保险柜有可能面临被丢失的风险，如果有人偷盗了保险柜，因为没有钥匙也无法取出里面的文件。整个过程可以用图 1.2来描述。



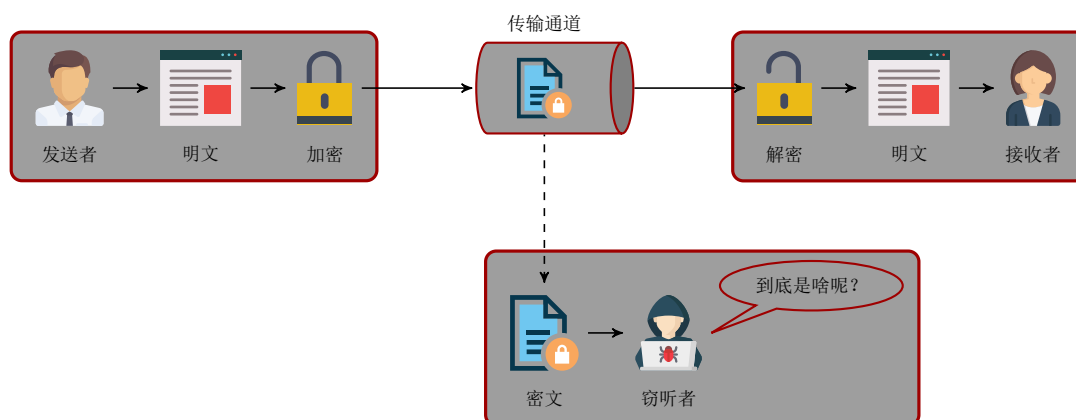


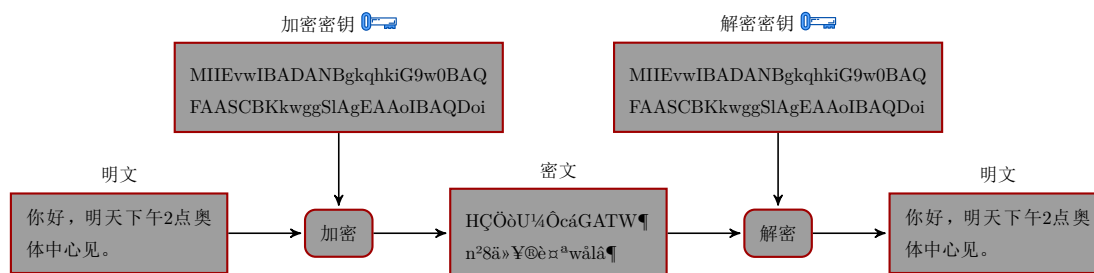
图1.2 消息加密和解密的过程

1.2 对称加密算法和非对称加密算法

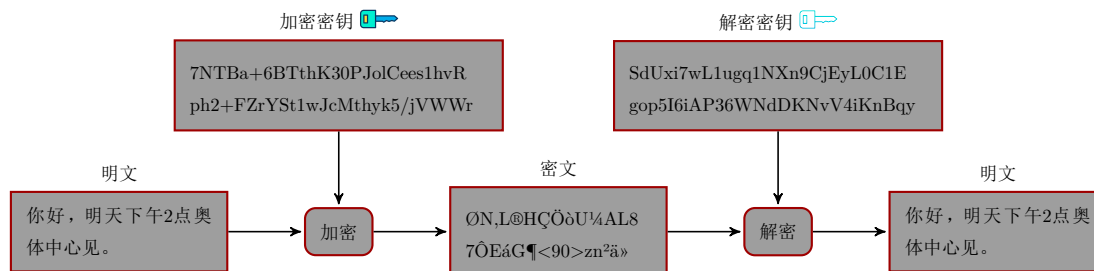
在上面的例子中，如果保险箱在运输的过程中被坏人偷盗，坏人有可能使用物理破坏的方法撬开保险箱，取出里面的机密文件。那么，显而易见，保险箱越坚固，坏人就越难破坏保险箱拿到里面的机密文件。所以，保险箱的坚固程度就决定了里面机密文件的安全性到底有多高。

在网络通信领域，网络传输通道是极不可靠的，信息加密后的密文在传输的过程中，存在被窃听者窃取的风险，加密算法也要确保即使密文被窃取也不能在现实的时间内被破解，这也正是加密算法的魅力所在。

从原理上，加密算法被分成两大类，即对称加密算法 (symmetric encryption algorithm) 和非对称加密算法 (asymmetric encryption algorithm)。它们最主要的区别在于密钥 (key) 的使用方式不同。区别于现实生活中的“钥匙”，密码算法中的密钥是一串很长的看起来非常杂乱无章的字符序列。



对称加密算法中，加密密钥和解密密钥相同



非对称加密算法中，加密密钥和解密密钥不同

图1.3 对称加密和非对称加密



对称加密算法在加密和解密时使用了相同的密钥，因为加密方和解密方使用了相同的密钥，任何人拿到了密钥就能解密加密后的消息从而获取到原始的数据。因此，对称加密算法的密钥必须在加密方和解密方两者同时妥善保管，不能泄露给任何未经授权的第三方。

相反，非对称加密算法则在加密和解密时使用了不同的密钥。而且，在非对称加密算法下，加密密钥通常被公开，但解密密钥需要由接收者私自妥善保管。据此特点，非对称加密算法又常常被称为公钥加密算法 (public key encryption)。

非对称加密算法是在1976年，由狄菲 (Whitfield Diffie) 与赫尔曼 (Martin Hellman) 两位学者以单向函数与单向暗门函数为基础，提出了“非对称密码体制即公开密钥密码体制”的概念，开创了密码学研究的新方向。现代计算机和互联网中的安全体系，很大程度上都依赖于公钥加密算法。

1.3 其他密码技术

加密算法为消息提供了机密性，但信息安全远不止于此，还有更多的问题需要解决。例如，如何保证数据的一致性，确保数据没有被恶意篡改过；如何对信息的来源进行判断，能对伪造来源的信息进行甄别。本节，我们来初步了解一下密码学工具箱中，除加密算法以外的其他几种密码技术。

1.3.1 单向散列函数

有时候，接收者希望能够验证消息在传递的过程中，没有被篡改过，即入侵者不会用假消息冒充合法消息而达到某些非法的目的。

我们经常会发现，在互联网上下载免费软件的时候，有安全意识的软件发布者会在发布软件的同时发布该软件的散列值 (hash)。散列值就是用单向散列函数 (one-way hash function) 计算出来的。这样，下载该软件的人可以自行计算所下载文件的散列值与发布者所发布的散列值进行比较。如果两个散列值一致，就说明下载的软件与发布者所发布的软件是相同的。软件发布者通过发布散列值的方法，可以防止有人在软件里植入一些恶意程序来侵害下载该软件的人的计算机系统。

单向散列函数所保证的并不是机密性，而是完整性 (integrity)。散列值通常又称为哈希值、校验和 (checksum)、指纹 (fingerprint) 或消息摘要 (message digest)。

1.3.2 消息认证码

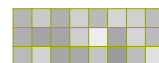
为了确认消息是否来源于所期望的对象，可以使用消息认证码 (message authentication code) 技术。通过消息认证码，不但能够确认消息是否被篡改，而且能够确认消息是否来自于所期望的通信对象。也就是说，消息认证码不仅能够保证完整性，还能够提供认证机制。

1.3.3 数字签名

我们先来看一个例子：供应商给采购方发来邮件，内容是“该商品的采购价格是10万元”。由于这封邮件涉及到数额巨大的交易，如果你是采购人员，肯定会特别小心，一定要核实该邮件确实来自你联系的供应商。仅仅靠邮件发送者的Email地址是不足以判断这封邮件的实际来源，因为邮件的发送者很容易被伪装 (spoofing)。

另一方面，还有这样一种可能，这封邮件确实是来自于采购方所联系的供应商。但是，供应商后来又反悔想提高采购价格，于是便谎称“我当时根本就没发送过那封邮件”。像这样事后否认自己做过某件事情的行为，称为抵赖 (repudiation)。现代商战中，大量充斥着这种案例。

当然，还有一种风险，就是供应商发给采购方的邮件在传输过程中，被别有用心的人篡改，将采购费改成了20万元。数字签名是一项能够同时防止伪装、抵赖和篡改等威胁的密码技术。当供应商对邮件的内容加上数字签名之后再通过邮件一起发送，采购方则可以通过对数字签名 (digital signature) 进行验证 (verify) 来检测出邮件是否被伪装和篡改，还能够防止供应商事后抵赖。



1.3.4 伪随机数生成器

伪随机数生成器 (Pseudo Random Number Generator, PRNG) 用于在系统需要随机数的时候, 通过一系列种子值计算出来的伪随机数。因为生成一个真正意义上的“随机数”对于计算机来说是不可能的, 伪随机数也只是尽可能地接近其应具有随机性, 但是因为有“种子值”, 所以伪随机数在一定程度上是可控可预测的。随机数在密码技术中承担了重要的职责, 例如在访问HTTPS加密站点时进行的TLS通信, 会生成一个仅用于当前通信的临时密钥 (即会话密钥), 这个密钥就是基于伪随机数生成器产生的。如果生成的随机数的算法不够好, 窃听者就有可能推测出密钥, 从而带来通信机密性下降的风险。

1.4 信息安全所面临的威胁及对策

回顾一下, 我们前面初步介绍了六种密码技术:

- 对称加密算法
- 非对称加密算法 (公钥加密算法)
- 单向散列函数
- 消息认证码
- 数字签名
- 伪随机数生成器

我们同时讨论了每种技术所解决的具体问题, 这里把前面的内容再梳理一遍, 用图 1.4所示的思维导视图总结了信息安全所面临的潜在威胁以及针对各种安全威胁所能采用的密码技术及对策。我们没有把伪随机数生成器画在图里面, 是因为它通常渗透在其他五种密码技术中使用, 发挥了非常重要的作用。我们把这六种密码技术统称为密码学家的工具箱。

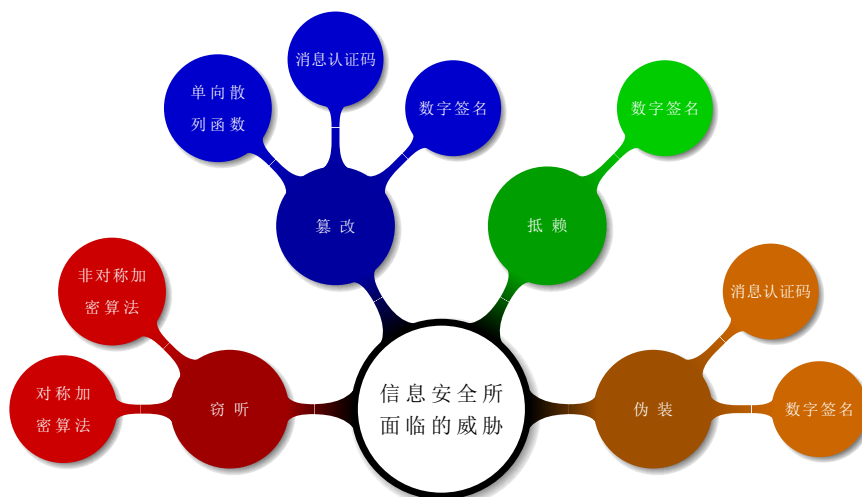


图1.4 信息安全所面临的威胁及其相应的密码技术对策思维导视图

从图 1.4中, 我们可以看到, 有些密码技术可以用来解决信息安全中的多种威胁, 例如, 数字签名可以防止篡改、伪装和抵赖, 但不提供保密。对于某些面临的威胁, 也可能存在多种应对的密码技术, 例如为了防止窃听导致信息被泄露, 可以使用对称加密算法或非对称加密算法。但是每种密码技术都有着各自的特点, 适用于不同的场景。后面章节, 我们会对这些密码技术进行深入的探讨, 逐个揭开它们的神秘面纱。



1.5 密码与信息安全常识

随着信息技术的飞速发展，计算机的计算能力和存储能力正在以惊人的速度不断提升，我们所熟知的摩尔定律到目前为止仍然成立。大数据和物联网应用正在渗透到社会组织的每一个细胞，几乎对所有行业产生颠覆性和革命性的影响。产业的发展环境逐步成熟，网络基础设施支撑能力大幅提升，网络通信的数据量正在呈现指数级爆炸式增长。在人们的生活越来越依赖互联网的时代，信息安全在网络通信中发挥的作用尤为重要，密码技术为保障信息安全提供了全方位的技术支持。本小节从最佳实践的角度阐述我们应该怎么合理地利用密码技术来保障信息的安全。

1.5.1 任何时候都不要尝试发明新的加密算法

刚接触密码技术的软件开发人员，经常会出现这样的想法：我自己设计一个不对外公开的密码算法不就可以保障信息的机密性了吗？这种想法是绝对错误的。加密系统的保密性只应建立在对密钥的保密上，不应该取决于加密算法的保密，这是密码学中的金科玉律。任何时候，我们都不要尝试自己去发明新的加密算法，因为对加密算法的保密是困难的。对手可以用窃取、购买的方法来取得算法、加密器件或者程序。如果得到的是加密器件或者程序，可以对它们进行反向工程而最终获得加密算法。如果只是密钥失密，那么失密的只是和此密钥有关的情报，日后通讯的保密性可以通过更换密钥来补救；但如果是加密算法失密，而整个系统的保密性又建立在算法的秘密性上，那么所有由此算法加密的信息就会全部暴露。

1.5.2 不要使用低强度的密码

很多人对密码的使用有这么一个误区：就算密码强度再低也比不用密码更安全吧。其实，这种想法是非常危险的。与其使用低强度的密码，还不如从一开始就不使用密码。这主要源于用户容易通过“密码”这个词获得一种“错误的安全感”。“信息被加密了”这一事实并不能和信息安全划上句号。攻击者使用暴力穷举（brute-force）等攻击方法就可能破解低强度的密码。

1.5.3 信息安全也是一门社会性课题

有了密码技术，信息安全就能完全得到保证吗？答案是否定的。密码技术只是信息安全的一部分，在信息安全的背景下，社会工程学（social engineering）攻击是一种操纵相关人员泄露出机密信息的攻击方法，建立在使人决断产生认知偏差的基础上，有时候这些偏差被称为“人类硬件漏洞”。犯罪分子利用社会工程学的手法进行诱骗，使受害者不会意识到被利用来攻击网络。当人们没有意识到他们拥有的信息的价值的时候，并不会特意地保护他们所得知的信息，社会工程学正是利用了这一点。

本书不会详细讨论社会工程学攻击，但是为了让大家提高安全意识，防患于未然，特列举以下一些流行的社会工程学攻击：

- 伪装：犯罪分子通过伪装成各种角色来骗取访问权限。例如，伪装成一个看门人、雇员或者客户来获取物理访问权限；冒充贵宾、高层经理或者其他有权或进入计算机系统并察看文件的人。
- 偷窥：通过偷窥方式在他人输入密码时收集他的密码。甚至寻找在垃圾箱中记录密码的纸、电脑打印的文件、快递信息等，往往也可以找到有用的信息。
- 钓鱼：钓鱼涉及虚假邮件、聊天记录或网站设计，模拟与捕捉真正目标系统的敏感数据。比如伪造一条上来自银行或其他金融机构的需要“验证”您登陆信息的信息，来冒充一条合法的登陆页面来骗取你的登录密码。
- 引诱：攻击者可能使用能勾起你欲望的东西引诱你去点击，可能是一场音乐会或一部电影的下载链接，也有可能是你“中奖”需要兑换礼品的链接，或者是商品大力打折的促销链接。一旦点击了这些链接，你的计算机设备或网络就会感染恶意软件以便于犯罪分子进入你的系统。



上面提到的这些攻击手段，都与密码的强度毫无关系。信息安全是一个复杂的系统性工程，其安全程度往往取决于系统中最薄弱的环节。通常，最薄弱的环节不是密码，而是人类自己。“道高一尺魔高一丈”，信息安全上的漏洞和人性上脆弱的环节也不断被不法分子发掘，我们唯有不断的增强自己的安全意识和时刻保持清醒才能更好地防患于未然。

1.6 本章小结

本章，我们初步了解了密码世界里常用的密码技术，并介绍了使用哪种密码技术来应对信息安全中存在的威胁。我们后在后续章节中更详细地介绍每种密码技术的细节，并为应用开发者介绍怎么在工程中使用各种密码技术。



第2章 密码的历史典故

从密码学发展历程来看，可分为古典密码和现代密码两类。古典密码有着悠久的历史，是以字符为基本加密单元的密码。而现代密码则以信息块为基本的加密单元。古典密码和现代密码的分水岭大致就是在计算机问世的时候。本章将重点回顾古典密码的发展历史并分享一些和密码相关的有趣典故和历史事件。

2.1 中国古代人民怎么加密？

早期加密算法主要使用在军事中，中国历史上最早关于加密算法的记载出自于周朝兵书《六韬·龙韬》中的《阴符》和《阴书》。其中《阴符》记载了：

太公曰：“主与将，有阴符，凡八等。有大胜克敌之符，长一尺。破军擒将之符，长九寸。降城得邑之符，长八寸。却敌报远之符，长七寸。警众坚守之符，长六寸。请粮益兵之符，长五寸。败军亡将之符，长四寸。失利亡士之符，长三寸。诸奉使行符，稽留，若符事闻，泄告者，皆诛之。八符者，主将秘闻，所以阴通言语，不泄中外相知之术。敌虽圣智，莫之能识。”

简单来说，阴符是以八等长度的符来表达不同的消息和指令，属于密码学中的替代法（图 2.1），在应用中是把信息转变成敌人看不懂的符号，但知情者知道这些符号代表的含义。

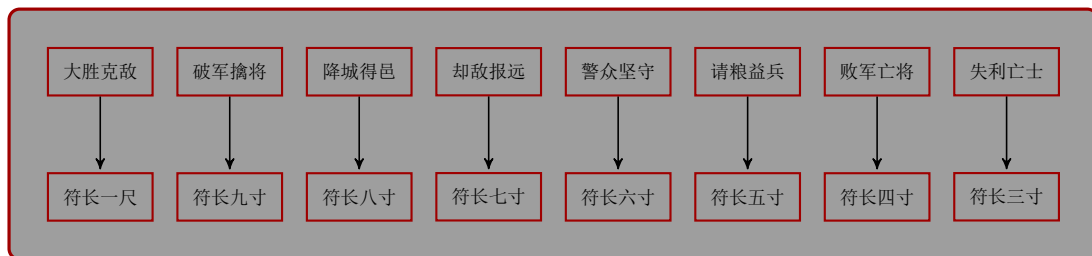


图2.1 阴符所蕴含的加密原理：替换法

阴符只能表述最关键的八种信号，无法表达丰富的含义和传递更具体的消息。所以，《阴书》又作了补充：

武王问太公曰：“引兵深入诸侯之地，主将欲合兵，行无穷之变，图不测之利，其事烦多，符不能明；相去辽远，言语不通。为之奈何？”太公曰：“诸有阴事大虑，当用书，不用符。主以书遗将，将以书问主。书皆一合而再离，三发而一知。再离者，分书为三部。三发而一知者，言三人，人操一分，相参而不相知情也。此谓阴书。敌虽圣智，莫之能识。”

阴书作为阴符的补充，所有密谋大计，都应当用阴书，而不用阴符。国君用阴书向主将传达指示，主将用阴书向国君请示问题，这种阴书都是一合而再离（把一封书信分为三个部分）、三发而一知（派三个人送信，每人负责其中的一部分）。阴书运用了文字拆分法直接把一份文字拆成三份（图 2.2），由三种渠道发送到目标方手中。敌人只有同时截获三份内容才可能破解阴书上写的内容。

无论是阴符，还是阴书，都有着一定的局限性。一是有可能被对方截获而难以达到传递消息的目的，二是有可能被对方破译内容并被对方将计就计加以利用。因此，并不是“敌虽圣智，莫之能识”。张献忠袭取襄阳就说明了这一点。

崇祯十三年七月，张献忠率领起义军突破明军防线，进入四川，杨嗣昌亦率明军十万尾随追击。面对强敌，张献忠挥师东进，于次年二月进入湖北兴山、当阳。在东进途中，起义军活捉了由襄阳（今湖北襄樊市）回四川的杨嗣昌的军使。张献忠从其口中得知杨嗣昌大营所在地襄阳城防空虚，决定奔袭襄阳。他杀掉使者，搜出所携带的兵符，挑选了二十八名起义军战士，换上明军的衣



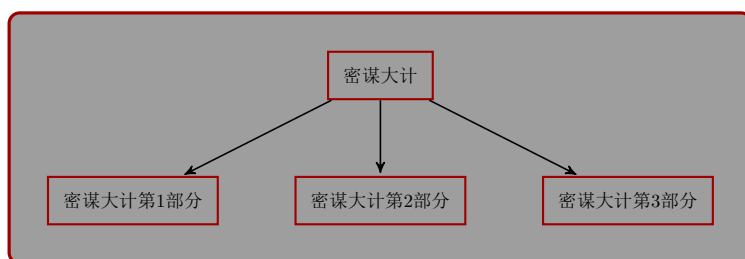


图2.2 阴书所蕴含的加密原理：文字分拆法

服，持兵符先行。张献忠自己则亲率二千精骑，随后跟进，一昼夜急行三百里，直扑襄阳。伪装成明军的起义军士兵到达襄阳时正是夜间，他们自称是督师杨嗣昌派来调运军械的，并出示兵符。守城明军用小筐吊上兵符，细心查验，完全吻合，才命开门放入。城门刚打开，二十八名起义军战士一涌而入，挥刀砍杀守门明军，占领城门。张献忠率领的后续部队恰好赶到，顺利入城。一时杀声震天，明军惊慌失措，被迫投降。起义军杀死襄王朱翊铭，降俘明军数千人，占领襄阳，杨嗣昌闻讯呕血而死。此战表明，无论是阴符还是阴书，都不是万无一失的。

2.2 凯撒密码

我们把视角切换到世界历史的长河中，看看古罗马时期凯撒大帝是怎么使用加密算法对军事信息进行加密的。根据罗马早期纪传体作者盖乌斯·苏维托尼乌斯的记载，恺撒大帝的加密策略很简单，就是把字母按照字母表顺序向后移动几位，但是偏移量（offset）只有他和将军知道，如果移动后超过了字母表中的最后一个字母（对于英文字母表而言就是Z），就回到字母表的第一个字母重新开始下一轮。以英文字母表为例，在偏移量为3的情况下，A将会替换为D，B将会被替换为E，W会被替换为Z，X会被替换为A；明文HELLO会被转换为密文KHOOR。这种加密方法又被称为移位加密。

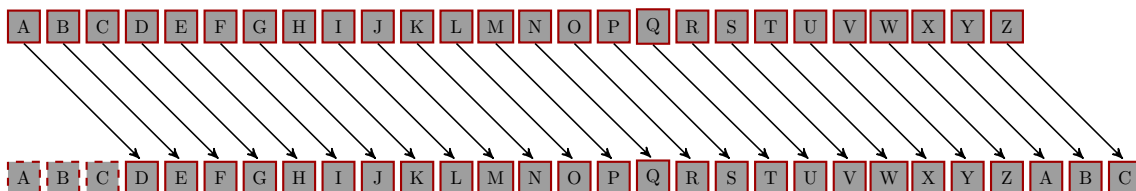


图2.3 凯撒密码的加密原理：替换法

凯撒密码的解密方法也很一目了然，只需要将密文中的每个字母向相反方向平移规定的偏移量便可解密出明文。恺撒密码的加密、解密算法还能够通过同余的数学方法进行计算。首先将字母表中的字母按顺序用数字代替， $A = 0, B = 1, \dots, Z = 25$ 。此时偏移量为 k 的加密算法的数学公式即为：

$$E_k(x) = (x + k) \bmod 26$$

解密算法的数学公式可以表示为：

$$D_k(x) = (x + 26 - k) \bmod 26$$

从加密和解密数学公式可以看出，当偏移量 $k = 13$ 时（字母表内所有字母数量的一半），凯撒密码加密和解密算法的公式完全相同，这是一种特殊的凯撒密码的变种算法，被称为ROT13。ROT13在英文网络论坛常常用作隐藏八卦、妙句、谜题解答以及某些脏话的工具，目的是逃过版主或管理员的匆匆一瞥。因为ROT13的加密和解密计算公式完全相同，很明显，文字经过两次ROT13加密之后，会恢复成原来的文字。

凯撒密码中的偏移量就相当于加密算法中的密钥。这个偏移量必须由发送者和接收者事先约定好。那么，当接收者以外的人窃取到用凯撒密码加密后的密文之后，是不是就无法破解这个密文了呢？或者换句话说，凯撒密码能够被破解吗？



破解密码的复杂度很大程度上取决于密钥空间 (keyspace) 的大小, 所谓密钥空间是指密钥的取值范围到底有多大。凯撒密码中的密钥是偏移量 k , 其取值范围为0至25的整数, 共26种可能的取值, 密钥空间非常有限。攻击者往往可以采用暴力破解 (brute-force attack) 的方法就可以轻而易举地破解凯撒密码。

假设发送者和接收者之间约定的偏移量为3, 那么明文CRYPTOGRAPHY加密后的密文则为FUBSWRJUDSKB。当第三方窃听到密文之后, 由于凯撒密码的密钥空间只有26种可能的取值, 窃听者可以使用穷举搜索 (exhaustive search) 的方法对每种可能的密钥取值尝试一遍:

```
k = 0: FUBSWRJUDSKB => FUBSWRJUDSKB
k = 1: FUBSWRJUDSKB => ETARVQITCRJA
k = 2: FUBSWRJUDSKB => DSZQUPHSBQIZ
k = 3: FUBSWRJUDSKB => CRYPTOGRAPHY
k = 4: FUBSWRJUDSKB => BQXOSNFQZOGX
k = 5: FUBSWRJUDSKB => APWNRMEPYNFW
k = 6: FUBSWRJUDSKB => ZOVMQLDOXMEV
k = 7: FUBSWRJUDSKB => YNULPKCNWLDU
k = 8: FUBSWRJUDSKB => XMTKOJBMVKCT
k = 9: FUBSWRJUDSKB => WLSJNIALUJBS
k = 10: FUBSWRJUDSKB => VKRIMHZKTIAR
k = 11: FUBSWRJUDSKB => UJQHLGYJSHZQ
k = 12: FUBSWRJUDSKB => TIPGKFXIRGYP
k = 13: FUBSWRJUDSKB => SHOFJEWHQFXO
k = 14: FUBSWRJUDSKB => RGNEIDVGPEWN
k = 15: FUBSWRJUDSKB => QFMDHCUFODVM
k = 16: FUBSWRJUDSKB => PELCGBTENCUL
k = 17: FUBSWRJUDSKB => ODKBFASDMBTK
k = 18: FUBSWRJUDSKB => NCJAEZRCLASJ
k = 19: FUBSWRJUDSKB => MBIZDYQBKZRI
k = 20: FUBSWRJUDSKB => LAHYCXP AJYQH
k = 21: FUBSWRJUDSKB => KZGXBWOZIXPG
k = 22: FUBSWRJUDSKB => JYFWAVNYHWOF
k = 23: FUBSWRJUDSKB => IXEVZUMXGVNE
k = 24: FUBSWRJUDSKB => HWDUYTLWFUMD
k = 25: FUBSWRJUDSKB => GVCTXSKVETLC
```

纵览所有尝试的破解, 就会发现只有当 $k = 3$ 的时候, 密文FUBSWRJUDSKB才可以解密出有意义的字符序列 CRYPTOGRAPHY, 即“密码学”的英文单词。因此, 凯撒密码是一种极其不安全的加密方法, 可以被攻击者在很快的时间内破解, 无法保护重要的秘密。

2.3 简单替换密码

2.3.1 什么是简单替换密码

凯撒密码通过将明文中的每个字符按照在字符表中的顺序平移固定数量的字符数来生成密文。由于字符偏移量的取值空间极其有限, 致使凯撒密码能被轻而易举地破解。我们也提到了密钥空间这个概念, 凯撒密码就是因为过小的密钥空间可以被攻击者使用暴力破解的方法在非常快的时间内被破解。你可能意识到凯撒密码这种通过平移字符来实现字符替换的方法过于公式化, 如果把这种映射用随机化的方式打乱, 是不是就完美了呢? 这就是我们接下来要讨论的简单替换密码。

简单替换密码将字母表中的26个字母, 分别与其他字母建立一一映射的关系, 这种映射关系不像凯撒密码那样通过平移字符这种线性化的方法, 而是用一个映射表来描述明文字符和密文字符之间的映射关系, 这种映射表也称为字符替换表。为了更直观地展示字符之间的映射关系, 我们把明文



中的字符都用小写字母表示，密文中的字符都用大写字母表示。图 2.4 就是一个简单的字符替换表。

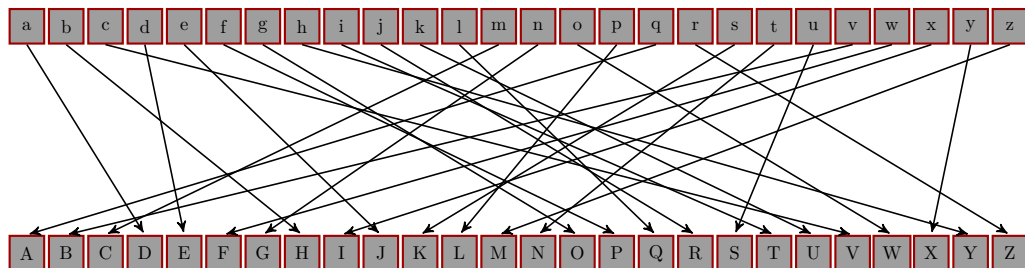


图2.4 简单替换密码的映射表

显然，图 2.4 表示的字符替换关系不像凯撒密码那么有规律，明文字符和密文字符之间的映射看起来是无章可循的。可以说，凯撒密码是简单字符替换密码的一个特例。为了更好地展示明文字符和密文字符之间的替换关系，我们对图 2.4 稍作转换，图 2.5，但仍然保持字符之间原来的映射关系。

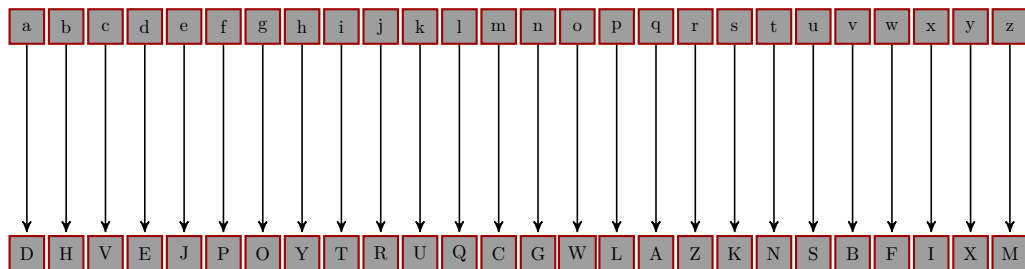


图2.5 变换后的简单替换密码的映射表

凯撒密码可以用暴力破解来破译，但简单替换密码则不然。前面，我们提到，密码算法被破译的困难程度取决于密钥空间的大小。我们来看看简单替换密码的密钥空间。明文字母中的a可以对应A, B, ..., Z这26个字母中的任意一个（26种），b可以对应除了a所对应的字母以外的剩余25个字母中的任意一个（25种）。以此类推，我们可以计算出简单替换密码的密钥空间大小是：

$$26 \times 25 \times 24 \times \dots \times 1 = 403291461126605635584000000$$

这个数字约等于 400×10^{24} ，密钥的数量如此巨大，用暴力破解进行穷举搜索就非常困难了。我们假设以当前（2018年11月）排名第一的Summit超级计算机在峰值性能下每秒约200亿次浮点计算的速度来遍历密钥的话，要遍历完所有的密钥也需要花费超过6千万年的时间。这还是用我们当前最顶级的超级计算机的峰值计算速度来遍历的，普通的家用计算机将要花费几百亿年的时间才能遍历完所有的密钥。由此可见，简单替换密码的密钥空间是足够大的。

2.3.2 用频率分析的方法破解简单替换密码

超大的密钥空间让破译简单替换密码看起来变得不可能，但密码破译工作者发现用频率分析的密码破译方法，使破译简单替换密码成为可能了。

在任何一种书面语言中，不同的字母或字母组合出现的频率各不相同。而且，对于以这种语言写的任意一段文本，都具有大致相同的特征字母分布。比如，在英语中，字母e出现的频率很高，而x出现的较少。类似地，字母组合st、ng、th以及qu等双字母组合出现的频率非常高，nz、qj组合则极少。表 2.1是人们从大量的英文文章中统计出的字母频率。



字母	频率	字母	频率
e	11.1607%	m	3.0129%
a	8.4966%	h	3.0034%
r	7.5809%	g	2.4705%
i	7.5448%	b	2.0720%
o	7.1635%	f	1.8121%
t	6.9509%	y	1.7779%
n	6.6544%	w	1.2899%
s	5.7351%	k	1.1016%
l	5.4893%	v	1.0074%
c	4.5388%	x	0.2902%
u	3.6308%	z	0.2722%
d	3.3844%	j	0.1965%
p	3.1671%	q	0.1962%

表2.1 英文字母出现的频率表

简单替换密码的密钥空间如此巨大，但它的弱点也是显而易见的，就是明文相同的字母在转换为密文后总是被同一个字母所替换。我们参考这个英文字母频率表来实际尝试破译一段密文。现在，假设我们得到下面一段经过简单替换密码加密过后的密文，其明文都是小写英文字母。

EAQASBAEEAQECPENFECQMRQFENABDCQECQSENFBQZQSNBECQUNBLEAFRKKQSECQFZNBVFPBPLSSADFAKAR
 ESPVQARFKASERBQASEAPWQPSUFVPVNBFEFPQPAKESARQZQFPBLOIAGGAFNBVBLECEALNQEAFZQQGBAUAS
 QPBLOIPFZQQGEAFPIDQBLECQCQPSEPTCQPBLECECARFPBLBPERSPZFCATWFECPEKZQFCNFCQNSEAENFPTAB
 FRUUPENABLQHAREZIEAQDNFCLEALNQEAFZQQGEAFZQQGGQSTCPBTQEALSQPUIECQSQFECQSROKASNBECPEF
 ZQQGAKLQPECDCPELSQPUPFUPITAUDCQBDQCPHQFCRKKZQLAKKECNFUASEPZTANZURFEVNHQRFGRPFQECQSQFE
 CQSQFGQTEECPEUPWQFTPZPUNEIAKFAZABVZKNQKASDCADARZLOQPSECQDCNGFPBLFTASBFAKENUQECAGGSQFF
 ASFDSABVECGQSARLUPBFTABERUQZIECQGPBVFAKLNFGSNXLZAHQECQZPDFLQZPIECQNBFAZQBTAQAKKNTQPB
 LECQFGRSBFECPEGPENQBEUQSNEAKECRBDASECIEPWQFDCQBCQNUFQZKUNVCECNFMRNQRERFUPWQDNECPOPSQO
 ALWNBDCADARZLKPSLQZFQOPSEAVSRBEPBLFDQPERBLQSPDQPSIZNKQOREECPEECQLSQPLAKFAUQECNBVPKEQS
 LQPECECQRBLNFTAHQSQLTARBESIKSAUDCAFQOARSBBAESPHQZZQSSQERSBFGXXXZFECQDNZZPBLUPWQFRFSP
 ECQSQPSECAFQNZZFDQCPHQECBPBKZIEAAECQSFECPEDQWBADBAEAKECRFTABFTNQBTLAQFUPWQTADPSLFAKR
 FPZZPBLECRFECQBPENHQCRQAKSQFAZRENABNFNTWZNQLAQSDNECECQGPZQTPFEAKECARVCEPBLQBEQSGSNFQ
 FAKVSQPEGNETCPBLUAUQBEDNECECNFSQVPSLECNSTRSSQBEFERSBPDSIPBLZAFQECQBPQUAKPTENAB

首先，我们统计这段密文中各个字母出现的次数和频率，结果如表 2.2所示。

根据密码研究工作者总结出来的字母频率表 2.1，字母e的出现频率远高于其他字母。经统计后，我们发现密文中字母Q的出现频率最高。我们先暂且假设字母Q就是由e变换而来的，这样，我们把密文中的Q替换回e，就得到下面的字母序列。

EAOeASBAEEAOeECPENFECeMReFENABDCeCeSeNFBAOZeSNBECeUNBLEAFRKKeSECeFZNBVFPBPLSSADFAKAR
 ESPVeARFKASERBeASEAPWePSUFVPVNBFEFPePAKESARQZeFPBLOIAGGAFNBVeBLECeUEALNeEAFZeeGBAUAS
 ePBLOIPFZeeGEAFPIDeeBLECeCePSEPTCePBLECeECARFPBLBPERSPZFCATWFECPEKZeFCNFCeNSEAENFPTAB
 FRUUPENABLeHAREZIEAOeDNFCLEALNeEAFZeeGEAFZeeGGeSTCPBTeeEALSePUIECeSeFECeSROKASNBECPEF
 ZeeGAKLePECDCPELSePUPFUPITAUeDCeBDeCPHeFCRKKZeLAKKECNFUASEPZTANZURFEVNHHeRFGRPFeeCeSeFe
 CeSeFGeTEECPEUPWeFTPZPUNEIAKFAZABVZKNKeKASDCADARZLOePSECeDCNGFPBLFTASBFAKENUeECAGGSefF
 ASFDSABVECeGSARLUPBFTABERUeZIECeGPBVFAKLNFGSNXLZAHHeECeZPDFLeZPIECeNBFAZeBTeAKAKKNTePB
 LECeFGRSBFECPEGPENeBEUeSNEAKECRBDASECIEPWQFDCeBBeCNUFeZKUNVCECNFMRNeERFUPWeDNECPOPSeO
 ALWNBDCADARZLKPSLeZFOPSeAVSRBEPBLFDePERBLESPDePSIZNKeOREECPEECeLSePLAKFAUeECNBVPKEeS
 LePECECeRBLNFTAHSeSeLTARBESIKSAUDCAFeOARSBBAESPHeZZeSSeERSBFGXXXZeFECeDNZZPBLUPWeFRFSP



字母	次数	频率	字母	次数	频率
Q	137	12.47%	K	34	3.09%
E	117	10.65%	D	28	2.55%
A	93	8.46%	U	28	2.55%
P	84	7.64%	T	24	2.18%
F	82	7.46%	G	22	2.00%
C	75	6.82%	O	15	1.36%
S	68	6.19%	I	14	1.27%
B	65	5.91%	V	14	1.27%
N	53	4.82%	W	10	0.91%
L	42	3.82%	H	8	0.73%
Z	41	3.73%	X	3	0.27%
R	40	3.64%	M	2	0.18%

表2.2 密文中各英文字母出现的次数和频率

ECeSOePSECAFeNZZFDeCPHeECPBKZIEAAECeSFECPEDeWBADBAEAKECRFTABFTNeBTeLaeFUPWeTADPSLFAKR
 FPZZPBLECRFECeBPENHeCRaKSeFAZRENABNFFNTWZNeLaeSDNECECeGPZeTPFEAKECARVCEPBLLeBeSGSNFe
 FAKVSePEGNETCPBLUAUeBEDNECECNFSeVPSLECeNSTRSSeBEFERSBPDSIPBLZAFeECeBPUEAKPTENAB

英文文章中，以字母e结尾的单词，the的出现频率极高，对上面的这段字符序列，进一步统计发现ECe出现了27次，远远高于其他以e结尾的包含3个字母的字符串的出现次数。我们进一步假定t被替换成了E，h被替换成了C，于是，我们继续将上面字符序列中E和C分别替换回t和h，得到：

tAOeASBAttAOethPtNFtheMReFtNABDhetheStNFBaoZeSNBtheUNBLtAFRKKeStheFZNBVFpBLPSSADFAKAR
 tSPVeARFKAStrBeASAtPWePSUFVPNBftPFepAKtSARoZefPBLoiAGGAFNBVeBLtheUtalNetAFZeeGBAUAS
 ePBLoiPFZeeGtAFPIDeeBLthehePStPThePBLthethARFPBLBPtRSPZFhATWFthPtKZeFhNFheNStAtNFPTAB
 FRUUPtNABLeHARtZiTAoeDNFhLtALNetAFZeeGtAFZeeGGeStHPTetALSePUPitheSeFtheSROKASNBthPtF
 ZeeGAKLePthDhPtLSePUFUPITAUEdheBDehPheFhRKKZeLAKKthNFUASTpZTANZURFtVNHeRFGPRFetheSeft
 heSeFGeTtthPtUPWeFTPZPUNTIAKFAZABVZNKeKASDhADARZLOePStheDhNGFPBLFTASBFaktNUethAGGSeFF
 ASFDSABVtheGSARLUPBFTABtRueZiTheGPBFVAKLNFSGNXLAHetheZPDFLeZPiTheNBFAZeBTeAKAKKNTePB
 LtheFGRSBFthPtGPTNeBtUeSntAKthRBDASthItPWeFDheBhehNUFeZKUNVhthNFMNNetRFUPWeDNthPOPSeo
 ALWNBdHADARZLKPSLeZFoePStAVSRBtPBLFDePTrBLESPDePSIZNKeOrtthPttheLSePLAKFAUethNBVPKteS
 LePththeRBLNFTAHeSeLTARbtSIKSAUDhAfeOARSBBAtSPHeZZeSSetRSBfGRXXZeFtheDNZZPBLUPWeFRFSP
 theSOePSthAfeNZZFDehPHethPBKZiTAatheSFthPtDeWBADBAAtAKthRFTABFTNeBTeLaeFUPWeTADPSLFAKR
 FPZZPBLthRFtheBPtNHheReAKSeFAZRtNABNFFNTWZNeLaeSDNththeGPZeTPFtAKthARVhtPBLLeBteSGSNFe
 FAKVSePtGnTThPBLUAUeBtDNththNFSeVPSLtheNSTRSSeBtFtRSBPDSIPBLZAFetheBPUEAKPTtNAB

进一步分析，我们发现thPt也多次出现，英文中单词that出现的频率也是特别高的。同时，我们发现P在这段密文中出现的频率也是极高的，我们几乎可以不假思索地猜测a被替换成了P。把P替换回a，我们得到：

tAOeASBAttAOethatNFtheMReFtNABDhetheStNFBaoZeSNBtheUNBLtAFRKKeStheFZNBVFabLaSSADFAKAR
 tSaVeARFKAStrBeASAtaWeaSUFaVanBftaFeaAKtSARoZefaBLoiAGGAFNBVeBLtheUtalNetAFZeeGBAUAS
 eaBLoiafZeeGtAfaiDeeBLtheheaStaTheaBLthethARfaBLBatRSaZfhaTWfthatKZeFhNFheNStAtNFaTAB
 FRUuatNABLeHARtZiTAoeDNFhLtALNetAFZeeGtAFZeeGGeStHABTetALSeaUaitheSeFtheSROKASNBthatF
 ZeeGAKLeathDhatLSeaUFUaITAUEdheBDehaHeFhRKKZeLAKKthNFUASTaZTANZURFtVNHeRFGaRFetheSeft
 heSeFGeTtthatUaWeFtaZaUNTIAKFAZABVZNKeKASDhADARZLOeaStheDhNGfaBLFTASBFaktNUethAGGSeFF
 ASFDSABVtheGSARLUaBFTABtRueZiTheGaBVfAKLNFSGNXLAHetheZaDFLeZaitheNBFAZeBTeAKAKKNTeaB
 LtheFGRSBFthatGatNeBtUeSntAKthRBDASthItaWeFDheBhehNUFeZKUNVhthNFMNNetRFUaWeDNthaOaSeO
 ALWNBdHADARZLKasLeZFoeaStAVSRBtaBLFDeatRBLesaDeaSiZNKeOrtthattheLSeaLAKFAUethNBVaKteS



LeaththeRBLNFTAHeSeLTARBtSIKSAUDhAFeOARSBBAtSaHeZZeSSetRSBFGRRXXZeFtheDNZZaBLUaWeFRFSa
theSOeaSthAFenZZFDehaHethaBKZItAatheSFthatDeWBADBAAtAKthRFTABFTNeBTeLaeFUaWeTADaSLFAKR
FaZZaBLthRFtheBatNHheReAKSeFAZRtNABNFFNTWZNeLaeSDNththeGaZeTaFtAKthARVhtaBLeBteSGSNFe
FAKVSeatGNtThaBLUAUeBtDNththNFSeVaSLtheNSTRSSeBtFtRSBaDSIaBLZAFetheBaUeAKaTtNAB

继续猜测，theSe会不会是there呢，Leath会不会是death呢，于是，我们用r和d分别替换回S和L，得到：

tAOeArBAttAOethatNFtheMReFtNABDhethertNFBAOZerNBtheUNBdtAFRKKertheFZNBVFabDarrADFAKAR
traVeARFKArtrBeArtAtaWearUFaVaNBftaFeaAKtrARozFaBdOIAGGAFNBVeBdtheUtAdNetAFZeeGBAUAr
eaBdOIaFZeeGtAFaIDeeBdtheheartaTheaBdthethARFaBdBatRraZfHATWFthatKZeFhNFheNrtAtNFaTAB
FRUUatNABdeHARtZItAOeDNFhdtAdNetAFZeeGtAFZeeGGerThaBTetAdreaUaIthereFtherROKArNBthatF
ZeeGAKdeathDhatdreaUFUaITAUEdheBDehaHeFhRKKZedAKKthNFUArtaZTANZURFtVNHeRFGaRFethereFt
hereFGeTtthatUaWeFTaZaUNTIAKFAZABVZNKeKArDhADARZdOearthDhNGFaBdFTArBFaKtNUethAGGreFF
ArFDrABVtheGrARdUaBFTABtRUeZIttheGaBVFAKdNFGGrNXdZAHetheZaDFdeZaItheNBFAZeBTeAKAKKNTeaB
dtheFGRrBFthatGatNeBtUerNtAKthRBDArthItaWeFDheBhehNUFeZKUNVhthNFMNRnetRFUaWeDNthaOareO
AdWNBdHADARZdKardeZFoeartAVrRBtaBdFdeatRBderaDearIZNKeORtthatthedreadAKFAUethNBVaKter
deaththeRBdNFTAHeretARBtrIKrAUDhAFeOARrBBatraHeZZerretRrBFGRRXXZeFtheDNZZaBdUaWeFRFra
therOearthAFenZZFDehaHethaBKZItAatherFthatDeWBADBAAtAKthRFTABFTNeBTedAeFUaWeTADardFAKR
FaZZaBdthRFtheBatNHheReAKreFAZRtNABNFFNTWZNeAerDNththeGaZeTaFtAKthARVhtaBdeBterGrNFe
FAKVreatGNtThaBdUAUeBtDNththNFreVardtheNrTRrreBtFtRrBaDrIaBdZAFetheBaUeAKaTtNAB

结合Dhether和Dhat，我们推测D是由w替换过来的。进一步，DNth极有可能就是with，以此类推，DNZZ可能是will，NF可能是is。用w、i和l分别替换D、N和Z，得到：

tAOeArBAttAOethatiFtheMReFtiABwhethertiFBAOleriBtheUiBdtAFRKKertheFliBVFaBdarrAwFAKAR
traVeARFKArtrBeArtAtaWearUFaVaiBFtaFeaAKtrARoleFaBdOIAGGAFiBVeBdtheUtAdietAFleeGBAUAr
eaBdOIaFleeGtAFaIweeBdtheheartaTheaBdthethARFaBdBatRralFhATWFthatKleFhiFheirtAtiFaTAB
FRUUatiABdeHARtliItAOewiFhdtAdietAFleeGtAFleeGGerThaBTetAdreaUaIthereFtherROKArIBthatF
leeGAKdeathwhatdreaUFUaITAUewheBwehaHeFhRKKledAKKthiFUArtaITailURFtViHeRFGaRFethereFt
hereFGeTtthatUaWeFTalaUitIAKFAlABVliKeKarwhAwARldOearthewhiGFaBdFTArBFaKtiUethAGGreFF
ArFwrABVtheGrARdUaBFTABtRUelltheGaBVFAKdiFGriXdlAHethelawFdelaItheiBFaleBTeAKAKKiTeaB
dtheFGRrBFthatGatieBtUeritAKthRBwArthItaWeFwheBhehiUFelKUivhthiFMRIetRFUaWewithaOareO
AdWiBwhAwARldKardelFOeartAVrRBtaBdFweatRBderawearIlikeORtthatthedreadAKFAUethiBvaKter
deaththeRBdiFTAHeretARBtrIKrAUwhAFeOARrBBatraHellerretRrBFGRRXXleFthewillaBdUaWeFRFra
therOearthAFeillFwehaHethaBKlItAatherFthatwewBAwBatAKthRFTABFTieBTedAeFUaWeTAwardFAKR
FallabDthRFtheBatiHheReAKreFAlRtiABiFFiTWliedAerwiththeGaleTaFtAKthARVhtaBdeBterGriFe
FAKVreatGitThaBdUAUeBtwiththiFreVardtheirTRrreBtFtRrBawrIaBdlAFetheBaUeAKaTtiAB

靠近句尾的地方出现了withthisreVard，这个可能是with this regard，也可能是with this reward。但是我们可以立即排除后者，因为我们在上一步已经推测了D是由w替换过来的，所以，我们推测V是由g替换过来的。现在，我们把已经推测出来的字母放到表 2.3 中，如表 2.3 所示。

到此为止，排在前五的高频字母只剩下A还没有推测出来，那我们对照字母频率表 2.1，发现高频字母中只有o和n还没有被反推出来。我们大胆地假设，A就是由o或者n替换过来的，然后，我们通过whA很快排除n，所以，我们推测A是由o替换过来的，继续还原字母序列，我们得到：

toOeorBottoOethatiFtheMReFtioBwhethertiFBoOleriBtheUiBdtoFRKKertheFliBgFaBdarrowFoKoR
trageorRFKortRBeortotaWearUFagaiBFtaFeaoKtroRoleFaBdOIogGoFiBgeBdtheUtodietoFleeGBoUor
eaBdOIaFleeGtoFaIweeBdtheheartaTheaBdthethoRfaBdBatRralFhoTWFthatKleFhiFheirtotiFaToB
FRUUatioBdeHoRtliItOewiFhdtodietoFleeGtoFleeGGerThaBTetodreaUaIthereFtherROKoriBthatF
leeGoKdeathwhatdreaUFUaIToUewheBwehaHeFhRKKledoKKthiFUortalToilURFtgiHeRFGaRFethereFt
hereFGeTtthatUaWeFTalaUitIoKPolobgliKeKorwhoworldOearthewhiGFaBdFTorBFoKtiUethoGGreFF
orFwroBgtheGroRdUaBFTtoBtRUelltheGaBgFoKdiFGriXdlOethelawFdelaItheiBFoleBTeoKoKKiTeaB



字母	次数	频率	字母	次数	频率
Q ← e	137	12.47%	K	34	3.09%
E ← t	117	10.65%	D ← w	28	2.55%
A	93	8.46%	U	28	2.55%
P ← a	84	7.64%	T	24	2.18%
F ← s	82	7.46%	G	22	2.00%
C ← h	75	6.82%	O	15	1.36%
S ← r	68	6.19%	I	14	1.27%
B	65	5.91%	V	14	1.27%
N ← i	53	4.82%	W	10	0.91%
L ← d	42	3.82%	H	8	0.73%
Z ← l	41	3.73%	X	3	0.27%
R	40	3.64%	M	2	0.18%

表2.3 使用频率分析方法已经推测出来的字母

dttheFGRrBFthatGatieBtUeritoKthRBworthItaWeFwheBhehiUFelKUighthiFMRietRFUaWewithaOareO
odWiBhowoRldKardelF0eartogrRBtaBdFweatRBderawearIliKeORtthatthedreadoKFOuethiBgaKter
deaththeRBdiFTtoHeredToRBtrIKroUwhoFeOorRBotraHellerretRrBFGRRXXleFthewillaBdUaWeFRa
therOearthoFeillFwehaHethaBKlIttootherFthatweWbowBotoKthRFTtoBFTieBTedoeFUaWeTowardFoKR
FallabDthRFtheBatiHehReoKreFolRtioBiFFiTWliedoerwiththeGaleTaFtoKthoRghtaBdeBterGriFe
FoKgreatGitThaBdUoUeBtwiththiFregardtheirTRrreBtFtRrBawrIaBdloFetheBaUeoKaTtioB

接下来，我们发现开头几个单词的组合toOeorBottoOethatistheMRestioB，这大概是to be or not to be that is the question吧，其中，b→O，n→B，q→M，u→R。进一步将这些字母替换进去，得到：

tobeornottobethatiftthequeftionwhethertifnoblerintheUindtoFuKKertheFlingFandarrowsFoKou
trageouFKfortuneortotaWearUFagainFtaFeaoKtroubleFandbIoGGofingendtheUtodietoFleeGnoUor
eandbIaFleeGtoFaIweendtheheartaTheandthethouFandnaturalFhoTWfthatKleFhiFheirtotiFaTon
FuUUationdeHoutlItobewiFhdtodietoFleeGtoFleeGGerThanTetodreaUaIthereFtherubKorinthatF
leeGoKdeathwhatdreaUFUaITtoUewhenwehaHeFhuKKledoKKthiFUortalToilUuFtgiHeuFGauFethereFt
hereFGeTtthatUaWeFTalaUitIoKPolongliKeKorwhowouldbearthewhiGFandFTornFoKtiUethoGGreFF
orFwrongtheGroudUanFTontuUelltheGangFoKdiFGriXdloHethelawFdelaItheinFolenTeoKoKKiTea
dttheFGurnFthatGatientUeritoKthunworthItaWeFwhenhehiUFelKUighthiFquietuFUaWewithabareb
odWinwhowouldKardelFbeartogruntandFweatunderawearIliKebutthatthedreadoKFOuethingaKter
deaththeundiFTtoHeredTountrIKroUwhoFebournnotraHellerreturnFGuXXleFthewillandUaWeFuFra
therbearthoFeillFwehaHethanKlIttootherFthatweWnownotoKthuFTonFTienTedoeFUaWeTowardFoKu
FallandthuFthenatiHehueoKreFolutioniFFiTWliedoerwiththeGaleTaFtoKthoughtandenterGriFe
FoKgreatGitThandUoUentwiththiFregardtheirTurrentFturnawrIandloFethenaUeoKaTtion

至此，我们的破译工作基本结束。我们基本上可以确定这段密文就是莎士比亚名著《哈姆雷特》中关于“生存和毁灭”的名段了。我们把原著和现在已经部分破译好的字母序列对比，就能确定所有字母的替换关系如图 2.6 所示。

明文如下：

tobeornottobethatisthequestionwhethertisnoblerinthemindtosuffertheslingsandarrowsfofou
trageousfortuneortotakearmsagainsta seaof troublesandbyopposingendthetodietosleepnomor
eandbyasleeptosayweendtheheartacheandthethousandnaturalshocksthatfleshisheirtotisacon
summationdevoutlytobewishdtodietosleeptosleepperchancetodreamaytherestherubforinthat



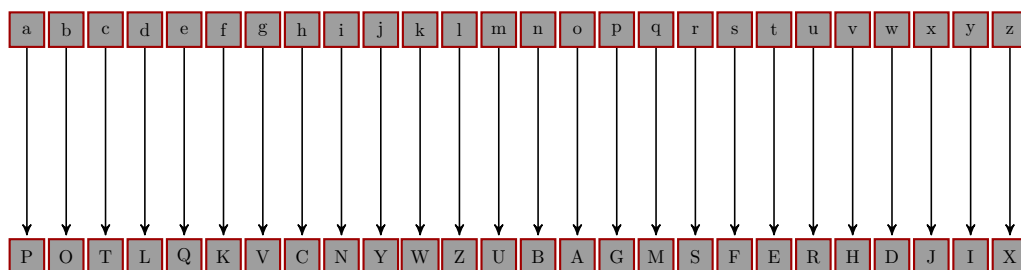


图2.6 变换后的简单替换密码的映射表

leepofdeathwhatdreamsmaycomewhenwehaveshuffledoffthismortalcoilmustgiveuspauseetherest
 herespectthatmakescalamityofsolonglifeforwhowouldbearthewhipsandscornsoftimethoppress
 orswrongtheproudmanscontumelythepangsofdisprizdlovethelawsdelaytheinsolenceofofficean
 dthespurnsthatpatientmeritofthunworthytakeswhenhehimselfmightthisquietusmakewithabareb
 odkinwhowouldfdardelsbeartogruntandsweatunderawearylifebutthatthedreadofsomethingafter
 deaththeundiscoveredcountryfromwhosebournnotravellerreturnspuzzlesthewillandmakesusra
 therbearthoseillswehavethanflytoothersthatweknownotofthusconsciencedoesmakecowardsofu
 sallandthusthenativehueofresolutionissickliedoerwiththepalecastofthoughtandenterprise
 sofgreatpitchandmomentwiththisregardtheircurrentsturnawryandlosethenameofaction

给明文补上空格和标点符号并断句之后，可读性就更好了：

To be, or not to be, that is the question:
 Whether 'tis nobler in the mind to suffer
 The slings and arrows of outrageous fortune,
 Or to take arms against a sea of troubles
 And by opposing end them. To die-to sleep,
 No more; and by a sleep to say we end
 The heart-ache and the thousand natural shocks
 That flesh is heir to: 'tis a consummation
 Devoutly to be wish'd. To die, to sleep;
 To sleep, perchance to dream-ay, there's the rub:
 For in that sleep of death what dreams may come,
 When we have shuffled off this mortal coil,
 Must give us pause-there's the respect
 That makes calamity of so long life.
 For who would bear the whips and scorns of time,
 Th'oppressor's wrong, the proud man's contumely,
 The pangs of dispriz'd love, the law's delay,
 The insolence of office, and the spurns
 That patient merit of th'unworthy takes,
 When he himself might his quietus make
 With a bare bodkin? Who would fardels bear,
 To grunt and sweat under a weary life,
 But that the dread of something after death,
 The undiscovere'd country, from whose bourn
 No traveller returns, puzzles the will,
 And makes us rather bear those ills we have
 Than fly to others that we know not of?
 Thus conscience does make cowards of us all,
 And thus the native hue of resolution



Is sicklied o'er with the pale cast of thought,
And enterprises of great pitch and moment
With this regard their currents turn awry
And lose the name of action.

通过上述破解过程，我们可以了解到利用频率分析破译简单替换密码可以从高频字母着手，同时利用高频单词查找线索。常用的词组也可能成为线索，同时密文越长越容易破解，因为长密文统计出来的字母频率表更接近密码工作研究者们总结出来的字母频率表。

早在公元九世纪，阿拉伯的密码破译专家就已经能够娴熟地运用统计字母出现频率的方法来破译简单替换密码，柯南·道尔在他著名的福尔摩斯探案《跳舞的小人》里就非常详细地叙述了福尔摩斯使用频率统计法破译跳舞人形密码（也就是简单替换密码）的过程。

2.4 复式替换密码: Enigma

Enigma这个名字在德语里是“谜”的意思，它是由德国人阿瑟·谢尔比乌斯 (Arthur Scherbius) 发明的一种能够进行加密和解密操作的机器。在刚刚发明之际，Enigma被用在商业用途，后来到了第二次世界大战期间，纳粹德国国防军使用Enigma并将其改良后用于军事用途。

2.4.1 Enigma的构造

Enigma加密机的外形如图 2.7所示，它是一种由键盘、齿轮、电池和灯泡所组成的机器，通过这一台机器就可以完成加密和解密两种操作。

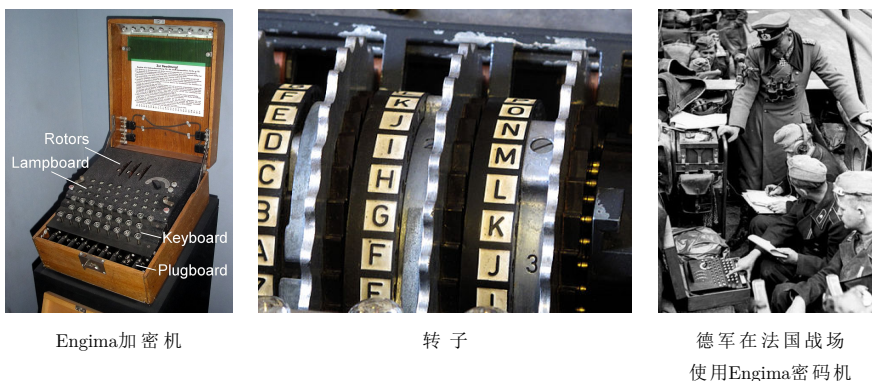


图2.7 Enigma

键盘上一共有26个按键，键盘排列和广为使用的计算机键盘基本一致，只不过为了使通讯尽量地短和难以破译，空格、数字和标点符号都被取消，而只有字母键。键盘上方就是“显示器” (Lampboard)，这可不是现今的计算机屏幕显示器，只不过是标示了同样字母的26个小灯泡。当键盘上的某个字母键被按下时，这个字母被加密后的密文字母所对应的小灯泡就亮了起来，就是这样一种近乎原始的“显示”。在显示器的上方是三个直径6厘米的转子 (Rotor)，转子是Enigma密码机最核心关键的部分。如果转子的作用仅仅是把一个字母转换成另一个字母，那就是等同于我们前一节介绍的简单替换密码。转子的巧妙之处在于它会旋转，每按下键盘上的一个字母键，相应加密后的字母在显示器上通过灯泡闪亮来显示，而转子就自动地转动一个字母的位置。这样，连续多次按下同一个字母键经过加密之后的密文字母都不相同。这就是Enigma难以被破译的关键所在，这不是一种简单替换密码。同一个字母在明文的不同位置时，可以被不同的字母替换，而密文中不同位置的同一个字母，又可以代表明文中的不同字母，字母频率分析法在这里丝毫无用武之地了。这种加密方式在密码学上也被称为复式替换密码。

但是如果连续键入26个字母，转子就会整整转一圈，回到原始的方向上，这时编码就和最初重复了。而在加密过程中，重复的现象就是最大的破绽，因为这可以使破译密码的人从中发现规律。于



是Enigma又增加了其他的转子，当前一个转子转动整整一圈以后，它上面有一个齿轮拨动下一个转子，使得它的方向转动一个字母的位置。而事实上，德军使用的Enigma有3个转子（德国防卫军版）或4个转子（德国海军M4版和德国国防军情报局版）。以Enigma密码机上配置了3个转子为例，重复的概率就达到了 $26 \times 26 \times 26 = 17576$ 个字母之后。

除此以外，在第一个转子之前和最后一个转子之后分别加上了一个接线板和反射器。接线板允许操作员设置各种不同的线路。接线板上的每条线都会连接一对字母，其作用就是在电流进入转子前改变它的方向。例如，将A插口和F插口连接起来，当操作员按下A键时，电流就会流到F插口（相当于按下了F键）再进入转子。电流进入转子前方向被改变，增强了Enigma的保密性。接线板上最多可以同时接13条线。

反射器和转子的显著区别在于它并不转动，它仅仅将最后一个转子的其中两个触点连接起来。乍一看这么一个固定的反射器好像没什么用处，它并不增加可以使用的编码数目，其精妙之处在于，让电流重新折回转子，把它和解密联系起来就会看出这种设计的别具匠心了。为了解释Enigma密码机的工作原理，我们用图 2.8和图 2.9来分别说明第一次和第二次按下A键的时候，Enigma是怎么加密的。

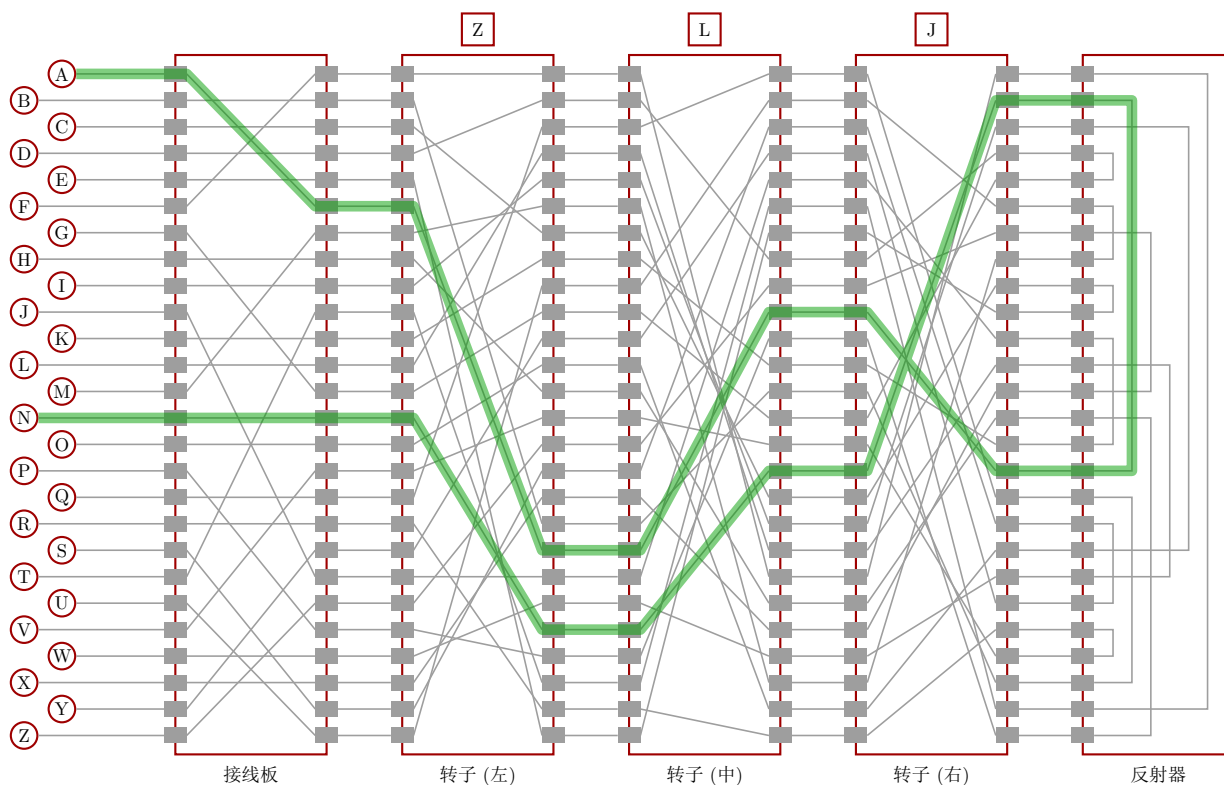


图2.8 Enigma电路布线示意图：第一次按下A键

首先，我们假设左、中、右三个转子的位置分别对应字母Z、L、J，如图 2.8 所示。字母键A按下时，电流先流到接线板上的A插口，由于接线板上A插口和F插口连接起来了，电流方向被改变，从F插口流出后进入到左边第一个转子的F插口。之后，依次经过所有转子，每个转子都会对电流的方向进行转换，即对字母进行替换。当电流从右边最后一个转子的P插口出来之后，经过反射器改变方向进入最后一个转子的B插口。此后，电流沿相反方向依次经过所有转子，最后从接线板N插口出来，点亮 N灯泡。这个就是加密的整个过程。在当前的设置下，如果这时按的不是A键而是N键，那么电流信号恰好按照前面A键被按下时的相反方向同行，最后到达A灯泡。换句话说，在这种转子的设置下，反射器使得解密过程完全重现加密过程。



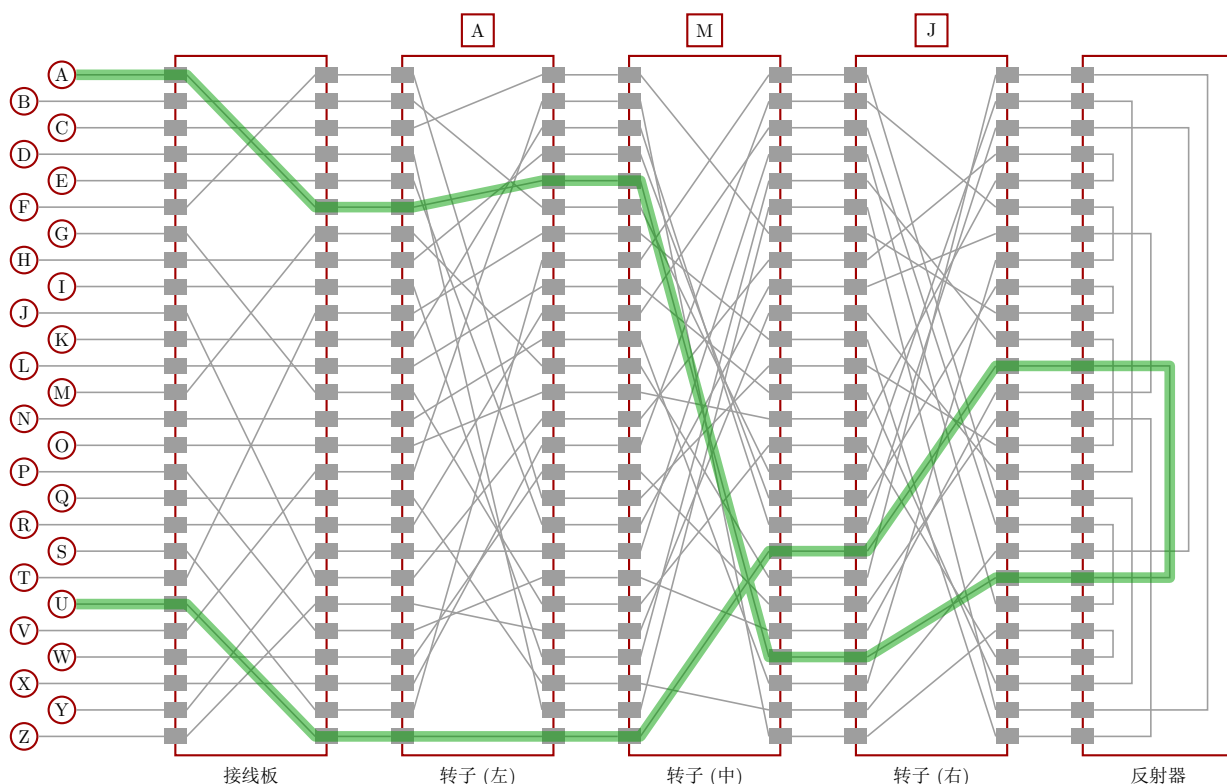


图2.9 Enigma电路布线示意图：第二次按下A键

再次按下字母键A，左边的转子转动一格回到字母A的位置，同时带动中间的转子转动一格到M的位置，右边的转子不动，仍然停留在J的位置，如图 2.9所示。在此时的转子的设置下，按下A键，U灯泡亮起。同样，如果这时按下的是U键，则点亮的是A灯泡。反射器再次完美地使得解密过程重现了加密过程。

从数学的角度，Enigma对每个字母的加密和解密过程可以看作由多步字符替换而组合在一起的过程。我们用 P 表示接线板的连线所对应的字符替换， L 、 M 、 R 分别表示左、中、右3个转子所对应的字符替换， U 表示反射器所对应的字符替换。其中接线板和反射器对应的字符替换 P 和 U 是一经设置就不再变化的。三个转子对应的字符替换 L 、 M 、 R 则会随着字符在明文消息中的位置不同发生变化，我们用下标 k 来表示它们在第 k 个字符的替换。另外，当电流经过反射器后折回沿反方向经过转子和接线板的过程正好是之前字符替换的反操作，我们用上标 -1 来表示这些字符替换的反操作。因此，明文中第 k 个字符 x_k 被加密后的字符 $E_k(x)$ 可以用如下数学公式表示：

$$E_k(x) = P^{-1} L_k^{-1} M_k^{-1} R_k^{-1} U R_k M_k L_k P x_k$$

从上述加密过程对应的数学公式可以看出，Enigma构造具有完美的对称性。解密和加密具有相同的组合过程。密文中第 k 个密文字符 $E_k(x)$ 被解密还原出明文字符 x_k 可以用相同的数学公式表示：

$$x_k = P^{-1} L_k^{-1} M_k^{-1} R_k^{-1} U R_k M_k L_k P E_k(x)$$

2.4.2 Enigma的加密过程

发送者和接收者需要各自拥有一台Enigma密码机。发送者用Enigma对明文加密，记录生成的密文并通过无线电发送给接收者。接收者收到密文后用自己的Enigma解密，还原出明文。

发送者和接收者会事先收到一份叫做国防军密码本的册子，这个册子中记载了发送者和接收者所使用的每日密码。Enigma的加密过程如图 2.10所示，具体描述如下：



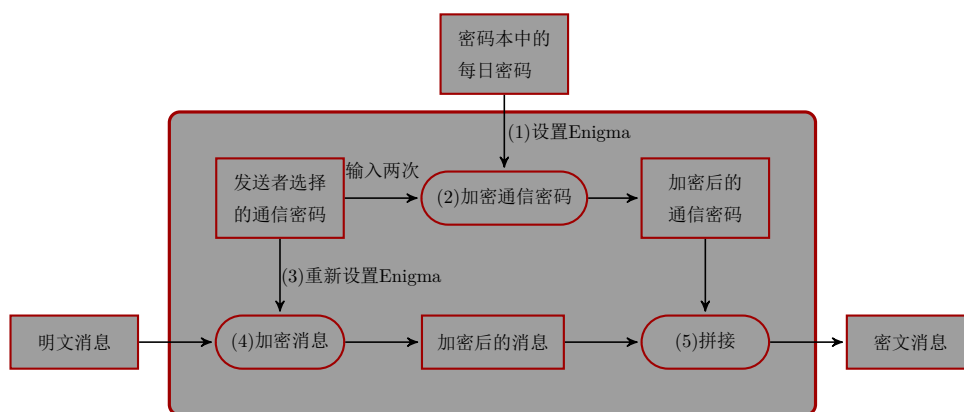


图2.10 Enigma的加密过程

(1) 设置Enigma

发送者查阅国防军密码本，找到当天的每日密码，并按照该密码设置Enigma，具体来说，这个每日密码描述了如果操作接线板上的接线并设置3个转子排列顺序和每个转子的初始位置。

(2) 加密通信密码

接下来，发送者要想出3个字母，并将其加密。这3个字母称为通信密码。通信密码的加密也是用Enigma完成的。假设发送者选择的通信密码是cat，则发送者需要在Enigma的键盘上输入两次该通信密码，即catcat。发送者观察亮起的灯泡对应的字符并记录这6个字母加密后的密文，我们用大写字母来假设得到的密文字母是PCVTAM。

(3) 重新设置Enigma

接下来，发送者根据通信密码重新设置Enigma。通信密码中的3个字母就代表了3个转子的初始位置。也就是说，左、中、右三个转子分别转到c、a、t的位置。

(4) 加密消息

发送者从键盘上逐字输入明文消息的字符，并从灯泡中读取所对应的字母并记录下来。

(5) 拼接

最后，发送者将加密后的通信密码和加密后的消息拼接在一起，通过无线电发送给接收者。

2.4.3 Enigma的解密过程

接收者收到密文消息之后，解密过程如图 2.11所示，具体操作步骤如下：

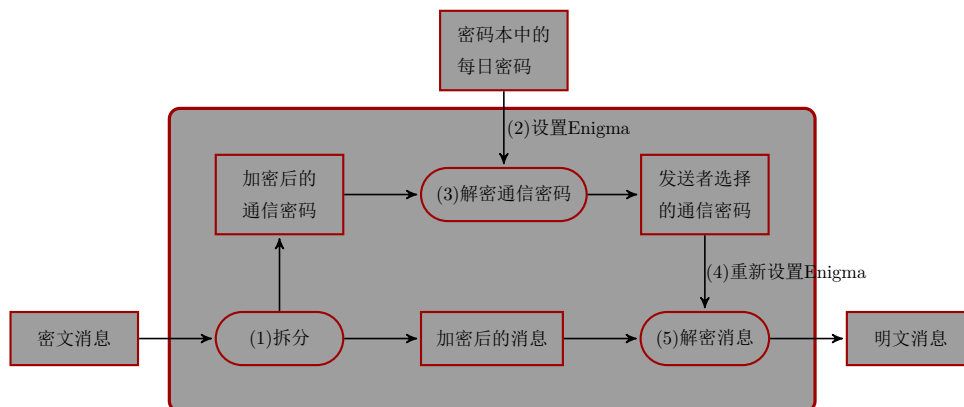
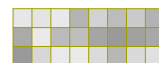


图2.11 Enigma的解密过程



(1) 拆分

接收者将密文消息拆分成两个部分，即开头的6个字母PCVTAM和剩下的字母序列。

(2) 设置Enigma

像发送者一样，接收者查阅国防军密码本，找到当天的每日密码，并按照该密码设置Engima。

(3) 解密通信密码

开头的6个字母PCVTAM即加密后的通信密码，接收者用Enigma对其进行解密，得到catcat。因为catcat是cat重复两次的组合，这样，接收者也可以判断密文消息在通信的过程是否发生错误。

(4) 重新设置Enigma

接收者根据解密后的通信密码cat重新设置Enigma三个转子的初始位置。

(5) 解密消息

接收者用当前Enigma的设置，对密文消息剩下部分的字母序列进行解密，得到明文消息内容。

2.4.4 每日密码和通信密码

通过前面对Enigma加密和解密过程的描述，我们注意到Enigma中出现了每日密码和通信密码这两种不同的密钥。在Enigma中，每日密码被用来加密通信密码，而不是用来加密消息的，消息是用通信密码加密的。也就是说，每日密码是一种用来加密密钥的密钥。这样的密钥，被称为密钥加密密钥 (Key Encrypting Key, KEK)。KEK在现代加密算法中依然被广泛使用。后面，我们在介绍混合密码系统时还会多次遇到这一概念。

2.4.5 Enigma的弱点

我们已经了解了Enigma的加密和解密过程，相比较于简单替换密码，Engima的确要复杂得多。但我们仍然能找到Enigma的一些弱点。

明文中的字母被Enigma加密之后永远不会被替换成该字母本身。 Enigma反射器在电流重新进入转子之前，改变了电流方向，无论接线板怎么接线以及三个转子的顺序和每个转子的旋转位置如何改变，输入的字母都绝对不会被替换成该字母本身。第二次世界大战中，英国军队的密码破译者截获了一段 Enigma的密文，他们发现在密文中字母L从未出现。密码破译者根据这一事实推测出明文是一段只有字母L的文字。发送者的目的是将毫无意义的明文加密发送以干扰密码破译者。发送者本想干扰密码破译者，没想到却反而为破译者提供了线索。

通信密码的弱点。 通信密码太短，被加密后只有6个字母。密码破译者可以知道，密文开头的6个字母就是通信密码被连续输入两次而加密的。而且，Enigma在加密通信密码这一重要步骤中，绝大部分情况下只有最左边的转子会旋转，只有当左边的转子设置到U之后的字母时，才可能带动中间的转子旋转。这个特点也可能被密码破译者利用。

国防军的每日密码本也是一个弱点。 国防军的每日密码本是使用Enigma的必要操作手册。因为发送者和接收者都得使用这个密码本，如果这个密码本落到敌人手里，就必须作废这个已经派发到全军的密码本，而不得不重新制作新的密码本。同时，如何安全地把这个密码本配送到全军中也是一个问题。这个话题，就是我们今后要在介绍现代密码通信时要详细探讨的密钥配送问题。

2.4.6 Enigma的破译

Enigma在当时被认为是一种无法破译的密码机，德军的一份对Enigma的评估写道：“即使敌人获取了一台同样的机器，它仍旧能够保证其加密系统的保密性。” Enigma的设计并不依赖Enigma的构造（相当于加密算法），只要不知道Enigma的设置（相当于密钥），就无法破译密码。Enigma的这个设计理念已经契合了现代密码体系的思想，即加密系统的保密性只应建立在对密钥的保密上，不应该取决于加密算法的保密。Enigma的设置由每日密码所决定，具体表现为3个转子的排列顺序、每个转子的初始位置、以及接线板连线的状况。我们先来看看要暴力破解，需要实验多少种可能性：



- 3个转子的排列顺序存在6种可能性；
- 3个转子初始位置存在 $26 \times 26 \times 26 = 17,576$ 种可能性；
- 接线板上两两交换6对字母的可能性则异常庞大，有100,391,791,500种。

于是一共有 $17576 \times 6 \times 100,391,791,500$ ，其结果大约为10,000,000,000,000! 即一亿亿种可能性！这样庞大的可能性，换言之，即便能动员大量的人力物力，要想靠暴力破解法来逐一试验可能性，那几乎是不可能的。

1931年11月8日，法国情报人员通过间谍活动搞到了Engima的操作和内部线路的资料，但是法国还是无法破译它，因为Enigma的设计要求就是要在机器被缴获后仍具有高度的保密性。当时的法军认为，由于凡尔赛条约限制了德军的发展，也就没有花费人力物力去破译它。与法国不同，第一次世界大战中新独立的波兰的处境却很危险，西边的德国根据凡尔赛条约割让给了波兰大片领土，德国人对此怀恨在心，而东边的苏联也在垂涎着波兰的领土。所以波兰需要时刻了解这两个国家的内部信息。在科学的其他领域，我们说失败乃成功之母；而在密码分析领域，我们则应该说恐惧乃成功之母。这种险峻的形势造就了波兰一大批优秀的密码学家。Enigma最终由波兰密码学家马里安·雷杰夫斯基 (Marian Rejewski) 破译。

雷杰夫斯基深知“重复乃密码大敌”。在Enigma密码中，最明显的重复莫过于每条电文最开始的那六个字母，它由三个字母的密钥重复两次加密而成。德国人没有想到这里会是看似固若金汤的Enigma防线的弱点。雷杰夫斯基每天都会收到一大堆截获的德国电报，所以一天中可以得到许多这样的六个字母串，它们都由同一个当日密钥加密而成。通过分析这些电文的前六个字母串，雷杰夫斯基总结出Enigma的数量巨大的密钥主要是由接线板来提供的，如果只考虑转子的排列顺序和它们的初始位置，只有 $6 \times 17576 = 105,456$ 种可能性。虽然这还是一个很大的数字，但是把所有可能性都试验一遍，已经是一件可以做到的事情了。雷杰夫斯基和同事根据情报复制出了Enigma样机，并在Enigma的基础上设计了一台能自动验证所有 $26 \times 26 \times 26 = 17,576$ 个转子位置的机器，为了同时试验三个转子的所有可能的排列顺序，就需要6台同样的机器，这样就可以试遍所有的 $6 \times 17576 = 105,456$ 种转子排列顺序和初始位置。所有这6台Enigma和为使它们协作的其他器材组成了一整个大约一米高的机器，能在两小时内找出当日密钥。

2.5 本章小结

我们在本章回顾了历史上一些经典的加密算法及其典故，从中我们知道，古典密码多使用替换法进行加密和解密，并详细介绍了简单替换和复式替换两种加密方法。同时，我们还讨论了这些古典密码所面临的问题，并尝试用暴力破解和频率分析的方法分别破译了凯撒密码和简单替换密码。通过本章的学习，我们了解到：

1. 暴力破解适用于破译密钥空间较小的加密算法，频率分析可以用于破解简单的单表替换密码。
2. 加密系统的保密性只应建立在对密钥的保密上，不应该取决于加密算法的保密。





第3章 对称加密算法

古典加密算法那些针对字符进行转换的加密方式在计算机技术所主导的信息时代已经没什么实用价值了。计算机的操作对象并非针对文字,而是由0和1组成的比特序列。无论是文字、图像、声音、视频还是程序,在计算机中都是用比特序列来表示的。将这些现实世界中的东西转换成比特序列的操作被称为编码(encoding)。现代加密技术是以编码后的比特序列为基本加密单元的。现代加密算法总体上分为两大类:对称加密算法和非对称加密算法,如图3.1所示。

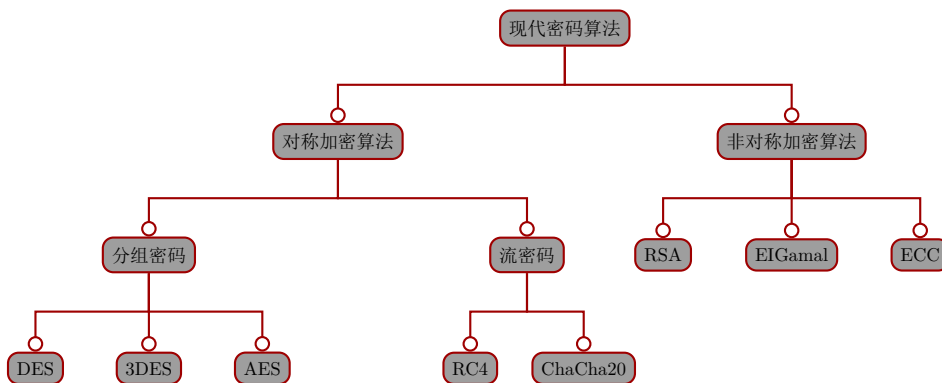


图3.1 现代加密算法

其中,对称加密算法还分为流密码和分组密码。本章将具体介绍几种常用的对称加密算法,包括RC4和ChaCha20两种流密码,以及DES、3DES和AES三种分组密码。

3.1 流密码和分组密码

在对称加密算法中,按照加密方式的不同,可以分为流密码和分组密码。

在流密码加密方式下,明文称为明文流,以比特序列的方式表示。加密时候,先由种子密钥生成一个密钥流。然后利用加密算法把明文流和密钥流进行加密,产生密文流。流密码一般以1比特、8比特、或32比特为单位进行加密和解密。加密过程所需要的密钥流由种子密钥通过密钥流生成器产生。

流密码的主要原理是通过随机数发生器产生性能优良的伪随机序列,使用该序列加密明文流得到密文流。若密钥流是无周期、无限长随机序列,则每一个明文都对应一个随机的加密密钥,理论上,流密码是“一次一密”密码体制,也就是绝对安全的。实际应用中密钥流都用有限存储和复杂逻辑的电路产生,此时它的生成器只有有限个状态,这样,它早晚要回到初始状态而呈现出一定长度的周期,其输出也就是周期序列。所以,实际应用中的流密码不会实现“一次一密”密码体制,但若生成的密钥流周期够长,随机性好,其安全强度还是能保证的。因此,密钥流生成器的设计是流密码的核心,流密码的安全强度取决于密钥流的周期、复杂度、随机(伪随机)性等。流密码涉及许多理论知识,提了很多设计原理,得到了广泛分析,但很多研究成果并没有全部公开,可能是因为目前流密码主要用于军事和外交。日前,公开的流密码算法主要有RC4、SEAL等。

不同于流密码加密方式,分组密码(又称块密码)在把明文变成二进制比特序列后,将其划分成若干个固定长度的组,不足位用0补全。分组长度通常是64比特、128比特、192比特和256比特等。分组密码对每个分组逐个依次进行加密操作。分组长短影响密码强度,分组长度不能太短也不能太长。既要便于操作与运算,又要保证密码的安全性。本章将详细介绍几种分组密码算法:DES、3DES、AES和Blowfish。分组密码算法一次只能处理固定长度的分组,但是我们需要加密的明文长度可能会超过分组密码的分组长度,这时就需要对分组密码算法进行迭代,以便将一段很长的明文全部加密,迭代的方法就称为分组密码的分组模式(mode)。我们将把分组密码的模式留在下一章进行讲解。



3.2 RC4流式加密算法

RC4 (Rivest Cipher 4) 是由罗纳德·李维斯特 (Ron Rivest) 在1987年开发出来的, RC4开始时是商业密码, 没有公开发表出来, 但是在1994年9月份的时候, 它被人匿名公开在了密码朋克 (Cypherpunks) 邮件列表上, 随后又传播到了互联网的许多站点被公开, RC4也就不再是商业秘密了, 只是它的名字“RC4”仍然是一个注册商标。RC4已经成为一些常用的协议和标准的一部分, 如1997年的WEP和2003/2004年无线卡的WPA; 1995年的SSL, 以及后来1999年的TLS。让它如此广泛分布和使用的主要因素是它不可思议的简单和速度, 不管是软件还是硬件, 实现起来都十分容易。

RC4算法的原理很简单, 包括初始化算法 (Key Scheduling Algorithm, KSA) 和伪随机子密码生成算法 (Pseudo Random Generation Algorithm, PRGA) 两大部分。

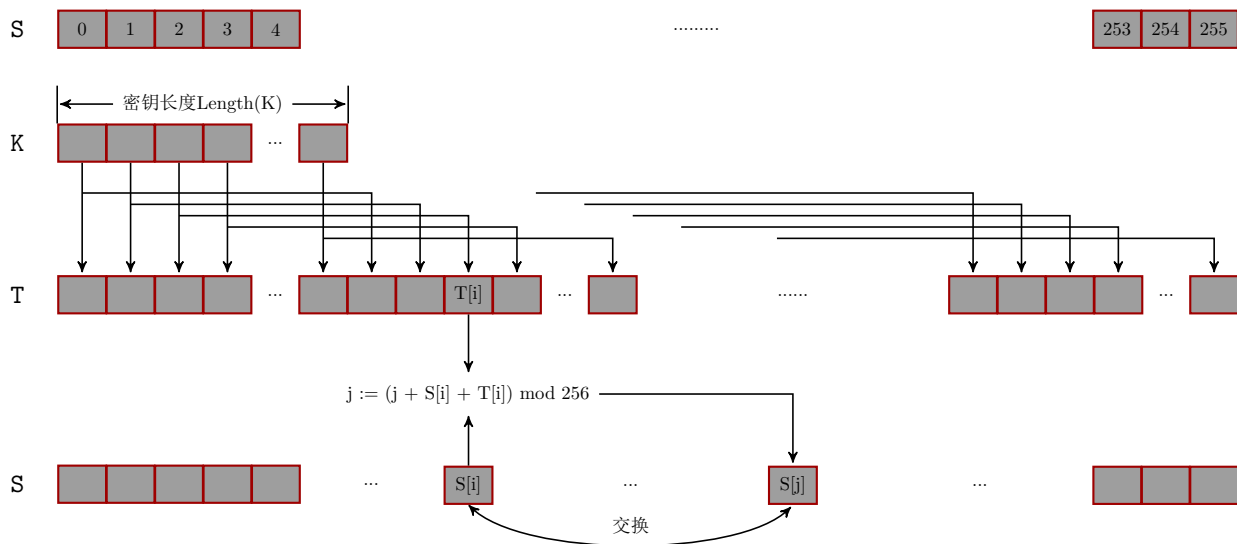


图3.2 RC4初始化算法对S盒的初始化过程示意图

3.2.1 初始化算法

RC4初始化阶段, 由种子密钥K对一个称之为S盒的数组进行初始化。我们用Length(K)来表示种子密钥的长度, 并假设S盒的长度为256, RC4初始化过程可以用图 3.2和如下伪代码来描述。

RC4初始化算法的伪代码描述

```
for i from 0 to 255 do
    S[i] := i
    T[i] := K[i mod Length(K)]

j := 0
for i from 0 to 255 do
    j := (j + S[i] + T[i]) mod 256
    swap values of S[i] and S[j]
```

第一个for循环对S盒和数组T进行初始化。S盒中按顺序依次填入索引序号, 种子密钥K中的字节序列依次循环填入数组T来完成对T的初始化。第二个for循环将S盒中的数值搅乱, i确保S中的每个元素都得到处理, j的选择通过种子密钥K来保证S的搅乱是随机的。

3.2.2 伪随机密码生成算法

在RC4伪随机密码生成算法阶段, 基于初始化好的S盒, 对明文流加密生成密文流。算法是一个迭代的过程, 图 3.3给出了一次迭代过程的示意图, 整个过程用如下伪代码描述。





```
i := 0
j := 0
while (inputByte = readByte(inStream)) != EOF:
    i := (i + 1) mod 256
    j := (j + S[i]) mod 256
    swap values of S[i] and S[j]
    k := (S[i] + S[j]) mod 256
    outputByte = inputByte ^ S[k]
    writeByte(outStream, outputByte)
```

PRGA生成密文流的过程中,只采用了异或运算,由于异或运算具有对称性,即由 $a \oplus b = x$,可以得到 $x \oplus b = a$ 和 $x \oplus a = b$ 。因此,RC4解密也使用同一套算法。

OpenSSL是密码工作者和开发人员的必备利器，它是一个开放源代码的软件库包，整个软件包大概可以分成三个主要的功能部分：

- 作为一个基于密码学的安全开发包，OpenSSL提供的功能相当强大和全面，囊括了主要的密码算法、常用的密钥和证书封装管理功能以及SSL协议，并提供了丰富的应用程序供测试或其它目的使用。在介绍完每项具体的现代密码技术之后，本书会使用OpenSSL来展示具体的例子，帮助读者加深对密码技术的理解。

首先，我们使用openssl对字符串hello进行加密，指定加密算法为RC4。

```
key=DEADBEEF000000000000000000000000
```

echo命令输出的字符串hello通过管道传给openssl作为输入的明文字符串被加密，-ne 选项控制echo输出的字符串最后不自动添加换行符。openssl的enc子命令用于加密，这里，我们通过-rc4选项指定了RC4算法，并指定了密钥是DEADBEEF，openssl会自动将密钥长度补齐到128个比特位。-e是控制加密的选项参数，-nosalt表示不给密码加盐（后面，我们会详细讨论加盐密码）。-out将加密后的密文输出到指定的文件中。

接下来，我们来看一下加密后的密文到底长得啥样。

```
root@localhost:~# cat hello.rc4 | xxd
00000000: 7d5c e5fc 3b                                }\..;
```

我们将hello.rc4密文文件的内容读出来，并用xxd显示十六进制格式的密文字节，可以看到密文中有5个字节，依次对应明文hello的5个字符。

最后，我们继续使用openssl命令对密文文件进行解密，并将解密后的结果输出到屏幕终端上。

```
root@localhost:~# openssl enc -rc4 -K "DEADBEEF" -d -nosalt -p -in hello.rc4
key=DEADBEEF0000000000000000000000000000
hello
```

在命令行中，我们使用-d选项来表明解密操作，和加密操作一样，继续指定了RC4密码算法和密钥。并用-in选项从文件hello.rc4中读入密文，最终还原出明文字符串hello。

3.2.4 针对RC4的破解

RC4算法是对称的加密算法，加密和解密的步骤都是众所周知的、固定不变的。唯一的保密性来自于种子密钥。理论上，这个种子密钥只有通讯双方知道，但是如果第三方从某种途径获得了这个种子密钥，那么第三方可以毫不费力地用RC4来解密他截获到的密文了。

理论上来说，RC4算法是很难被破解的。RC4中用到的种子密钥是长度在1到256之间可变的无类型字符序列，每个字节有256种可能的取值，那么RC4算法的种子密钥的可能性就是 $256 + 256^2 + 256^3 + \dots + 256^{256} \approx 256^{257}$ 种可能性，量级在 10^{600} 以上。因此，针对RC4算法的暴力破解几乎是不可能的。

研究人员早前发现可以利用RC4中的统计偏差，导致可对加密信息中的一些伪随机字节能进行猜测。在2013年，科学家利用这个漏洞设计了一次攻击实验：他们在2000小时内猜出一个基础身份认证cookie中包含的字符。后来技术改进后，研究人员只需约75小时猜解就能得到94%的准确率。

另外，新型针对WPA-TKIP网络的猜解攻击大概只需要花上1个小时的时间，攻击者可以任意注入、解密数据包。这项技术不仅可以解密cookie和WiFi数据包，其他高速传输的加密数据流也有可能被解密。技术是通过向加密负载中注入数据，如每个认证 cookie或者WiFi数据包中的标准头部。攻击者会通过组合所有可能的值，通过使用统计偏差找出最有可能的组合。研究人员表示，现在RC4加密已经不安全了，建议完全停止使用。在2015年由RFC 7465禁止在所有版本的TLS中使用。

3.3 DES加密算法

1973年，美国国家标准局（现在的美国国家标准技术研究所NIST, National Institute of Standards and Technology）开始研究除国防部外的其它部门的计算机系统的数据加密标准，于1973年5月和1974年8月先后两次向公众发出了征求加密算法的公告。加密算法要达到的目的主要为以下四点：

- 提供高质量的数据保护，防止数据未经授权的泄露和未被察觉的修改；
- 具有相当高的复杂性，使得破译的开销超过可能获得的利益，同时又要便于理解和掌握；
- 密码体制的安全性应该不依赖于算法的保密，其安全性仅以加密密钥的保密为基础；
- 实现经济，运行有效，并且适用于多种完全不同的应用。



1977年1月，美国政府采纳IBM公司设计的方案作为非机密数据的正式数据加密标准 (DES, Data Encryption Standard)，DES被确定为美国联邦信息处理标准 (FIPS, Federal Information Processing Standard)，随后在国际上广泛流传开来。

DES算法因为包含一些机密设计元素，相对短的密钥长度以及怀疑内含美国国家安全局 (NSA, National Security Agency) 的后门而在开始时有争议，DES因此受到了强烈的学院派式的审查，并以此推动了现代的块密码及其密码分析的发展。DES现在已经不是一种安全的加密方法，1999年1月，distributed.net与电子前哨基金会合作，在22小时15分钟内即公开破解了一个 DES密钥。由于DES的密文可以在短时间内被破译，因此现在不应该再使用它了，在2001年，DES作为一个标准已经被AES所取代。但是，作为现代密码技术的开山之作，我们还是要详细地介绍它的技术细节，其中，部分结构上的设计仍然被用于其他密码算法中。

3.3.1 DES的总体结构

DES是一个分组加密算法，以64位为分组对数据加密，加密和解密用的是同一个算法。它的密钥表面上是64位的，然而只有其中的56位被实际用于加密计算，其余8位可以被用于奇偶校验，并在加密算法中被丢弃。因此，DES的有效密钥长度仅为56位。

DES算法的整体结构如图 3.4所示。在DES加密的过程中 (图的左半部分)，有16个相同的被称为轮 (round) 的处理过程，并在首尾各有一次置换，分别称为初始置换 (IP, Initial Permutation) 与最终置换 (FP, Final Permutation)。首尾两次置换是互逆的，即最终置换可以还原初始置换后的原始数据分组，反之亦然。

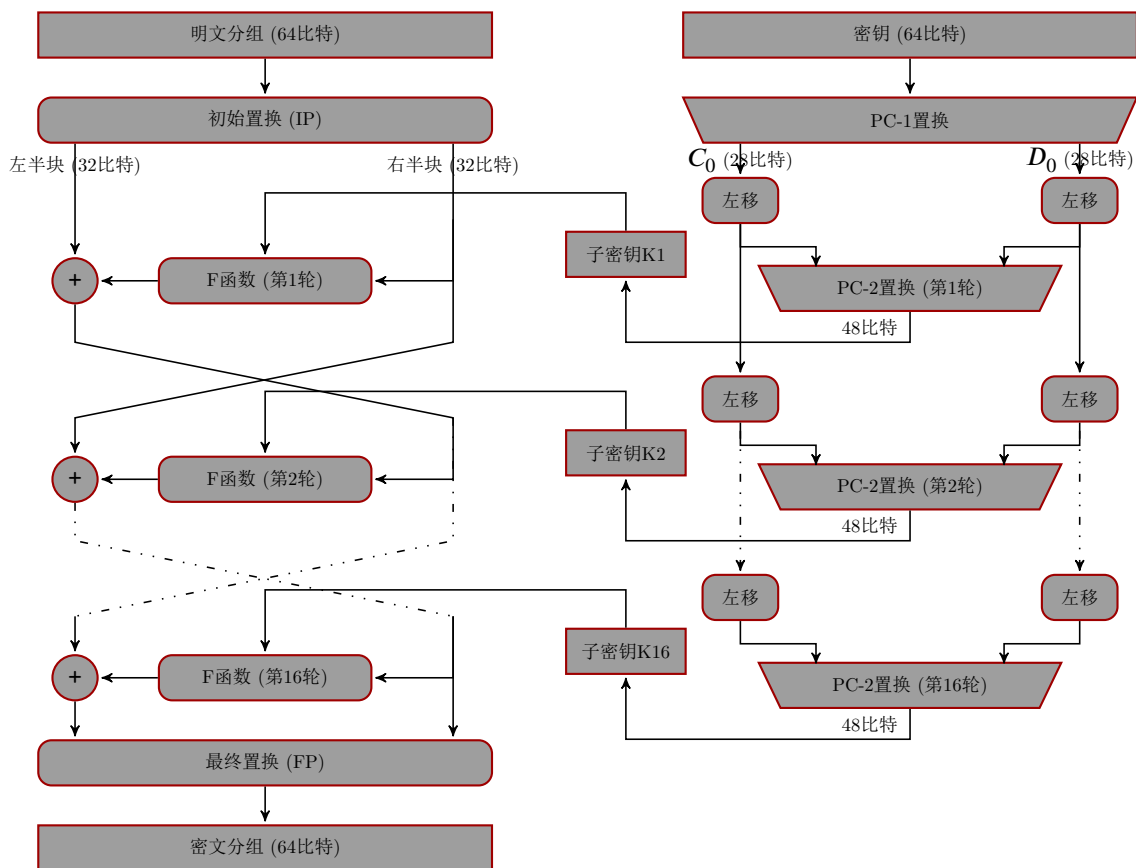


图3.4 DES总体结构图

64位的数据分组在进入第一轮处理之前，被分成两个32位的半块分别处理。随后，每轮处理完后，左右两个半块交叉作为下一轮处理的输入，这种多轮反复交叉的方式被称为Feistel网



络，是由Horst Feistel设计发明的。这一结构不仅被用于DES，在其他很多密码算法中也有应用。Feistel网络中每轮处理的F函数对数据分组的右半块与该轮处理的子密钥进行运算。然后，F函数的输出与左半块进行异或操作之后，再与右半块交叉作为下一轮的处理的两个输入半块。这种结构保证了加密和解密过程足够相似——唯一的区别在于在解密时子密钥是以反向的顺序作用的，而剩余部分则完全相同。这样的设计大大简化了算法的实现，尤其是硬件实现，因为不需要区分加密和解密算法。

Feistel网络中每轮处理所需要的子密钥由密钥调度算法(3.3.4)根据密钥生成(图3.4的右半部分)。下面，对DES每个处理的细节一一进行详细的介绍。

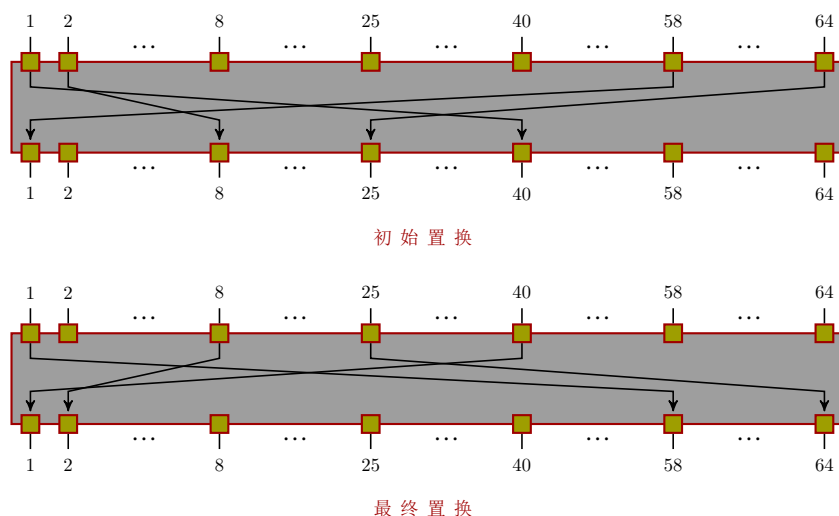


图3.5 初始置换和最终置换

3.3.2 初始置换和最终置换

64位明文分组经过初始置换，生成重新排列之后的64位数据分组。如图3.5所示，这个初始置换对明文分组的64个比特重新排列。排列后的数据分组分成左右两个各32位的半块，输入到Feistel网络中经过16轮F函数的运算，最后由最终置换再次重新排列，输出64位密文分组。由图3.5可以看出，初始置换和最终置换是互逆的，这就满足了加密和解密对称性的必要条件。

3.3.3 Feistel网络

从图3.4所示的DES的总体结构图中，我们了解到明文分组经过初始置换之后，分成两个32位的半块进入Feistel网络。Feistel网络是由16轮F函数串联组成的计算过程。在Feistel网络中，每一轮都需要使用一个不同的子密钥，我们会随后的3.3.4密钥调度这一小节再详细介绍子密钥的生成算法。这里，我们要注意的是，每一轮子密钥的长度均为48位，它和32位的右半块作为F函数的输入，经过计算处理输出长32位的比特序列再和32位的左半块进行异或计算。该异或计算得到的32位比特序列，在进入下一轮处理之前和右半块进行对调。

总结一下，一轮的具体计算步骤如下：

1. 将输入数据分组分成左右两个均为32位的半块。
2. 右半块不经任何处理直接作为F函数的输入。
3. F函数根据右半块和该轮子密钥，计算出长32位的比特序列。
4. 将上一步得到的比特序列和左半块进行异或计算，其结果作为下一轮处理的右半块。
5. 该轮右半块不经任何处理直接发送到下一轮，作为下一轮处理的左半块。



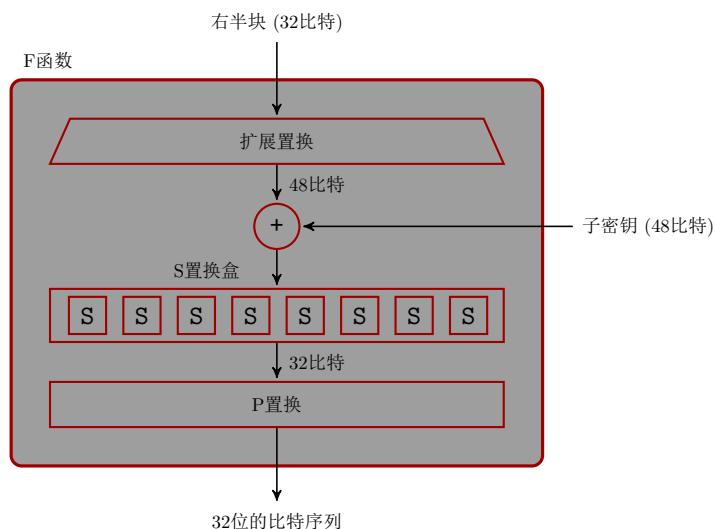


图3.6 F函数计算过程图

Feistel网络一轮计算中最核心的部分是F函数，包含了多个装置：扩展置换、S盒置换和P置换。其计算过程可以用图 3.6来描述，具体如下：

1. 右半块32位的比特序列，经过扩展置换扩展为48位的比特序列。
2. 右半块被扩展后的48位比特序列和子密钥进行异或计算，得到另一个48位的比特序列。
3. 在上一步异或计算得到的48位比特序列，经过S置换盒后，输出32位的比特序列。
4. 最后，P置换把32位的比特序列再次打乱，输出F函数的处理结果。

下面，我们来详细介绍F函数中每个装置的具体操作方法。

(1) 扩展置换

扩展置换将32位的半块扩展到48位。如图 3.7所示，32位的半块在置换前被分成 8个4位的小分组，置换后输出8个6位的小分组。输出的每个6位的小分组中有4位直接对应输入小分组，首末2位则分别来自于前后两个相邻输入小分组中紧邻的首末位。

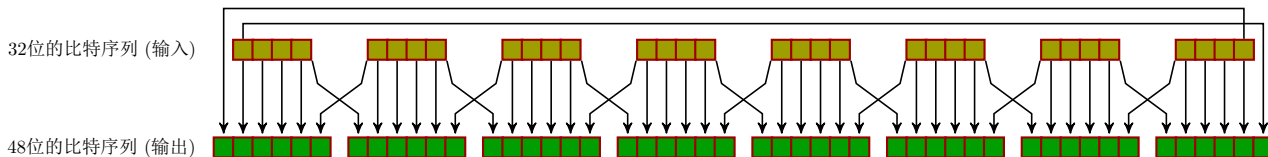


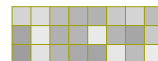
图3.7 扩展置换原理图

(2) S置换

扩展置换后输出的48位比特序列和子密钥经过异或计算之后，由S置换转换回32位的比特序列。S置换操作前，先将48位的比特序列分成8个6位的小分组。S置换操作包含8个S盒，每个S盒分别负责一个小分组的转换。S盒以查找表的方式提供非线性的变换将它的6个输入位变成 4个输出位。

图 3.8解释了将6位比特序列置换成4位比特序列的操作原理。S盒中内置了一个4行16列的替换表，表中元素均为0至15的数（4个比特）。6位比特序列的首尾两个比特组成行索引值，中间4个比特组成列索引值，用于定位替换表中的元素，作为输出的比特序列。S盒提供了DES的核心安全性——如果没有S盒，密码会是线性的，很容易破解。

(3) P置换



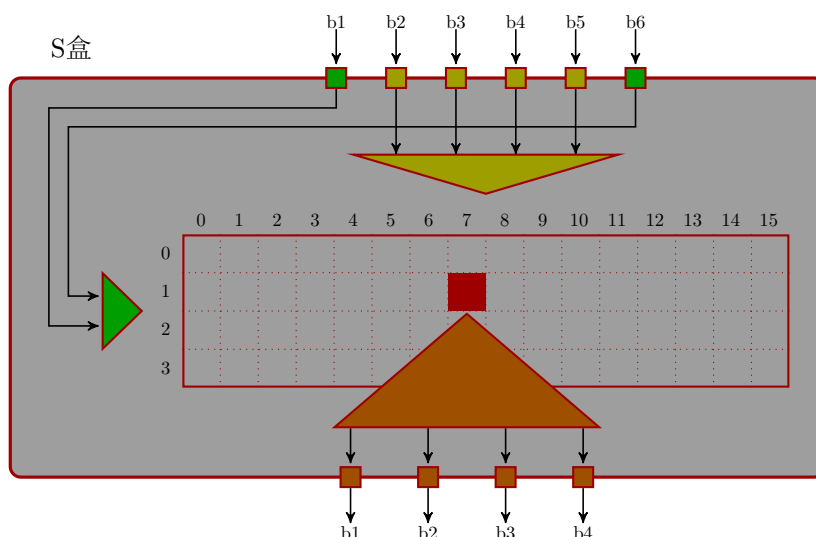


图3.8 S盒将6位的比特序列转换为4位的比特序列

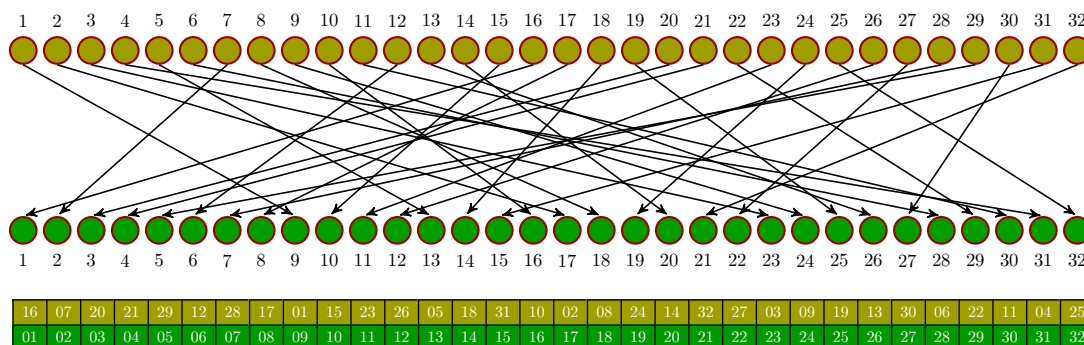


图3.9 P置换原理图

P置换本质上是一个替换表，将S盒置换后输出的32位比特序列按照图 3.9重新排列为另一个 32位的比特序列。P置换后的32位比特序列就是F函数的输出。

至此，我们已经完成了对F函数的详细探讨。尽管F函数稍显复杂，包含多次置换和非线性变换等操作，我们化繁为简，对每个操作进行了详细的解释，以帮助大家更容易理解这些细节。最后，我们来回顾和总结一下Feistel网络的特点。

首先，Feistel网络的轮数不会对加解密产生影响，轮数可以任意增加，无论运行多少轮的计算，都不会发生无法解密的情况。

其次，每轮F函数的设计也不会对加解密产生影响。F函数可以不用被设计成可逆函数，即就算用F函数的输出不能计算出其输入，也不会出现无法解密的问题。所以，F函数可以被设计得任意复杂。

另外，加密和解密可以用完全相同的结构来实现，这也是Feistel网络的一个重要特征。在Feistel网络的一轮中，右半分组实际上没有做任何处理，这样就迫使加密算法不得不使用多轮迭代处理来提高密文的保密程度，这看起来非常浪费时间，但却保证了可解密性。因为完全没有被处理的右半分组，是解密过程中所必需的信息。由于加密和解密可以用完全相同的结构来实现，因此用于实现DES算法的硬件设备的设计也非常容易。

综上所述，无论多少轮数、采用什么F函数，Feistel网络都可以用相同的结构实现加密和解密，且加密的结果必定能正确解密。



3.3.4 密钥调度

密钥调度为DES加密算法生成Feistel网络16轮加密处理中每轮处理所需要的子密钥，其过程如图 3.4的右半部所示。DES密钥调度主要包含2种置换： $PC-1$ 置换和 $PC-2$ 置换。

(1) $PC-1$ 置换

图 3.10解释了 $PC-1$ 置换的原理，图中小圆圈内的数字表示密钥比特序列的位置。初始密钥的每第8个比特在 $PC-1$ 置换中被舍弃，随后重新排列生成56位的比特序列。因此，DES密钥实际的有效长度为56位。 $PC-1$ 置换完成后得到的56位比特序列分成两个28位的半密钥块 C_0 和 D_0 。在剩下的16轮子密钥生成过程中，两个半密钥块均先被循环左移1或2位（第1、2、9、16轮左移1位，其他轮左移2位）后，由 $PC-2$ 置换输出该轮的48位子密钥。

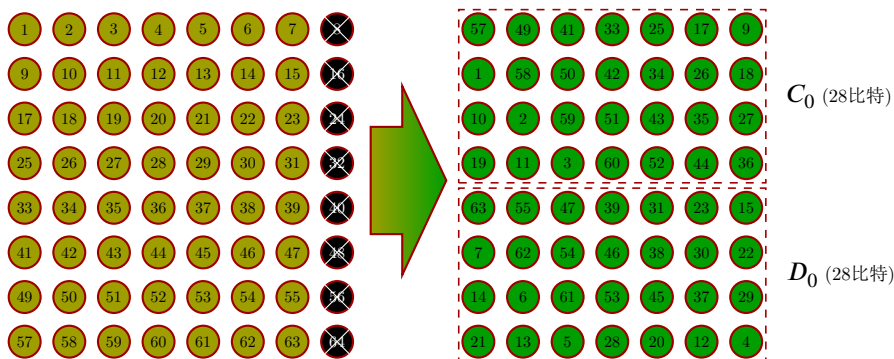


图3.10 $PC-1$ 置换

(2) $PC-2$ 置换

如图 3.11所示， $PC-2$ 置换对两个半密钥块 C_i 和 D_i 分别处理，将每个半密钥块的28位比特序列经挑选后重新排列成24位的比特序列，最后合并成48位子密钥 K_i 。

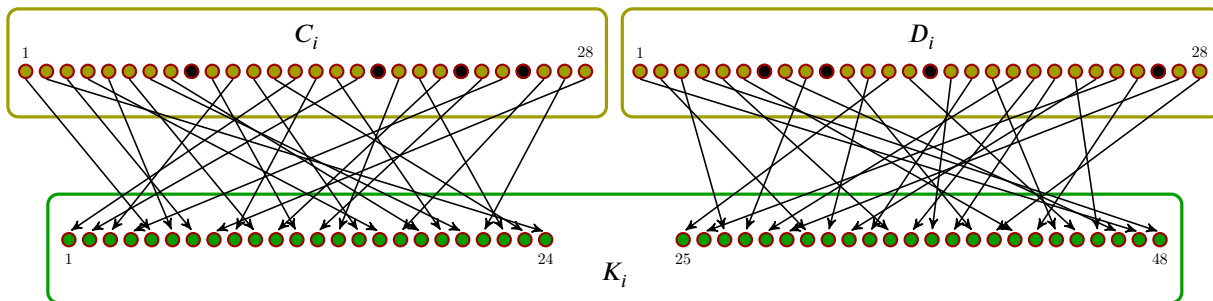


图3.11 $PC-2$ 置换

3.4 3DES加密算法

由于计算机运算能力不断增强，DES密码的密钥长度（56位）变得容易被暴力破解。3DES (Triple DES, 三重DES) 为了增强DES密码的强度，将DES重复三次来增加DES的密钥长度以避免被暴力破解。因此，3DES并不是一种全新的分组密码算法。

3.4.1 3DES的总体结构

3DES加密算法的总体结构如图 3.12所示。在加密和解密过程中，都需要经过3次DES处理。3DES使用“密钥包”，其包含3个DES密钥， K_1 ， K_2 和 K_3 ，均为56位（除去奇偶校验位）。



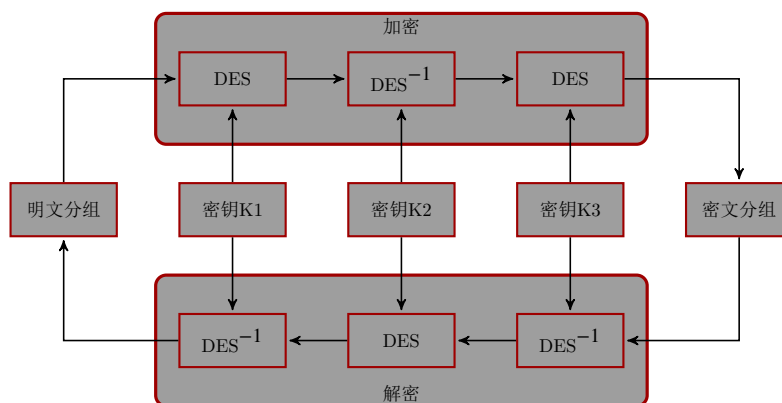


图3.12 3DES密码算法总体结构

在加密过程中，明文分组经过3次DES处理，即 $DES \rightarrow DES^{-1} \rightarrow DES$ ，这里，我们用 DES^{-1} 表示DES的解密。所以，在3DES加密过程中，实际的运算是： DES 加密 \rightarrow DES 解密 \rightarrow DES 加密。这看起来有点不可思议，但实际上却是有意为之。这个设计是IBM提出来的，目的是为了3DES能够向下兼容DES。当3DES的所有密钥 K_1 、 K_2 、 K_3 都相同时，3DES也就等同于普通的DES了。这是因为前两次DES处理（ DES 加密 \rightarrow DES 解密）完成后的结果就是最初的明文分组。

用 E 表示DES加密函数， D 表示DES解密函数，那么3DES的加密算法可以表示为：

$$ciphertext = E_{K_3}(D_{K_2}(E_{K_1}(plaintext)))$$

也就是说，使用密钥 K_1 进行DES加密，再用密钥 K_2 进行DES“解密”，最后以密钥 K_3 进行DES加密。

而解密算法则和加密过程正好相反，是以密钥 K_3 、密钥 K_2 、密钥 K_1 的顺序执行的，形式上可以表示为：

$$plaintext = D_{K_1}(E_{K_2}(D_{K_3}(ciphertext)))$$

即以密钥 K_3 解密，以密钥 K_2 “加密”，最后以密钥 K_1 解密。无论是加密还是解密，中间一步都是前后两步的逆。

3.4.2 密钥选项

3DES标准定义了三种密钥选项：

- 密钥选项1：三个密钥是独立的，即 $K_1 \neq K_2 \neq K_3$ 。
- 密钥选项2： K_1 和 K_2 是独立的，而 $K_3 = K_1$ 。
- 密钥选项3：三个密钥均相等，即 $K_1 = K_2 = K_3$ 。

密钥选项1的强度最高，拥有 $3 \times 56 = 168$ 个独立的密钥位。这种3DES密码算法称为DES-EDE3，EDE是加密 (Encryption) – 解密 (Decryption) – 加密 (Encryption) 3个单词首字母的缩写。

密钥选项2的安全性稍低，拥有 $2 \times 56 = 112$ 个独立的密钥位。该选项比简单的应用DES两次的强度较高，即使用 K_1 和 K_2 ，因为它可以防御中途相遇攻击 (Meet-in-the-Middle Attack)。这种3DES密码算法称为DES-EDE2。

密钥选项3等同于普通DES，只有56个密钥位。这个选项提供了与DES的兼容性，因为第1和第2次DES操作相互抵消了。



3.4.3 3DES密码现状

银行机构目前还在普遍使用3DES，并持续开发和宣传基于3DES的标准，例如EMV。EMV三个字母分别代表Europay、MasterCard与Visa，是制定该标准最初的三家公司。EMV是国际金融业界对于智慧支付卡与可使用晶片卡的POS终端机及自动柜员机（ATM）等所制定的标准。EMV智慧卡（也称为IC卡）的信息储存在集成电路中而非过去的磁条里，但大部分EMV卡背也有可以向下兼容的磁条。芯片可以和插入式读卡器交换数据，非接触式智能卡还可以使用射频识别（RFID）技术在一定距离范围内交换数据。符合EMV标准的支付卡叫做芯片卡。

此外，Microsoft OneNote和Microsoft Outlook 2007使用3DES密码保护用户数据。

尽管3DES还有在使用，但是其两个独立密钥的变种DES-EDE2已经在2015年正式退出。在经过一系列安全分析和针对3DES的实例攻击演示之后，2017年11月，NIST建议应当限制3DES用于对数量在 2^{20} 个以上的64位分组（8MB）的数据进行加密，因此，TLS、IPSec和大文件加密不应该再使用3DES。目前，NIST正在推动2030年彻底废除3DES的使用，但是，随着密码研究揭露的越来越多的漏洞，3DES可能会被加速退出历史舞台。

3.5 AES加密算法

AES的全称是Advanced Encryption Standard，意思是高级加密标准。它的出现主要是为了取代DES加密算法的，我们已经知道DES算法的有效密钥长度是56位，因此算法的理论安全强度是 2^{56} 。但二十世纪中后期计算机飞速发展，元器件制造工艺的进步使得计算机的处理能力越来越强，虽然出现了3DES的加密方法，但由于它的加密时间是DES算法的3倍多，64位的分组大小相对较小，所以还是不能满足对信息安全的要求。于是1997年1月2号，美国国家标准技术研究所（NIST）宣布希望征集高级加密标准，用以取代DES。在密码标准征集中，所有AES候选提交方案都必须满足以下标准：

- 分组大小为128位的分组密码。
- 必须支持三种密码标准：128位、192位和256位。
- 比提交的其他算法更安全。
- 在软件和硬件实现上都很高效。

AES也得到了全世界很多密码工作者的响应，先后有很多人提交了自己设计的算法。有5个候选算法进入最后一轮：Rijndael，Serpent，Twofish，RC6和MARS。最终经过安全性分析、软硬件性能评估等严格的步骤，Rijndael算法获胜。

3.5.1 Rijndael算法

Rijndael算法是由比利时学者Joan Daemen和Vincent Rijmen所提出的，算法的名字就由两位作者的名字组合而成。Rijndael的优势在于集安全性、性能、效率、可实现性及灵活性于一体。不同于DES，Rijndael使用的是代换—置换网络（SPN，Substitution-Permutation Network），而非Feistel结构。

严格地说，AES和Rijndael加密算法并不完全一样（虽然在实际应用中两者可以互换）。Rijndael算法支持更大范围的分组和密钥长度，介于128-256之间所有32的倍数均可，最小支持128位，最大256位，共25种组合。而AES标准支持的分组大小固定为128位，密钥长度有3种选择：128位、192位及256位，分别为AES-128、AES-192和AES-256。

3.5.2 Rijndael的加密

不同于DES加密算法的Feistel网络结构，Rijndael算法是基于代换—置换网络的迭代算法。但是，和DES加密算法类似，Rijndael加密过程中，明文数据也需要经过多个轮次的转换后方能生成密文，每个轮次的转换操作由轮函数定义。同时，Rijndael也包括一个密钥调度算法来根据种子密钥（用户密钥）编排生成多组轮密钥。



根据Rinjdael算法的定义，加密轮数会针对不同的分组及不同的密钥长度选择不同的数值。以word（字长）为单位，1 word等于 4个字节，用 N_b 表示分组的长度， N_k 表示密钥的长度，那么， $N_b = 4$ 就是 $4 \times 4 \times 8 = 128$ 比特，以此类推。Rinjdael给出的加密轮数 N_r 的选择如表 3.1所示。

N_r	$N_b = 4$ (AES)	$N_b = 6$	$N_b = 8$
$N_k = 4$	10 (AES-128)	12	14
$N_k = 6$	12 (AES-192)	12	14
$N_k = 8$	14 (AES-256)	14	14

表3.1 Rinjdael加密轮数 N_r

由于AES标准中分组大小固定为128位（ $N_b = 4$ ），以128位密钥长度（ $N_k = 4$ ）为例，即AES-128算法，加密的总体结构如图 3.13所示，AES-128需要10轮加密，图中左半部分给出了AES-128加密的过程，右半部分为轮密钥的生成过程。加密过程中，第0轮相当于预处理，不计算在轮数内。除最后一轮之外，第1至9轮，每轮的加密轮函数会根据轮密钥对数据依次进行如下 4种运算：

- SubBytes（字节代换）
- ShiftRows（行移位）
- MixColumns（列混淆）
- AddRoundKey（轮密钥加）

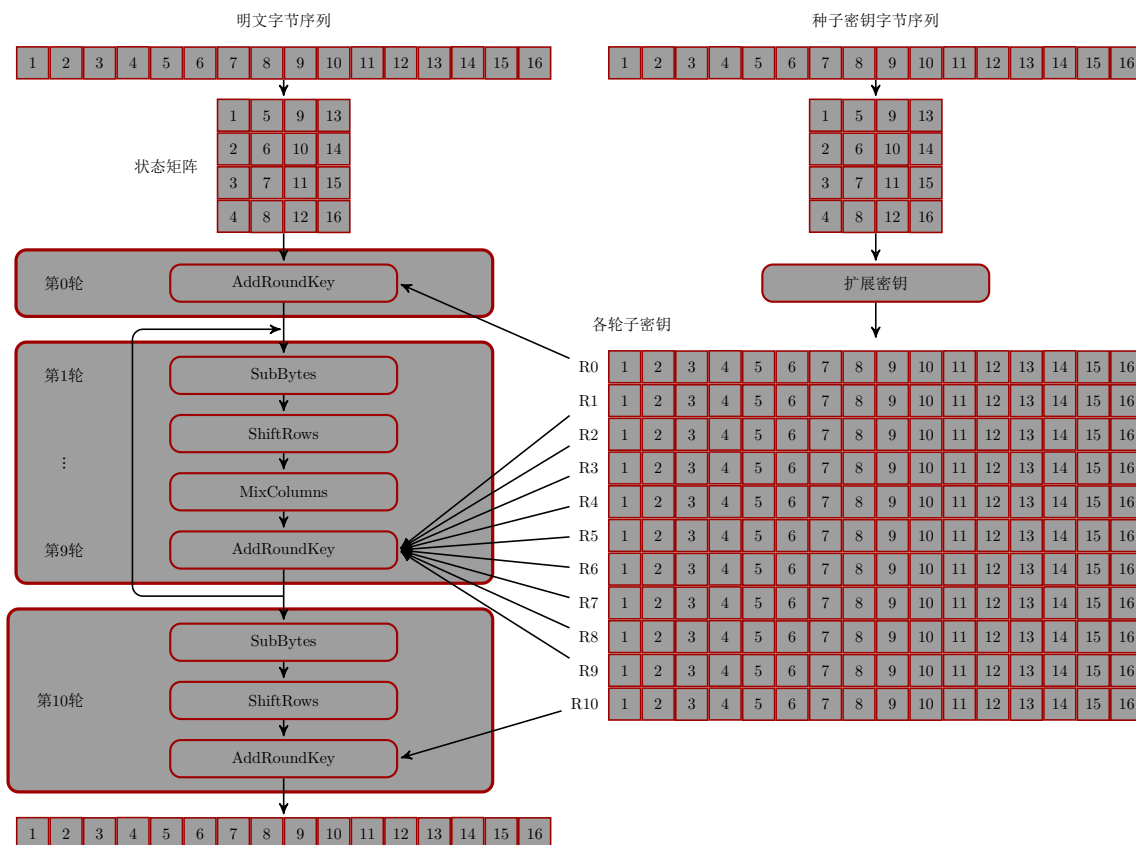


图3.13 AES算法总体结构

图 3.13中小方格中的数字代表该字节在分组中的位置。这16字节的明文分组以列为主序排列为一个 4×4 的状态矩阵，矩阵中的每个元素就是一个字节。由明文分组生成的状态矩阵称为初始状态。



初始状态矩阵经过第0轮的 **AddRoundKey** 运算后的状态矩阵，进一步经过10轮迭代处理生成密文分组。前一轮处理完后的输出状态作为下一轮处理的输入状态，第10轮加密完后的输出状态以列为主序排列为密文分组。

以下，对Rijndael算法中使用的4种运算进行详细的介绍。

AddRoundKey（轮密钥加）

在每次的加密循环中，都会由种子密钥产生一把轮密钥（通过Rijndael密钥调度算法产生）。在**AddRoundKey**运算中，轮密钥与状态矩阵中的每个字节元素依次相“加”，这里的加法运算实际上是逻辑运算中的异或计算，如图 3.14 所示。

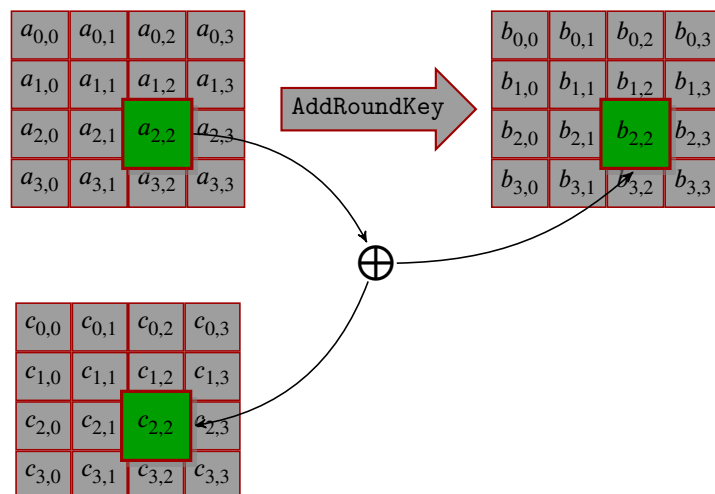


图3.14 AddRoundKey 运算

SubByte（字节代换）

在**SubBytes**步骤中，状态矩阵中的各个字节通过一个8比特的S盒进行替换，如图 3.15所示。字节代换是可逆的非线性变换，也是AES运算组中唯一的非线性变换。S盒是事先设计好的替换查询表。其设计不是随意的，而是根据设计原则严格计算求得，不然无法保证算法的安全性。

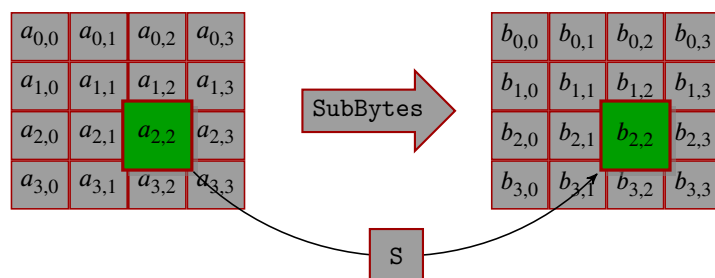


图3.15 SubBytes 运算

ShiftRows（行移位）

ShiftRows描述了作用在矩阵的行上的移位操作，如图 3.16。在此运算中，每一行都向左循环位移某个偏移量。在AES-128中，第1行维持不变，第2行里面的每个元素都向左循环移动一格。同理，第3行及第4行向左循环位移的偏移量就分别是2格和3格。



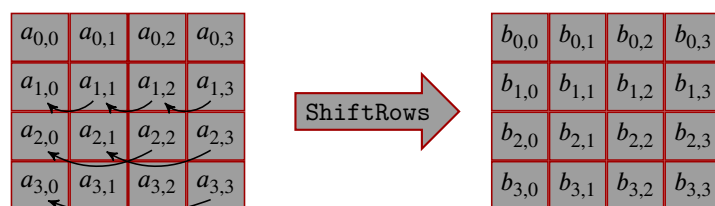


图3.16 ShiftRows运算

MixColumns (列混淆)

列混淆是Rijndael算法中最复杂的一步，其理论基础为近世代数的域论，实质是在有限域（亦称伽罗瓦域，Galois Field） $GF(2^8)$ 上的多项式乘法运算。伽罗瓦是法国的数学家，群论的创立者。用群论彻底解决了根式求解代数方程的问题，而且由此发展了一整套关于群和域的理论。在解释列混淆运算之前，我们有必要先简单了解一下域的概念和域中元素的四则运算。

一组元素的集合，以及在集合上的四则运算，构成一个域。其中加法和乘法必须满足交换、结合和分配的规律，这看起来跟我们平时算术中的乘法和加法运算非常类似。但是，域的一个非常重要的性质是加法和乘法具有封闭性，即加法和乘法的结果仍然是域中的元素。

伽罗瓦的重大发现在于有限域，也就是仅含有限多个元素的域。所以当我们说伽罗瓦域的时候，就是指有限域。 $GF(2^w)$ 表示含有 2^w 个元素的有限域。

$GF(2^8)$ 由一组从0x00到0xff的256个值组成，加上加法和乘法运算。 $GF(2^8)$ 加法就是异或（XOR）操作。然而， $GF(2^8)$ 的乘法有点复杂。AES的MixColumns运算及其逆运算需要知道怎样只用七个常量0x01、0x02、0x03、0x09、0x0b、0x0d和0x0e来相乘。所以这里不全面介绍 $GF(2^8)$ 的乘法，而只是针对这七种特殊情况进行说明。

首先，在 $GF(2^8)$ 中用0x01的乘法是特殊的；它相当于普通算术中用1做乘法并且结果也同样——任何值乘0x01等于其自身。

现在让我们看看用0x02做乘法。和加法的情况相同，虽然理论是深奥的，但最终结果十分简单。只要被乘的值小于0x80，这时乘法的结果就是该值左移1比特位。如果被乘的值大于或等于0x80，这时乘法的结果就是左移1比特位再用值0x1b异或。它防止了“域溢出”并保持乘法的乘积在范围以内。

一旦你在 $GF(2^8)$ 中用0x02建立了加法和乘法，你就可以用任何常量去定义乘法。用0x03做乘法时，你可以将0x03分解为2的幂之和。为了用0x03乘以任意字节b，因为 $0x03 = 0x02 \oplus 0x01$ ，因此：

$$\begin{aligned} b \otimes 0x03 &= b \otimes (0x02 \oplus 0x01) \\ &= (b \otimes 0x02) \oplus (b \otimes 0x01) \end{aligned}$$

同理，用0x0d去乘以任意字节b可以这样做：

$$\begin{aligned} b \otimes 0x0d &= b \otimes (0x08 \oplus 0x04 \oplus 0x01) \\ &= (b \otimes 0x08) \oplus (b \otimes 0x04) \oplus (b \otimes 0x01) \\ &= (b \otimes 0x02 \otimes 0x02 \otimes 0x02) \oplus (b \otimes 0x02 \otimes 0x02) \oplus (b \otimes 0x01) \end{aligned}$$

MixColumns运算的其它乘法遵循大体相同的模式，例如：



$$\begin{aligned}
b \otimes 0x09 &= b \otimes (0x08 \oplus 0x01) \\
&= (b \otimes 0x02 \otimes 0x02 \otimes 0x02) \oplus (b \otimes 0x01) \\
b \otimes 0x0b &= b \otimes (0x08 \oplus 0x02 \oplus 0x01) \\
&= (b \otimes 0x02 \otimes 0x02 \otimes 0x02) \oplus (b \otimes 0x02) \oplus (b \otimes 0x01) \\
b \otimes 0x0e &= b \otimes (0x08 \oplus 0x04 \oplus 0x02) \\
&= (b \otimes 0x02 \otimes 0x02 \otimes 0x02) \oplus (b \otimes 0x02 \otimes 0x02) \oplus (b \otimes 0x02)
\end{aligned}$$

基于以上关于有限域 $GF(2^8)$ 上的加法和乘法运算，图 3.17说明了MixColumns的原理，用矩阵乘法公式可以表示为：

$$\begin{bmatrix} b_{0,j} \\ b_{1,j} \\ b_{2,j} \\ b_{3,j} \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} a_{0,j} \\ a_{1,j} \\ a_{2,j} \\ a_{3,j} \end{bmatrix} \quad 0 \leq j \leq 3$$

由于矩阵乘法将一列中的各个字节利用有限域 $GF(2^8)$ 上的乘法运算和加法运算组合起来，输出状态矩阵相应列的每个字节都受来自输入列四个字节的影响。同样，每个输入列单个字节都会对输出列四个字节造成影响。因此，ShiftRows和MixColumns两步骤为这个密码系统提供了扩散性。

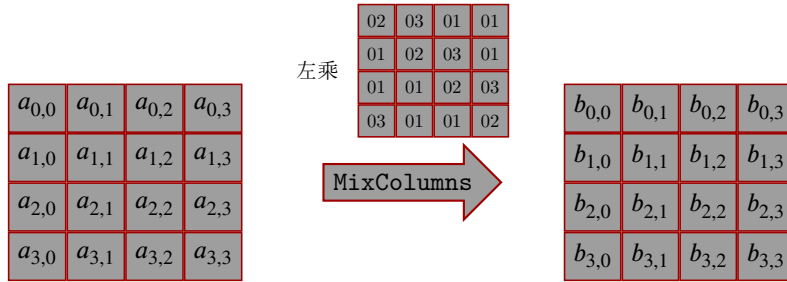


图3.17 MixColumns运算

3.5.3 Rijndael密钥扩展算法

密钥扩展算法是Rijndael的密钥编排实现算法，其目的是根据种子密钥（用户密钥）生成多组轮密钥。轮密钥用于每轮（包括第0轮）AddRoundKey运算中和状态矩阵进行按字节顺序的异或计算。我们继续用 N_b 、 N_k 、 N_r 来表示分组和密钥长度（以word为单位），以及加密轮数（参见表 3.1）。Rijndael需要扩展出 $N_r + 1$ 个子密钥参与每轮的AddRoundKey运算。由于每轮子密钥的长度和分组长度 N_b 一致，所以，需要扩展的所有子密钥的总长度 $N_w = N_b \times (N_r + 1)$ 个字。因为AES算法固定了分组长度为128位，即4个word，因此，AES算法中 $N_b = 4$ 。那么，AES加密算法需要根据种子密钥扩展的子密钥总长度 N_w 如表 3.2所示。

	N_b	N_k	N_r	N_w
AES-128	4	4	10	44
AES-192	4	6	12	52
AES-256	4	8	14	64

表3.2 AES扩展子密钥的word数 N_w

将每轮子密钥以word为单位排列成一个序列，用 w_i ($0 \leq i \leq N_w - 1$)表示子密钥序列中的第 i 个元素，那么，AES算法的密钥扩展就是计算每个 w_i 的值。 w_i 的计算可以用如下数学公式描述：



$$w_i = \begin{cases} [k_{4i} \ k_{4i+1} \ k_{4i+2} \ k_{4i+3}], & \text{if } 0 \leq i < N_k; \\ w_{i-N_k} \oplus G(w_{i-1}), & \text{if } i \geq N_k \text{ and } i \bmod N_k = 0; \\ w_{i-N_k} \oplus \text{SubWord}(w_{i-1}), & \text{if } i \geq N_k, N_k > 6, \text{ and } i \bmod N_k = 4; \\ w_{i-N_k} \oplus w_{i-1}, & \text{Otherwise.} \end{cases}$$

公式略显复杂，但是如果我们用图 3.18来解释AES-128的密钥扩展过程，则该计算就一目了然了。

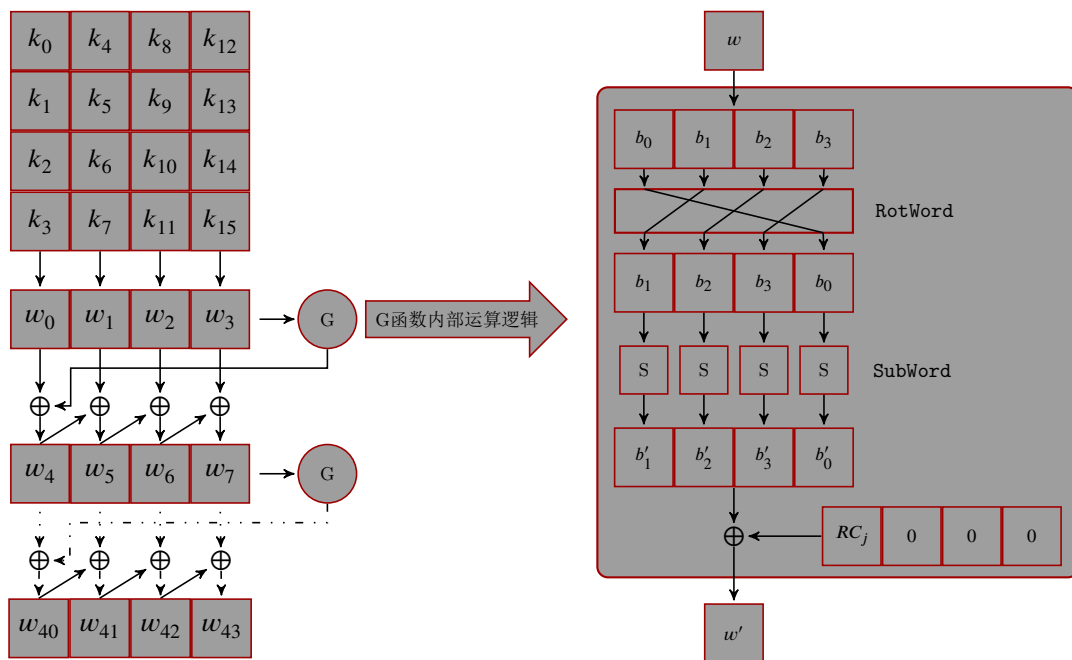


图3.18 AES-128密钥扩展原理图

首先，从公式中的第1个条件可知，前 N_k 个 w 元素由种子密钥初始化而来。对于AES-128， $N_k = 4$ ，将种子密钥每4个字节依序赋值给 w_0 、 w_1 、 w_2 和 w_3 。将每 N_k 个 w 元素排成一行，根据公式中的第2个条件，每行的第一个 w 元素由上一行的最后一个 w 元素经过G函数转换后与上一行的第一个 w 元素进行异或计算而来。当 $N_k > 6$ 时，即AES-128和AES-192，根据第4个条件，每行的其他 w 元素均由该行前一个 w 元素和该列上一行的 w 元素直接异或计算而来。第3个条件则适用于AES-256，也就是说，在AES-256中，每行的第5个 w 元素则由该行的前一个元素经过SubWord替换后与该列上一行的 w 元素异或计算而来。

在AES子密钥生成过程中，每行的最后一个 w 元素都要经过G函数转换后再和该行第一个 w 元素“相加”计算出下一行的第一个 w 元素。图 3.18的右半部分描述了G函数的计算过程，主要包括3个运算，分别为：RotWord、SubWord和一次异或计算。RotWord将 w 元素中的4个字节进行循环置换（以字节为单位循环右移）。SubWord则是根据加密过程中SubByte运算中的S盒对每个字节进行替换操作。而最后一步的异或计算则稍微繁琐，其复杂之处在于轮常量 RC_j ($1 \leq j \leq N_r$)的计算。 RC_j 由定义在有限域 $GF(2^8)$ 上的 $0x02^{j-1}$ 的值计算而来，根据前面讨论有限域 $GF(2^8)$ 上的乘法运算的规则，我们不难得出每轮的轮常量 RC_j 值如下表所示。

j	1	2	3	4	5	6	7	8	9	10
RC_j	0x01	0x02	0x04	0x08	0x10	0x20	0x40	0x80	0x1b	0x36

表3.3 AES密钥扩展G函数中轮常量 RC_j 的值

3.5.4 Rijndael的解密

上一节，我们已经对Rijndael的总体加密过程以及加密过程中所涉及的各种运算逐个进行了详细的说明。下面，我们来了解 Rijndael的解密。Rijndael的解密过程就是将图 3.13所示的加密过程颠倒过来。加密轮函数中的四种运算对应的逆运算分别为：InvSubBytes（图 3.19）、InvShiftRows（图 3.20）、InvMixColumns（图 3.21）和AddRoundKey。

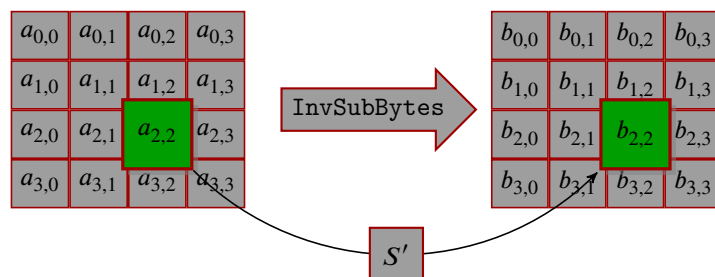


图3.19 InvSubBytes运算

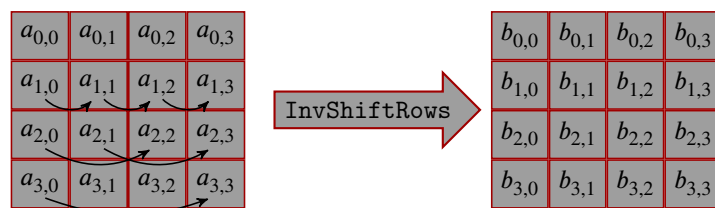


图3.20 InvShiftRows运算

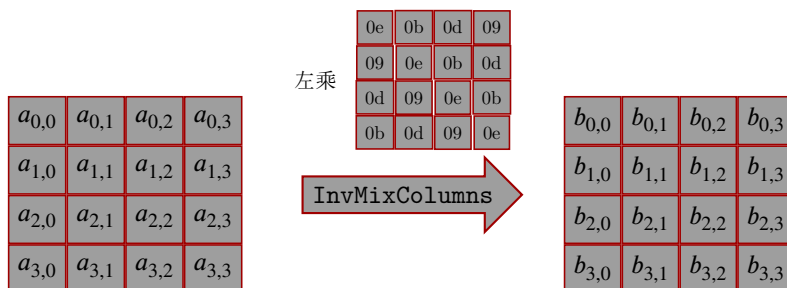


图3.21 InvMixColumns运算

在AddRoundKey运算中使用了异或计算，由于异或计算本身具有自反性，即 $a \oplus b \oplus b = a$ ，因此AddRoundKey的逆运算即是其自身。

InvSubBytes是SubBytes的逆运算。在InvSubBytes运算中用到的 S' 盒恰好是 SubBytes中S盒的反向字符替换表。

InvShiftRows是ShiftRows的逆运算。在InvShiftRows运算中作用在状态矩阵上的移位操作往反方向上移动相应的格数便实现了ShiftRows的逆运算。

InvMixColumns是MixColumns的逆运算。在有限域 $GF(2^8)$ 上，由于

$$\begin{bmatrix} 0e & 0b & 0d & 09 \\ 09 & 0e & 0b & 0d \\ 0d & 09 & 0e & 0b \\ 0b & 0d & 09 & 0e \end{bmatrix} \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} = \begin{bmatrix} 01 & 00 & 00 & 00 \\ 00 & 01 & 00 & 00 \\ 00 & 00 & 01 & 00 \\ 00 & 00 & 00 & 01 \end{bmatrix}$$



说明，这两个矩阵在有限域 $GF(2^8)$ 上互逆，因此，左乘逆矩阵便可实现MixColumns的逆运算。

3.5.5 Rijndael的安全性

AES是目前被广泛采用的对称加密算法。AES加密算法（使用128，192，和256位密钥的版本）的安全性，在设计结构及密钥的长度上都已到达保护机密信息的安全标准。最高机密信息的传递，则至少需要192或256位的密钥长度。

在密码学的意义上，只要存在一个方法，比穷举法还要更有效率，就能被视为一种“破解”。从应用的角度来看，这种程度的破解依然太不切实际，但理论上的突破有时仍可提供对漏洞模式的深入了解。

由于密钥长度每增加1位，密钥空间将增加2倍，并且如果密钥的每个可能值的机会均相等，则意味着平均蛮力密钥搜索时间将增加一倍。这意味着蛮力搜索的工作量随着密钥长度的增加而呈指数增长。密钥长度本身并不意味着针对攻击的安全性，因为已经发现具有很长密钥的密码容易受到攻击。

针对AES最大的争议则在于AES有一个相当井然有序的代数框架的数学结构。尽管相关的代数攻击尚未出现，但有许多学者认为，把安全性创建于未经透彻研究过的结构上是有风险的。不过，这也只是一种假设而已。到目前为止，尚无已知的实际攻击方法。不过，根据斯诺登的文件，美国国家安全局（NSA）正在研究基于 τ 统计信息的加密攻击是否可能有助于破坏AES。

3.6 Chacha20加密算法

ChaCha20是一种流式加密算法，它源自于Salsa20加密算法。2004年，ECRYPT启动了eSTREAM流密码计划的研究项目。丹尼尔·J·伯恩斯坦（Daniel J. Bernstein）在2005年设计出了Salsa20流式加密算法，被eSTREAM选中，成为最终胜出的7个算法之一。其特点是结构简单、易于实现，很安全（支持128比特或256比特的密钥），速度很快，也适合在嵌入式系统上实现。此外，Salsa20不受专利的约束。随后，在2008年，伯恩斯坦在Salsa基础上做了一些改动，发布了变种算法ChaCha，其目的是在不牺牲性能的前提下，增加每一轮的扩散。

ChaCha20由Google公司率先在Android移动平台中的Chrome浏览器和Google网站间的HTTPS通讯时代替RC4使用。ChaCha20通常和Poly1305消息认证码（Message Authentication Code, MAC）算法一起配合使用，合称为ChaCha20-Poly1305算法。在密码学上，这种把加密算法和MAC算法的组合称为关联数据的认证加密（Authenticated Encryption with Associated Data, AEAD）。比如，AES-GCM就是另外一种AEAD算法。

AEAD是一种同时具备保密性，完整性和可认证性的加密形式。AEAD产生的原因很简单，单纯的对称加密算法，其解密步骤是无法确认密钥是否正确的。也就是说，加密后的数据可以用任何密钥执行解密运算，得到一组疑似原始数据，而不确认密钥是否是正确的时候，也就也不知道解密出来的原始数据是否正确。因此，需要在单纯的加密算法之上，加上一层验证手段，来确认解密步骤是否正确。这里，我们先对AEAD有个初步的认知，后面，我们在介绍到MAC算法的时候，会重点讨论AEAD。本小节，我们主要介绍ChaCha20加密算法。

3.6.1 ARX运算和操作函数

Salsa20的核心是一个建立在基于Add-Rotate-XOR（ARX）操作的Hash函数之上，将一个64字节的输入序列转换为另一个64字节的输出序列。这个Hash函数产生的伪随机字节流和明文字节流进行异或运算生成密文字节流。这使得Salsa20具有了不同寻常的优势，它可以在恒定时间内定位到输出流中的任何位置。

前面我们都是采用自顶向下的方法介绍各个加密算法的，即先介绍算法的总体结构，然后深入到算法运算的细节。现在，我们使用自底向上的方法来介绍Salsa20加密算法，先介绍其中采用的各种运算，再介绍salsa20是如何将这些运算组合起来的。

ARX运算



在Salsa20算法内部，采用的3个基本运算是：32位模加、循环移位和异或计算。异或运算自然不必多说了，普遍应用在密码算法中。我们用符号 \boxplus 表示模加运算， w_1 和 w_2 两个word的32位模加运算则可以表示为：

$$w_1 \boxplus w_2 = (w_1 + w_2) \bmod 2^{32}$$

Salsa中的循环移位为循环左移操作，我们用 R^m 表示循环左移 m 位的运算， $R^m(w)$ 表示将32位的word w 循环左移 m 位。

QuarterRound函数

用 $w = (w_0, w_1, w_2, w_3)$ 表示由4个word组成的序列，那么QuarterRound(w)输出另外一个由4个word组成的序列 $z = (z_0, z_1, z_2, z_3) = \text{QuarterRound}(w)$ 。Salsa算法中定义的QuarterRound函数如下：

$$\begin{aligned} z_1 &= w_1 \oplus (R^7(w_0 \boxplus w_3)) \\ z_2 &= w_2 \oplus (R^9(z_1 \boxplus w_0)) \\ z_3 &= w_3 \oplus (R^{13}(z_2 \boxplus w_1)) \\ z_0 &= w_0 \oplus (R^{18}(z_3 \boxplus w_2)) \end{aligned}$$

随后，在2008年，伯恩斯坦发布的ChaCha密码家族对Salsa的QuarterRound函数进行了修改，其目的是增加每一轮的扩散以实现相同或稍微提升的性能。ChaCha算法中的QuarterRound函数如下：

$$\begin{aligned} w_0 &= w_0 \boxplus w_1; & w_3 &= w_3 \oplus w_0; & w_3 &= R^{16}(w_3); \\ w_2 &= w_2 \boxplus w_3; & w_1 &= w_1 \oplus w_2; & w_2 &= R^{12}(w_2); \\ w_0 &= w_0 \boxplus w_1; & w_3 &= w_3 \oplus w_0; & w_3 &= R^8(w_3); \\ w_2 &= w_2 \boxplus w_3; & w_1 &= w_1 \oplus w_2; & w_2 &= R^7(w_2); \\ (z_0, z_1, z_2, z_3) &= (w_0, w_1, w_2, w_3) \end{aligned}$$

图 3.22展示了Salsa和ChaCha算法的QuarterRound运算过程。经过比较，不难发现ChaCha算法中的QuarterRound运算对每个word更新2次，增加了扩散性。

ChaCha状态

ChaCha状态是一个以行为主序的 4×4 的矩阵（如图 3.23所示），矩阵中的每个元素是一个word。QuarterRound运算每次从状态矩阵上选取4个元素进行转换运算。为简化表示，我们用QuarterRound(a, b, c, d)来表示选取状态矩阵中元素下标分别为 a 、 b 、 c 和 d 的word进行QuarterRound运算，即QuarterRound(1, 5, 9, 13)表示对word序列(w_1, w_5, w_9, w_{13})进行QuarterRound运算。

如果选取的4个元素属于状态矩阵的某一列，就被称为ColumnRound；如果选取的4个元素排列在状态矩阵的某一条对角线上，则被称之为DiagnoalRound。例如，QuarterRound(1, 5, 9, 13)就是ColumnRound，QuarterRound(0, 5, 10, 15)和QuarterRound(2, 7, 8, 13)都属于DiagnoalRound。

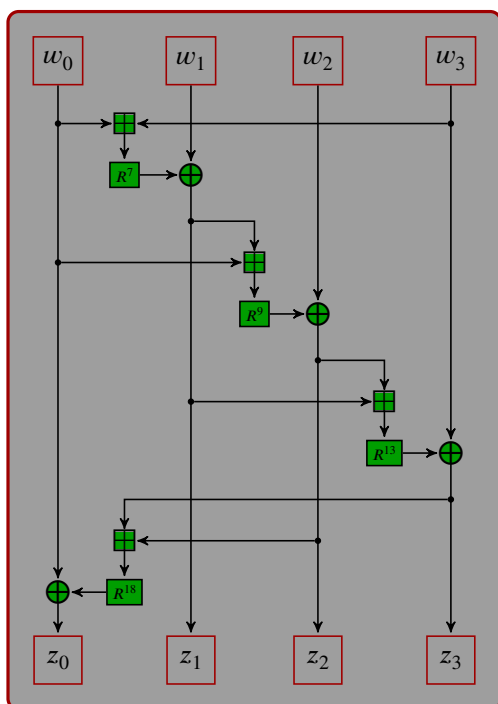
LittleEndian函数

在计算机内存中的字节排序分为两类：Big-Endian（大端）和Little-Endian（小端）。标准的Big-Endian和Little-Endian的定义如下：

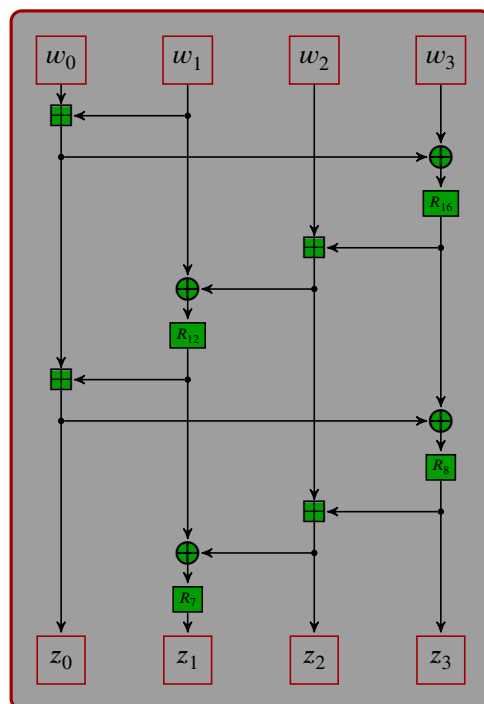
- Little-Endian就是低位字节排放在内存的低地址端，高位字节排放在内存的高地址端。
- Big-Endian就是高位字节排放在内存的低地址端，低位字节排放在内存的高地址端。

根据上述定义，对于字节序列 $b = (b_0, b_1, b_2, b_3)$ ，经过LittleEndian函数运算之后的word值可以表示为：





Salsa的QuarterRound运算



ChaCha的QuarterRound运算

图3.22 QuarterRound的运算过程

w_0	w_1	w_2	w_3
w_4	w_5	w_6	w_7
w_8	w_9	w_{10}	w_{11}
w_{12}	w_{13}	w_{14}	w_{15}

图3.23 ChaCha状态矩阵

$$\text{LittleEndian}(b) = 2^{24}b_3 + 2^{16}b_2 + 2^8b_1 + b_0$$

3.6.2 ChaCha20分组函数

ChaCha20分组函数通过对如下的输入数据形成状态矩阵，并对状态矩阵进行多轮QuarterRound运算后，计算出密钥流的分组。

- 256位的密钥：32个字节被划分成由8个32位的LittleEndian整数排列组成的word序列；
- 96位的随机常量（nonce）：12个字节被划分成3个32位的LittleEndian整数排列组成的word序列；
- 32位的分组计数器：4个字节被LittleEndian函数转换成一个word。

首先，根据图 3.24所示的过程初始化一个 4×4 的状态矩阵，共16个word、64个字节。以上三个输入数据的长度加起来共48个字节，缺少的16个字节由常量字符串“expand 32-byte k”中每个字母对应的ASCII编码后的字节序列补上。



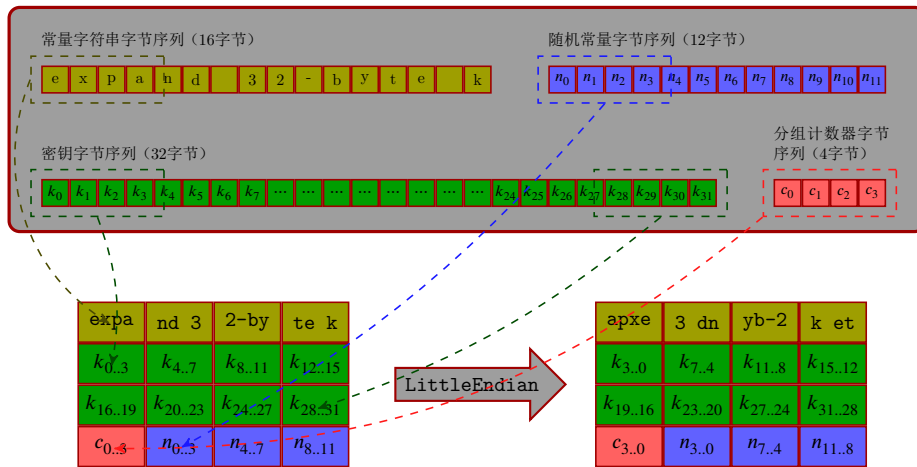


图3.24 ChaCha状态矩阵初始化过程

由图 3.24可知，ChaCha状态的初始化过程如下：

- 初始状态的前4个word，由常量字符串“expand 32-byte k”对应的ASCII编码后的字节序列，经 LittleEndian运算而来。
- 256位的密钥对应的字节序列经LittleEndian运算后，得到初始状态矩阵中的第5至12个字节，即 $w_4 \sim w_{11}$ 。
- 4字节的分组计数器的LittleEndian值填充到第13个字节。
- 96位的随机常量nonce所对应的字节序列经LittleEndian转换后，填充到初始状态矩阵的最后3个字节。

由此可知，一旦密钥和随机常量确定后，初始状态就取决于分组计数器的值了。在早期的ChaCha版本中，还存在64位分组计数器和 64位随机常量组合的方案，但在RFC8439中，已经舍弃了这种组合，确定了32位分组计数器的方案。这就意味着当密钥和随机常量确定后，最多有 2^{32} 个状态（分组），把每个64字节的状态当做512位的比特流，那么最多可以形成 $2^{32} \times 512 \div 8 \text{ bytes} = 256\text{GB}$ 不重复的密钥流，这在很多应用场合下，已经够用了。对于需要加密超过 256GB数据的应用，比如文件或磁盘加密等，推荐使用早期的64位分组计数器和64位随机常量组合的方案。

在确定了初始状态矩阵之后，ChaCha20算法执行20轮由ColumnRound和DiagnoalRound交替进行的运算。每轮运算执行4次QuarterRound计算。在轮计算为ColumnRound时，执行如下4次QuarterRound 计算，确保状态矩阵中的每个word都替换掉。

QuarterRound(0,4,8,12)
 QuarterRound(1,5,9,13)
 QuarterRound(2,6,10,14)
 QuarterRound(3,7,11,15)

在轮计算为DiagnoalRound时，执行的QuarterRound计算如下。4次QuarterRound计算同样确保状态矩阵中的每个word都被替换。

QuarterRound(0,5,10,15)
 QuarterRound(1,6,11,12)
 QuarterRound(2,7,8,13)
 QuarterRound(3,4,9,14)



在20轮计算以后，也就是经过对上述8个QuarterRound运算进行10次迭代之后输出的状态矩阵再和初始状态矩阵按元素进行32位模加计算所得到的结果状态矩阵，进一步使用LittleEndian序列化化成64字节的分组。整个过程参见示意图 3.25。

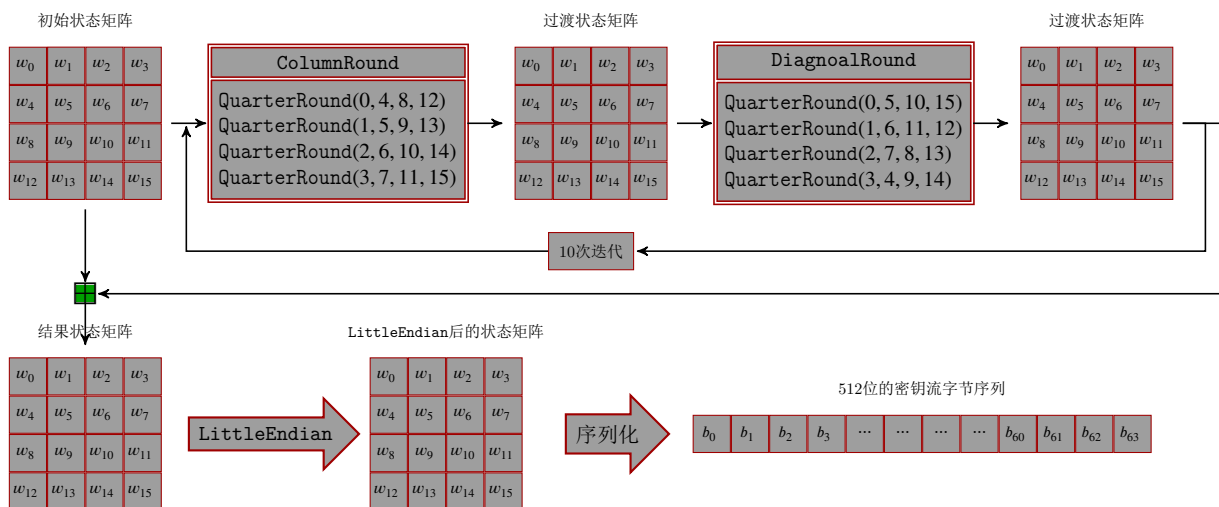


图3.25 ChaCha分组函数计算过程图

3.6.3 ChaCha20加密过程

ChaCha20是一种流式加密算法，通过连续调用分组函数来生成连续的密钥流。每次调用分组函数的种子密钥和随机常量nonce值都是固定不变的，只对分组计数器进行递增。这样，每次调用分组函数生成了不同的512比特的密钥流片段。分组函数的调用次数取决于明文数据的长度。明文数据字节序列和密钥流字节序列按位置顺序依次进行异或计算便得到了加密后的密文数据。由于异或计算具有自反性，解密和加密过程相同，将密文数据字节序列和分组函数生成的密钥流进行异或后，解密出明文数据。

3.6.4 ChaCha20的应用

在Google采用了ChaCha20-Poly1305来替代RC4完成互联网的安全通信之后，ChaCha20-Poly1305算法也以chacha20-poly1305@openssh.com成为OpenSSH中的一个新密码套件。

ChaCha20-Poly1305在IKE和IPsec中的使用已在RFC 7634中标准化。在RFC 7905中，ChaCha20-Poly1305已经被加入TLS扩展标准。

若CPU不支持AES指令集，ChaCha20可提供比AES更好的性能。因此，ChaCha20通常被广泛应用于基于ARM CPU的移动设备上。

3.7 本章小结

本章，我们详细介绍了几种对称加密算法。总体上，对称加密算法分为流密码和分组密码。

