经典密码学杂谈

卓泽滨

摘要：
本文首先介绍加密法的发展历程，后将结合案例介绍两种经典的加密法以及其对应的密码分析方法，最后将总结以上说谈，并简单的介绍其余的经典加密法。
关键词：加密法；密码分析；加密法发展史；Vigenere 加密法；keyword 加密法；

概述

“芦花滩上有扁舟， 俊杰黄昏独自游， 义到尽头原是命， 反躬逃难必无忧。”
这是一首出自《水浒传》中吴用留下的藏头反诗；电视剧《裂变》中汉奸“蝙蝠”也曾使用数字对应书本页面和文字的方法传递消息给日寇；电影《暗算》中更是提到了黄依依解决多种密文的具体情节；甚至连动画片《名侦探柯南》也出现了 skytale 加密的细节。事实上，密文不仅存在于荧幕中，而且深入到生活的方方面面，例如用于存储互联网消息的 cookie、以及互联网安全中常提到的数字签名、在银行等网站上填写个人信息时，都会用一定的手段将明文加密成密文传输到在远处的服务器中，可以说，在互联网的世界里，只要有比特流动，就一定会有加密的存在。为此，各大高校还设立了专门的学科，如密码学、密码分析学、密码史等。不得不说，密码的发展更数学密切相关、大多数的密码学家都兼任数学家的身份，而密码学，这一学科在战争时代更是快速地发展。以下将会介绍密码学的发展史、以及一些经典密码学的典型加密方法和其对应的解密方法的介绍、文章最后会简单地提及现代密码学的一些实现手段。

加密手段与解密手段

常见的加密手段有三种：编码法、加密法和夹带加密法；

编码法

编码法类似于计算机的编译[1]过程，在计算机编译过程中，以 ASCII 码为例，计算机会将字母 A 编译成 0x41，B 编译为 0x42 以此类推。我们之所以会很清楚 0x43 对应字母 C，

[1] 将源代码转化为目标文件的过程（二进制），此处用 16 进制表示 2 进制

是因为我们知道 ASCII 码表，假设我们对 ASCII 码表一无所知，那么 0x44 究竟代表什么我们就不得而知了。而编码法就是这样一种加密手段，即将明文按照一本写好的编码簿翻译成密文，在受到密文后，解读者需要一本一模一样的编码簿才能解读出具体的明文。

加密法

加密法则不同，一种常用的加密法是将字母转化为数字，例如用一下方法：

从右图中我们看出 a 对应的是 11，并且令底频率字母 j 等于 i，这样，每一个字母都有与之对应的数字，而我们可以对数字按照一定的算法进行运算即可得出相应的密文数字，此时，解读者要解读出这些密文，则需要知道形成这一密文的运算法则[2]，并更具这一运算法则进行逆运算即可得出明文。

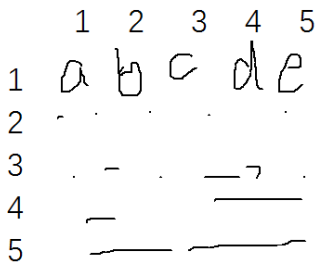


图 1

夹带加密法[3]

夹带加密法是指在不将明文处理成难以解读的密文的情况下，想办法将明文隐藏起来。例如小时候使用隐形笔写字就属于这种加密方法。又如吴用的藏头诗也属于这一类加密手段。历史资料记载，希腊人在传递消息的时候，曾经把奴隶剃成光头，并将文字刻在奴隶的头上，待到奴隶的头发长出来时，将奴隶送给对方，以便让对方获得信息。这也属于夹带法的范畴。

解密手段

常见的解密手段可分为三类，首先是密文攻击，在只有密文的情况下解读出明文叫密文攻击；
另外还有明文攻击，即在知道某段明文一定存在的情况下[4]，将密文解读成明文的手段；第三种是选取明文攻击，即事先泄露对方感兴趣的话题，使对方针对这一话题的信息加密成密文，这样在截获密文后，就可以清晰的知道密文中一定包含这一话题有关的内容。例如二战中美国暴露给日本中途岛的信息，从而得出 AC 对应中途岛这一结论，就是利用了选取明文攻击。

经典密码学的发展历史

[2] 即秘钥
[3] 也叫夹带法
[4] 不一定知道所对应的密文

加密法的出现可以追溯到 200BC 甚至更远，那个时候，希腊作家 polybius 发明了一种名为 polybius cipher 的加密法，这是一种单码加密[5]，其加密手段大致如图 1 所示，即将字母对应成数字，当然，可以乱序地将字母填充在该矩阵中得到不同的对应模式。早期的密码还有斯巴达克人所发明的一种 skytale 的加密工具如右图所示：

其加密原来是将纸条绕在一根加密棒上，同时在绕在上面的纸条上写下明文，展开后就可以得到乱序的密文了，本质上是一种夹带加密法。

希伯来人也发明了早期的加密法：如 atbah、atbash、albam 等。这些加密法原理上是将字母与某个密文符号一一对应起来的加密手段。后来罗马大帝凯撒在军事上曾用凯撒加密法对明文进行加密，防止密码泄露。凯撒加密法是将明文的字母按标准字母表向后移动若干位，得到一种对应模式；例如 a 对应 d、b 就对应 e、明文 zhuozebinK 对应 ckxrchelq；可以说这些加密法都很容易破解，例如对凯撒加密法 K，我们可以穷举所有的可能，可以得出密文 ckxrchelq 对应的所有字符串，则其中一定存在明文：可以看出，移动三位可以得到真正的明文。于是密码被破解。

1	bjwqbgdkp
2	aivpafcjo
3	zhuozebin
4	ygtnydahm
5	xfsmxczgl
6	werlwbyfk
7	vdqkvaxej

此后的几百年里，出现了一种名单码加密法[6]，其中的代表有关键词加密法，直到 14 世纪、15 世纪仍在欧洲广泛使用。这时候已经出现了对此种加密手段的解密方法。为了弥补这种不足，出现许多新的加密法。如仿射加密法、多文字加密法，这类加密法的解决原理与关键词加密法一样，并且安全性甚至较之更低。单码加密法持续的时间最久，甚至在美国南北战争期间也在使用。

直到 16 世纪中期，出现了多码加密[7]的加密方法。代表有 Vigenere 加密法，此类加密法如昙花一现，虽然安全性高，但在数学爆炸性发展的时代，Vigenere 加密法没过多久就被解决了。派生的多码加密法还有自动秘钥加密法、Nihilst 加密法、圆柱面加密法和回转轮加密法，其中在二战期间，联军使用的加密法就是回转轮加密法。

经典加密法的收官之作是多图加密法，这种加密方法将多个明文字母对应成多个密文，更具安全性。此后则进入现代加密法的时代。

[5] 即一个明文字母对应一个密文字母
[6] 同 5
[7] 即一个明文字母对应多个密文字母

代表性的加密法介绍

关键词加密法

keyword 加密法，要求首先选取一个关键词，去掉关键词内的重复字母，如选取“zhuozebin”后的变为“zhuoebin”，然后将 zhuoebin 放在开头（或任意位置开始），其余的位置按照标准字母表的顺序填入，并与标准字母表对应，如下：

Keyword Substitution

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
z	h	u	o	e	b	i	n	a	c	d	f	g	j	k	l	m	p	q	r	s	t	v	w	x	y

明文：

sunnysaidthatshelovemebutidontknowthatwhetheritistruthorjustatrick[8]

加密后：

qsjjxqzaornzrqnefktegehsraokjrdjkvrnzrvnerneparaqrpsrnkpcsqrzrpaud

解析

因为关键词是任意的，不可能穷举，我们可以采用频率分析的方法，解决此类的问题。因为一一对应的单码加密法保留了用此频率不变的特征，根据这一特征，我们可以统计出字母之间的对应关系。以下统计了各个词的频率，从柱形图中我们可以清晰地看出明显的波峰波谷，根据英语的用词频率，可以估计出以下对应关系

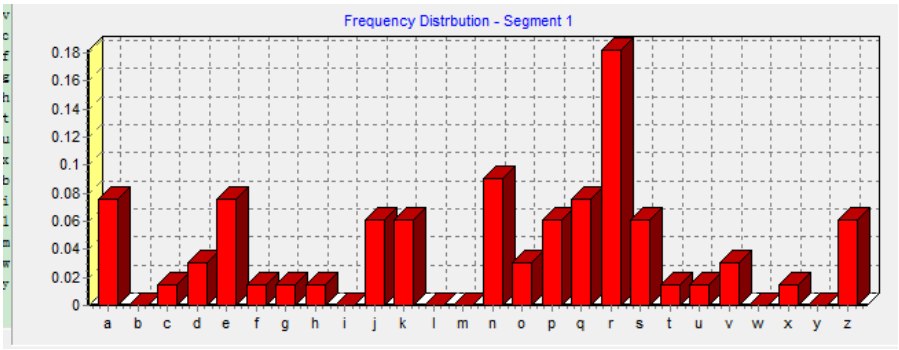
E	A	I	O	U	Y
r	e	z	j	k	k

Frequency Report

Report on segment 1

r	12	0.182
n	6	0.091
a	5	0.076
e	5	0.076
q	5	0.076
j	4	0.061
k	4	0.061
p	4	0.061
s	4	0.061
z	4	0.061
d	2	0.030
o	2	0.030
v	2	0.030
c	1	0.015
f	1	0.015
g	1	0.015
h	1	0.015
t	1	0.015
u	1	0.015
x	1	0.015
b	0	0.000
i	0	0.000
l	0	0.000
m	0	0.000
w	0	0.000
y	0	0.000

[8] 为了方便加密和解密，一般明文中不包括符号和空格



当然，由于截获的密文数量不足，统计未免有不当之处。也可以采用双联字母的方法估计，如字母组合：th、he、er 频率高，可以猜测高频率的组合可能是此三者之一，此外还有三联字母组合的方法确定对应模式。

Vigenere 加密法

这是一种多码加密法，一个明文字母对应多个密文字母，同样选取关键词，如“zhuoebin”，重复关键词并与标准顺序的字母表对应，即可形成一种加密方式。即：

zhuoebinzhuoebinzhuoebinzh

abcdefghyzklmnopqlstuvwxyz

这样明文与秘钥通过查找 Vigenere 表可以得出密文，如 za 为 z，hb 对应 i

这样明文为：

sunnysaidthatshelovemebutidontknowthatwhetheritistruthorjustatrick

密文为：

rbhbxwbqqsouhrlftbulgsayuqqnunymsxbuzaqvdximehacgsvvbunydirxbbehje

解析

IC 法估算关键词的长度

定义 IC 值为：$IC = \frac{\sum_{\alpha=a}^{z} f_\alpha (f_\alpha - 1)}{N(N-1)}$

其中 f α 是密文中 α 字母出现的频率，而 N 密文的长度[9]。

计算出 IC 之后，例如上例所得出的 IC 值为 0.04382，查表可值，keyword 的长度可能是：5,6,7,8；

Kasiski 法确定具体的关键词长度

从密文中看出，bu 一共出现了两次

rbhbxwbqqsouhrlftbulgsayuqqnunymsxbuzaqvdximehacgsvvbunydirxbbehje

根据数学推导，可以知道，重复字符串之间间隔的字符串长度为关键词长度的整数倍，据此我们可以推断出关键词长度为 16/2=8。

得知关键词长度后，将密文分为 8 组，第一组为第 1、 9、 17…组成，以此类推，这样就可以按频率法分别求出各自的对应关系即可。

结尾：

关于经典的密码技术还有很多；最典型的该属回转轮加密法了，也称为 Enigma 加密法。屏幕上的惊悚跌宕的剧情在现实中也是同样存在的。比如《比尔密码》就是属于这类的传奇。传说比尔（Beale）于 1817 年得到了一笔巨大的财富，并将一个盒子交给他的朋友，告诉他 10 年后没回来就打开盒子。结果 10 年后朋友打开时发现里面竟然是写满数据的纸。由此比尔密码花费了许多人毕生的生命。

自 2013 年 6 月棱镜计划破产后，网络安全成为众人关注的问题。如何才能防止自己的个人信息被监听？如何才能避免黑客窃取自己的账号密码？如何防止医院的医疗记录被盗？如何在战争中获取敌人的情报？都应该是值得思考的问题。

经典加密法远不止如此，而基于电子技术的现代加密法更是引人入胜。道高一尺魔高一

[9] 不计入空格和标点符号

丈，有密码就会有密码分析，人类的技术就在于挑战中进化，在数学中升华。

参考文献

[1]〔美〕Richard Spillman. 经典密码学与现代密码学. 清华大学出版社.

[2] Lori Andrews.I Know Who You Are and I Saw What You Did.中国友谊出版社.