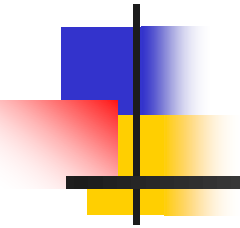

第2章 古典密码破译

古典密码

古典密码是密码学的渊源，这些密码大都比较简单，现在已很少采用了。然而，研究这些密码的原理，对于理解、构造和分析现代密码都是十分有益的。

古典密码

明文字母表和密文字母表相同，为：

$$\boldsymbol{Z}_q=\{0, 1, \ldots, q-1\}。$$

明文是长为L的字母串，以$\boldsymbol{m}$表示：

$$\boldsymbol{m}=(m_0\ m_1,\ldots, m_{L-1}),$$

其中每个$m_i\in Z_q$，$i=0$，1，…，$L-1$。

密文是长为L的字母串，以$\boldsymbol{c}$表示：

$$\boldsymbol{c}=(c_0, c_1, \ldots, c_{L-1}),$$

其中每个$c_i\in Z_q$，$i=0$，1，…，$L-1$。

古典密码

单表代换密码

单表代换密码是字母表到自身的一个可逆映射f,

$$f: \ \boldsymbol{Z}_q \rightarrow \boldsymbol{Z}_q。$$

令明文$\boldsymbol{m}=m_0m_1\ldots$，则相应密文为

$$\boldsymbol{c}=c_0c_1\ldots=f(m_0)f(m_1)\ldots$$

古典密码

1．移位代换密码 (Shift Substitution Cipher)

加密变换：$f(l)=(l+k) \bmod q$，$0 \le l < q$。

其中k为密钥，$0 \le k < q$。

解密变换：$f^{-1}(l)=(l-k) \bmod q$，$0 \le l < q$。

例如：**凯撒**(Caeser)密码是对英文26个字母进行移位代换的密码，其q=26。

三、古典密码

选择密钥*k*=3，则有下述代换表：

abcdefg hijklmn opqrst uvwxyz

DEFGHIJ KLMNOPQ RSTUVW XYZABC

明文：***m*** =Casear cipher is a shift substitution

密文：***c***=FDVHDU FLSKHU LV D VKLIW VXEVWLWXWLRQ

三、古典密码

2. 乘数密码(Multiplicative Cipher)：

加密变换：$f(I)=Ik \bmod q$，$0\le I<q$。

其中k为密钥，$0\le k<q$。显然，仅当$(k, q)=1$（即k与q互素）时，$f(I)$才是可逆变换。

解密变换：$f^{-1}(I)=Ik^{-1} \bmod q$，$0\le I<q$。

古典密码

我们知道，共有$\varphi(q)$个k满足： $0 \le k < q$， $(k, q)=1$。这就是说，乘数密码共有$\varphi(q)$个不同的密钥。

对于$q=26$，

$$\varphi(26)=\varphi(2\times 13)=\varphi(2)\times\varphi(13)=12,$$

即共有12个不同的密钥k=1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23和25。

此时对应的$k^{-1} \bmod q$=1, 9, 21, 15, 3, 19, 7, 23, 11, 5, 17和25。

三、古典密码

3. 仿射密码(Affine cipher)

加密变换： $f(l)=lk_1+k_0 \bmod q$，$0\le l<q$。

其中k_1, $k_2\in \boldsymbol{Z}_q$，$(k_1, q)=1$，以$[k_1, k_0]$表示密钥。当$k_0=0$时就得到乘数密码，当$k_1=1$时就得到移位密码。

$q=26$时可能的密钥数为$26\times\varphi(26)=26\times12=312$个。

古典密码

4. 多项式代换密码(Polynomial Substitute Cipher)

加密方程为：

$f(l)=k_t l^t+k_{t-1} l^{t-1}+\ldots+k_1 l+k_0 \mod q$。

其中，$k_t, \ldots, k_0 \in \boldsymbol{Z}_q$，$l \in \boldsymbol{Z}_q$。

前三种密码都可看作是它的特例。

古典密码

5. 密钥短语密码

选一个英文短语，称其为**密钥字**(Key Word)或**密钥短语**(Key Phrase)，如HAPPY NEW YEAR，去掉重复字母得HAPYNEWR。将它依次写在明文字母表之下，而后再将字母表中未在短语中出现过的字母依次写于此短语之后，就可构造出一个字母代换表，如下所示：

古典密码

$\mathcal{A}$：abcdefg hijklmn opqrst uvwxyz

$\mathcal{A}$'：HAPYNEWRBCDFGIJKLMOQSTUVKZ

这是一种易于记忆而又有多种可能选择的密码。用不同的密钥字就可得到不同的代换表。q=26时将可能有$26! \approx 4 \times 10^{26}$种。其中绝大多数代换都是好的。是一种灵活变化密钥的代换密码。

古典密码

用现代密码学的眼光观察单表代换密码

设单表代换密码用于多次一密的加解密方式，以下对其进行已知明文攻击。Eve截获了一段密文，并获得了该段密文所对应的明文。

一、只要密文中含有q个不同的字母（因此对应的明文中也含有q个不同的字母），则加密变换f被确定。

古典密码

二、对于移位代换密码，只要密文中含有1个字母（对应的明文中也含有1个字母，即1个明密文对），则密钥k被确定。

三、对于乘数密码，只要密文中含有1个字母（对应的明文中也含有1个字母，即1个明密文对），则密钥k被确定。

四、对于仿射密码，只要密文中含有2个不同的字母（对应的明文中也含有2个不同的字母，即2个明密文对），则密钥$[k_1, k_0]$被确定。

古典密码

五、对于多项式代换密码，只要密文中含有min{t+1, q}个不同的字母（对应明文中也含有min{t+1, q}个不同的字母），则密钥[k_t, ..., k_0]被确定。

综上所述，只要密文中含有至多q个不同的字母，单表代换密码体制就被攻破了。

古典密码

多表代换密码

多表代换密码是字母表$\boldsymbol{Z}_q=\{0, 1, \ldots, q-1\}$到自身的$d$个可逆映射$f_0$，$f_1$，…，$f_{d-1}$，在加密时循环排列使用。

令明文$\boldsymbol{m}=m_0m_1\ldots$，则相应密文为

$$\boldsymbol{c}=c_0c_1\ldots=f_0(m_0)f_1(m_1)\ldots f_{d-1}(m_{d-1})f_0(m_d)f_1(m_{d+1})\ldots$$

古典密码

1. 维吉尼亚密码

可逆映射f_0，f_1，…，f_{d-1}都是移位代换密码，分别使用密钥$(k_0, k_1, \ldots, k_{d-1})$。令明文$\boldsymbol{m}=m_0m_1\ldots$，则相应密文为

$$\boldsymbol{c}=c_0c_1\ldots=(m_0+k_0 \bmod q)(m_1+k_1 \bmod q)\ldots$$
$$(m_{d-1}+k_{d-1} \bmod q)$$
$$(m_d+k_0 \bmod q)(m_{d+1}+k_1 \bmod q)\ldots$$

古典密码

对维吉尼亚密码的讨论

设维吉尼亚密码用于多次一密的加解密方式，对其进行已知明文攻击。Eve截获了一段密文，并获得了该段密文所对应的明文。只要密文长度不小于d，密钥$(k_0, k_1, \ldots, k_{d-1})$就被确定。

若d充分大，大到不可能截获长度为d的密文（存储量和时间限制），甚至可以设d“接近无穷大”。此时当然可以抵抗已知明文攻击。

古典密码

（注解：当d“接近无穷大”时，维吉尼亚密码变成了现代密码中的一种，我们称之为流密码或序列密码。）

但问题是，通信伙伴之间怎样简单快速地协商极长的密钥序列$(k_0, k_1, \ldots, k_{d-1})$?

当然不能逐字母地协商密钥。因为，如果攻击者截获长度为d的密文是不可能的（存储量和时间限制），则通信伙伴协商出长度为d的密钥也是不可能的。

古典密码

一种办法是： 取(k_0, k_1, …, k_{d-1})为一本书，此时通信伙伴只需要相互告知该书的书名和版本号，因此使得密钥协商简单快速。

这种办法很容易进行局部破译。

设Eve截获了一段长度为l的密文，并获得了该段密文所对应的明文。 Eve因此也获得了密钥中长度为l的一段(k_0， k_1， …， k_{l-1})。 l与d相比当然是微不足道的。

古典密码

如果该段是名书名句，则Eve只需要找到该名书。

如果该段并不著名，也可以根据文字推断书名及书目类别，并到书店寻找该书。

也可以根据文字的特征，由上下文含义来推测后续密钥。比如$(k_0, k_1, \cdots, k_{l-1})$=information securit，则$k_l$=y；$(k_0, k_1, \cdots, k_{l-1})$=计算机病，则$k_l$=毒；等等。

古典密码

另一种办法是： 取(k_0, k_1, …, k_{d-1})为某个周期序列的一个周期，周期d极大，而用一个长度大约为$ln(d)$的“密钥种子”，采用公开算法来递归生成这个周期序列。

此时通信伙伴只需要相互告知“密钥种子”的值。

这是现代流密码的一般构造，存在着大量有待解决的工程实践问题、学术理论问题。

古典密码

2. 多字母代换密码

多字母代换密码是字母L维向量空间到自身的一个可逆映射f: $\boldsymbol{Z}_q^L \to \boldsymbol{Z}_q^L$; 即

$$f(m_0 m_1 \ldots m_{L-1}) = c_0 c_1 \ldots c_{L-1}。$$

令明文$\boldsymbol{m} = m_0 m_1 \ldots$，则相应密文为

$$\boldsymbol{c} = c_0 c_1 \ldots$$

$$= f(m_0 m_1 \ldots m_{L-1}) f(m_L m_{L+1} \ldots m_{2L-1}) \ldots$$

古典密码

对多字母代换密码的讨论

我们知道，字母L维向量空间$\boldsymbol{Z}_q^L$一共有q^L个向量。换句话说，多字母代换密码f实际上是一个单表代换密码，只不过“字母表”是$\boldsymbol{Z}_q^L$，有q^L个“字母”。这里的一个“字母”就是$\boldsymbol{Z}_q^L$中的一个L维向量。

如果f设计得好，则需要q^L个“密文字母”和其对应的“明文字母”才能确定f。（取$q=2$，$L=128$，则$q^L=2^{128}$是一个极庞大的数字。）

古典密码

因此，设计得好的多字母代换密码f能够极大地扩大已知明文攻击所需要的明文/密文的长度。

但问题是，多字母代换密码f怎样做到设计得好并且简单快速？（达到设计要求的密码是一种现代密码，称之为分组密码。）

显然不能使用真实的代换表，因为一个代换表需要q^L个对应规则：

$$x \to f(x),\ x \in \boldsymbol{Z}_q^L.$$

古典密码

古典的多字母代换密码f或者采用移位运算，或者采用线性运算，或者采用仿射运算，等等。这些孤立的运算都是简单快速的，但是在已知明文攻击之下都是非常不安全的。只需要比较短的密文和对应明文就可以确定密钥。

现代分组密码的设计思想是，f由若干简单运算组合而成。这些简单运算互相屏蔽，使得已知明文攻击很难成功地找出密钥。