密码学实验报告 实验三

2019 年 3 月 19 日

1 仿射密码

1.1 算法原理

仿射 Caesar 密码是对 Caesar 密码的一种推广，定义为：对于每一个明文p，用密文C代替，其中

$$C = E([a,b],p) = (ap+b)\bmod 26$$

容易推导出解密算法为：

$$p = D([a,b],C) = (C-b)*a^{-1}\bmod 26$$

其中a^{-1}可以通过扩展欧几里得算法求得。

同时我们还需要保证算法是单射的，即如果$p \neq q$，则$E(k,p) \neq E(k,q)$，这要求$\gcd(a,26)=1$，所以我们可以选取的密钥$[a,b]$是有限制的。

1.2 算法实现的伪代码

仿射密码加密算法 利用仿射密码加密

输入： 明文x，密钥a,b
输出： 密文y

```
function encrypt(x, a, b)
    if gcd(a, 26) ≠ 1 then
        return 无法加密
    end if
    for i ← 1 to length(x) do
        y[i] ← a * x[i] + b mod 26
    end for
    return y
end function
```

仿射密码解密算法 利用仿射密码解密

输入： 密文y，密钥a,b
输出： 明文x

```
function decrypt(y, a, b)
    if gcd(a, 26) ≠ 1 then
        return 无法加密
```

```
    end if
    for i ← 1 to length(y) do
        x[i] ← (y[i] − b) ∗ a^{−1} mod 26
    end for
    return x
end function
```

1.3 测试样例及运行结果

我们假设密钥$a = 25, b = 3$，尝试对明文

$$cryptography$$

进行加密，得到密文

$$BMFOKPXMDOWF$$

再使用相同的密钥对密文解密，得到相同的明文，证明算法正确。

2 Vigenere 密码

2.1 算法原理

Vigenere 密码是一种著名的多表代替密码，其代替规则集由 26 个 Caesar 密码的代替表组成，其中每一个代替表是对明文字母表移位 0~25 次后得到的代替单表。每个密码由一个密钥字母来表示，这个密钥字母用来代替明文字母 a，故移位 3 次的 Caesar 密码由密钥值 d 来代表。

Vigenere 密码的表述如下：假设明文序列为$P = p_0, p_1, \cdots, p_{n-1}$，密钥由序列$K = k_0, k_1, \cdots, k_{m-1}$构成，其中典型的是$m < n$。密码序列$C = C_0, C_1, \cdots, C_{n-1}$计算如下：

$$C = C_0, C_1, \cdots, C_{n-1} = E(K, P) = E[(k_0, k_1, \cdots, k_{m-1}), (p_0, p_1, \cdots, p_{n-1})]$$
$$= (p_0 + k_0)\text{mod } 26, (p_1 + k_1)\text{mod } 26, \cdots, (p_{m-1} + k_{m-1})\text{mod } 26,$$
$$(p_m + k_0)\text{mod } 26, (p_{m+1}, k_1)\text{mod } 26, \cdots, (p_{2m-1}, k_{m-1})\text{mod } 26, \cdots$$

因此，密钥的第一个字母模 26 加到了明文的第一个字母，接着是第二个字母，以此类推，直到前m个明文处理完毕。对于第二组的m个明文，重复使用密钥字母。继续该过程直到所有的明文序列都被加密完。加密过程的一般方程是

$$C_i = (p_i + k_{i \text{ mod } m})\text{mod } 26$$

类似地，解密过程的一般方程是

$$p_i = (C_i - k_{i \bmod m}) \bmod 26$$

2.2 算法实现的伪代码

Vigenere 密码加密算法 利用 Vigenere 密码加密

输入： 明文x，密钥key
输出： 密文y

```
function encrypt(x, key)
    m ← length(key)
    for i ← 1 to length(x) do
        y[i] ← x[i] + key[i mod m] mod 26
    end for
    return y
end function
```

Vigenere 密码解密算法 利用 Vigenere 密码解密

输入： 密文y，密钥key
输出： 明文x

```
function decrypt(y, key)
    m ← length(key)
    for i ← 1 to length(y) do
        x[i] ← y[i] − key[i mod m] mod 26
    end for
    return x
end function
```

2.3 测试样例及运行结果

我们使用课本上的样例对算法进行测试，假设密钥为$deceptive$，对明文

$$wearediscoveredsaveyourself$$

进行加密，得到密文

$$ZICVTWQNGRZGVTWAVZHCQYGLMGJ$$

再使用原密钥对密文进行解密，得到相同明文，证明算法正确。

3 Vernam 密码

3.1 算法原理

Vernam 密码体制的运算基于二进制数据，该体制可以简明地表述为

$$c_i = p_i \oplus k_i$$

其中p_i是明文第i个二进制位，k_i是密钥第i个二进制位，c_i是密文第i个二进制位，$\oplus$是异或运算符。密文是通过对明文和密钥的逐位异或而成的。根据异或运算的性质，解密过程为

$$p_i = c_i \oplus k_i$$

3.2 算法实现的伪代码

Vernam 密码加密解密算法 利用 Vernam 密码加密解密

输入： 二进制数据x，密钥key
输出： 二进制数据y

```
function vernam(x, key)
    for i ← 1 to length(x) do
        y[i] ← x[i] ⊕ key[i]
    end for
    return y
end function
```

3.3 测试样例及运行结果

由于 Vernam 密码通常用于二进制流数据加密，所以我们假设密钥为二进制数据

$$0010011000$$

明文为

$$1001101101$$

加密后得到相应的密文为

$$1011110101$$

再重新使用密钥对密文解密，输出与明文相同，证明算法正确。

4 单表替换密码的攻击算法

4.1 算法原理

单表代替密码虽然有很大的密钥空间，难以被穷举攻击破解，但是它带有原始字母使用频率的一些统计学特征，我们可以首先把密文中字母的相对频率统计出来，与英文字母的使用频率分布进行比较，如果已知消息足够长的话，便可以

得到密码对照表。但是在密文较短的情况下，可能需要使用二阶或三阶的频率特征来进行推测。

由于单表替换密码的攻击并不一定每次都能给出正确的答案，所以我们需要从可能的答案中挑选最符合的结果，要求算法能够按照可能性大小给出前N个可能的明文。这里我们还需要对**可能性**做一个定义：

根据文献中给出的英文字母频率分布表，英文文本中出现的字母按照频率从高到低排列为：etaoinshrdlcumwfgypbvkjxqz。在每次给当前频率最大的明文字母挑选相应密文字母时，不一定会选择频率最大的那个密文字母，而可能会选取其他的字母，在每次选取时，假设当前频率最高的字母频率为p_m，而挑选得到的字母频率为p_k，我们有$p_k \leq p_m$，在选择该字母后得到的对照表的可能性计算公式如下：

$$P_t = P_{t-1} * [1 - p_k * (p_m - p_k)]$$

该公式的意义在于，给予当前频率最高的密文字母最大的可能性，而挑选其他字母会得到一些惩罚，即降低答案的可能性，同时此次挑选对于整个对照表的影响取决于当前字母的频率。每当完成一个明密文对照表，同时会得到一个相应的可能性，按照可能性大小给出最终的答案即可。

4.2 算法实现的伪代码

单表替换密码攻击算法 返回前N个可能的明文

输入： 密文C，待选的明文数K

输出： 按照可能性大小排序后前N个可能的明文P

```
function statistic(text)
    for i ← 1 to 26 do
        freq[i] ← 0
    end for
    for i ← 1 to length(text) do
        freq[text[i]] ← freq[text[i]] + 1
    end for
    for i ← 1 to 26 do
        freq[i] ← freq[i]/length(text)
    end for
    return freq
end function
function singleTableCracker(C, K)
```

$prob\_stat \leftarrow statistic(C)$
$prob\_table \leftarrow 'etaoinshrdlcumwfgypbvkjxqz'$

$candidates \leftarrow [[\{\}, prob\_stat, 1.0]]$

while $candidates[0][1] \neq \{\}$ **do**
for $i \leftarrow 1$ to $length(candidates)$ **do**
for $j \leftarrow 1$ to $length(candidates[i][1])$ **do**
$newobj \leftarrow candidates[i]$
$newobj[2] \leftarrow 1 - newobj[1][i](newobj[1][0] - newobj[1][i])$

$newobj[0][newobj[i]] \leftarrow prob\_table[length(newobj[0])]$

if $length(candidates) > K$ **then**
break
end if
end for
end for
end while
for $i \leftarrow 1$ to $length(candidates)$ **do**
for $j \leftarrow 1$ to $length(C)$ **do**

$P[j] \leftarrow candidates[i][0][C[j]]$

end for
end for
return P
end function

4.3 测试样例及运行结果

作为测试样例，我们选取了童话《卖火柴的小女孩》文本片段作为待加密的明文，片段内容如下：

her little hands were almost numbed with cold oh a match might afford her a world of comfort if she only dared take a single one out of the bundle draw it against the wall and warm her fingers by it she drew one out rischt how it blazed how it burnt it was a warm bright flame like a candle as she held her hands over it it was a wonderful light it seemed really to the little maiden as though she were sitting before a large iron stove with burnished brass feet and a brass ornament at top the fire burned with such blessed influence it warmed so delightfully the little girl had already stretched out her feet to warm them too butthe small flame went out the stove vanished she had only the remains of the burntout match in her hand

使用自定义的单表代替密码加密后的密文如下：

BYI ARQQAY BXLKJ GYIY XACZJQ LSCUYK GRQB VZAK ZB X CXQVB CRMBQ XOOZIK BYI X GZIAK ZO VZCOZIQ RO JBY ZLAD

KXIYK QXPY X JRLMAY ZLY ZSQ ZO QBY USLKAY KIXG RQ XMXRLJQ QBY GXAA XLK GXIC BYI ORLMYIJ UD RQ JBY KIYG ZLY ZSQ IRJVBQ BZG RQ UAXHYK BZG RQ USILQ RQ GXJ X GXIC UIRMBQ OAXCY ARPY X VXLKAY XJ JBY BYAK BYI BXLKJ ZFYI RQ RQ GXJ X GZLKYIOSA ARMBQ RQ JYYCYK IYXAAD QZ QBY ARQQAY CXRKYL XJ QBZSMB JBY GYIY JRQQRLM UYOZIY X AXIMY RIZL JQZFY GRQB USILRJBYK UIXJJ OYYQ XLK X UIXJJ ZILXCYLQ XQ QZE QBY ORIY USILYK GRQB JSVB UAYJJYK RLOASYLVY RQ GXICYK JZ KYARMBQOSAAD QBY ARQQAY MRIA BXK XAIYXKD JQIYQVBYK ZSQ BYI OYYQ QZ GXIC QBYC QZZ USQQBY JCXAA OAXCY GYLQ ZSQ QBY JQZFY FXLRJBYK JBY BXK ZLAD QBY IYCXRLJ ZO QBY USILQZSQ CXQVB RL BYI BXLK

我们设置$K = 3$，即希望得到前 3 个可能性最大的明文，对密文进行破解得到以下结果：

结果 1（可能性 1.00）

her little hands were almost numged with cold oh a match miyht afford her a world of comfort if she onlk dared tape a sinyle one out of the gundle draw it ayainst the wall and warm her finyers gk it she drew one out rischt how it glazed how it gurnt it was a warm griyht flame lipe a candle as she held her hands ojer it it was a wonderful liyht it seemed reallk to the little maiden as thouyh she were sittiny gefore a larye iron stoje with gurnished grass feet and a grass ornament at tox the fire gurned with such glessed influence it warmed so deliyhtfullk the little yirl had alreadk stretched out her feet to warm them too gutthe small flame went out the stoje janished she had onlk the remains of the gurntout match in her hand

结果 2（可能性 0.73）

her little hands were almost numged with cold oh a match miyht afford her a world of comfort if she onlj dared tape a sinyle one out of the gundle draw it ayainst the wall and warm her finyers gj it she drew one out rischt how it glabed how it gurnt it was a warm griyht flame lipe a candle as she held her hands oxer it it was a wonderful liyht it seemed reallj to the little maiden as thouyh she were sittiny gefore a larye iron stoxe with gurnished grass feet and a grass ornament at toq the fire gurned with such glessed influence it warmed so deliyhtfullj the little yirl had alreadj stretched out her feet to warm them too gutthe small flame went out the stoxe xanished she had onlj the remains of the gurntout match in her hand

结果 3（可能性 0.64）

her little hands were almost numged with cold oh a match miyht afford her a world of comfort if she onlp dared tabe a sinyle one out of the gundle draw it ayainst the wall and warm her finyers gp it she drew one out rischt how it glazed how it gurnt it was a warm griyht flame libe a candle as she held her

hands ojer it it was a wonderful liyht it seemed reallp to the little maiden as thouyh she were sittiny gefore a larye iron stoje with gurnished grass feet and a grass ornament at tox the fire gurned with such glessed influence it warmed so deliyhtfullp the little yirl had alreadp stretched out her feet to warm them too gutthe small flame went out the stoje janished she had onlp the remains of the gurntout match in her hand

可以看出，该单表代替密码的攻击方案有较好的攻击效果，并且可以依照可能性大小给出合理的待选答案，在之后，我们还可以利用二阶或三阶的统计特征，来进一步筛选答案，优化我们对答案可能性的定义。

5 Hill 密码的攻击

5.1 算法原理

Hill 密码属于多表代替密码，利用了线性代数的知识对明文进行加密。该加密算法将m个连续的明文字母替换成m个密文字母，这是由m个线性等式决定的，在等式里每个字母被指定为一个数值，例如$m = 3$，系统可以描述为

$$c_1 = (k_{11}p_1 + k_{21}p_2 + k_{31}p_3) \mod 26$$
$$c_2 = (k_{12}p_1 + k_{22}p_2 + k_{32}p_3) \mod 26$$
$$c_3 = (k_{13}p_1 + k_{23}p_2 + k_{33}p_3) \mod 26$$

用行向量和矩阵表示如下：

$$(c_1 \quad c_2 \quad c_3) = (p_1 \quad p_2 \quad p_3)\begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \mod 26$$

或

$$\boldsymbol{C} = \boldsymbol{PK} \mod 26$$

这里$\boldsymbol{C}$和$\boldsymbol{P}$是长度为3的行向量，分别代表密文和明文，$\boldsymbol{K}$是一个3×3矩阵，代表加密密钥。运算按模 26 执行。

解密则需要用到矩阵$\boldsymbol{K}$的逆，把逆矩阵$\boldsymbol{K^{-1}}$应用到密文上，则可以恢复明文。

用一般术语，Hill 密码系统可以表示如下：

$$\boldsymbol{C} = E(\boldsymbol{K}, \boldsymbol{P}) = \boldsymbol{PK} \mod 26$$

$$\boldsymbol{P} = D(\boldsymbol{K}, \boldsymbol{C}) = \boldsymbol{CK^{-1}} \mod 26 = \boldsymbol{PKK^{-1}} = \boldsymbol{P}$$

Hill 密码可以做到完全隐蔽单字母频率特性。实际上，Hill 密码体制加密用的密钥矩阵越大所隐藏的频率信息就越多。

尽管 Hill 密码足以抗唯密文攻击，但是它较易被已知明文攻击破解。对于一

个$m \times m$的 Hill 密码，假如有m个明密文对，每个长度都是m，对于第j个明密文对，定义$\boldsymbol{P}_j = (p_{1j}, p_{2j}, \cdots, p_{mj})$和$\boldsymbol{C}_j = (c_{1j}, c_{2j}, \cdots, c_{mj})$，使得对每个$\boldsymbol{C}_j$和$\boldsymbol{P}_j$，都有$\boldsymbol{C}_j = \boldsymbol{P}_j\boldsymbol{K}$，其中$\boldsymbol{K}$是未知的矩阵形密钥。现在定义两个$m \times m$的矩阵$\boldsymbol{X} = (p_{ij})$和$\boldsymbol{Y} = (c_{ij})$。那么我们可以得出矩阵等式$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{K}$，若$\boldsymbol{X}$可逆，则可得$\boldsymbol{K} = \boldsymbol{X}^{-1}\boldsymbol{Y}$。若$\boldsymbol{X}$不可逆，那么可以另找$\boldsymbol{X}$和对应的$\boldsymbol{Y}$，直至得到一个可逆的$\boldsymbol{X}$。

5.2 算法实现的伪代码

Hill 密码加密算法 利用 Hill 密码加密

输入： 明文x，密钥key，维数m
输出： 密文y

```
function encrypt(x, key, m)
    for i ← 1 to length(x) by m do
        for j ← 1 to m do
            mx[j] ← x[i + j]
        end for
        my ← mx * key
        for j ← 1 to m do
            y[i + j] ← my[j]
        end for
    end for
    return y
end function
```

Hill 密码解密算法 利用 Hill 密码解密

输入： 密文y，密钥key，维数m
输出： 明文x

```
function decrypt(y, key, m)
    for i ← 1 to length(y) by m do
        for j ← 1 to m do
            my[j] ← y[i + j]
        end for
        mx ← my * key^-1
        for j ← 1 to m do
            x[i + j] ← mx[j]
        end for
    end for
    return x
end function
```

Hill 密码攻击算法 对 Hill 密码实施已知明文攻击

输入： 明文x，密文y，维数m

输出： 密钥key

```
function cracker(x, y, m)
    for s ← 0 to length(x)/m − 1 do
        for i ← 1 to m do
            for j ← 1 to m do
                mx[i, j] ← x[(s + i) * m + j]
                my[i, j] ← y[(s + i) * m + j]
            end for
            if det(mx) ≠ 0 then
                key ← mx^(−1) * my
                return key
            end if
        end for
    end for
    return 破解失败
end function
```

5.3 算法执行效率分析

我们通过使用不同阶数的密钥矩阵对相同长度密文进行加密，比较算法的执行效率，实验结果如下表所示。

表 1 不同阶数矩阵的执行效率（单位：毫秒）

密钥矩阵阶数	执行时间
4	4.38
5	21.86
6	138.64
7	1122.95
8	11336.25

从表中可以看出，随着矩阵阶数的增长，破解密码所需的时间将会按照指数级增长，由于破解时需要对矩阵进行求逆操作，所以算法复杂度会比较高，约为指数量级，因此对于超大型的密钥矩阵，破解将会耗费较长的时间。

5.4 测试样例及运行结果

作为测试，我们使用三阶密钥矩阵对如下明文进行加密：

$$Today\ is\ 2019/3/19, this\ is\ a\ sunny\ day, we\ feel\ happy\ \text{:-)}$$

加密后的密文为：

$jzn`p[T;] + 9|@\{5@Cc\#`(P/E6@\backslash * HaK[/('h\&Zo,lCENByC\^YRwES + \%wP;q)$

利用密钥重新对密文进行解密，我们得到与明文相同的文本，证明我们的加解密算法是正确的。接着，我们使用以上的明密文对，采取已知明文攻击尝试得到密钥矩阵，破解算法返回的可能密钥矩阵是：

$$\begin{pmatrix} 3 & 9 & 1 \\ 6 & 5 & 1 \\ 3 & 4 & 9 \end{pmatrix}$$

该矩阵与我们之前使用的密钥矩阵相一致，证明我们的破解算法是正确的。

6 总结

在本次实验中，我们实现了多种传统加密技术，以及对传统加密方案的破解算法。传统加密技术曾起到过显著的作用，有着重要的历史意义，通过实验我们进一步熟悉了传统加密技术，体会到了其中使用的代替和置换的重要思想，在编程时，我进一步思考了可以对传统加密技术的改进方法。本次实验的难点在于对传统加密方案的攻击，首先对于单表代替密码的攻击，我们需要对答案的可能性进行界定，即定义一个量化标准来得到不同答案的可能性，然后利用英文字母的分布频率来计算可能的前N个代替表；其次，在 Hill 密码中，我们利用 Python 类的机制，实现了矩阵类和模 95 剩余类环，进一步编写了 Hill 密码的加密解密以及已知明文的攻击方法，由于矩阵不一定可逆，我们需要进行多次尝试，直到得到答案，我们实现的算法可以对 95 个可打印的 ASCII 码字符进行加解密以及已知明文攻击。最后我们还探究了攻击算法的执行效率，发现矩阵越大需要消耗更多的计算时间。