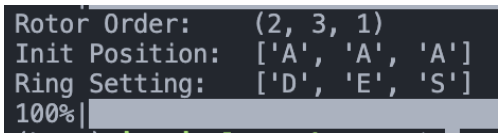
现代密码学 · Hw1

计01　容逸朗　2020010869

Part1

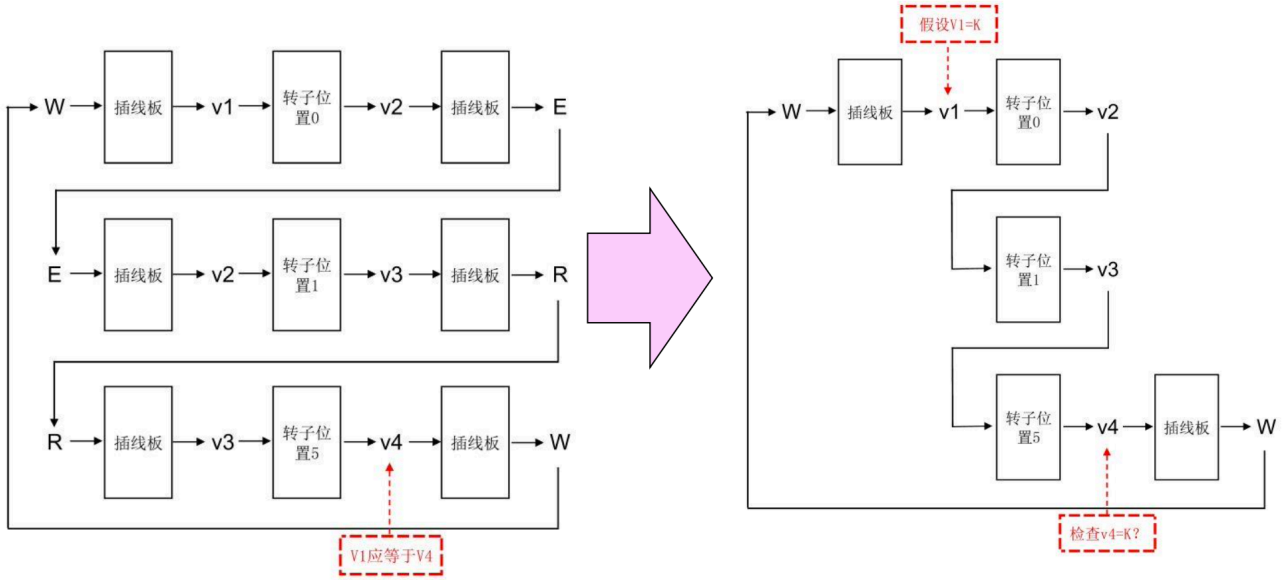
(1)

- 攻击方法是找到密文构成的字母链，同时枚举不同配置下得到的字母链，若两者比较后无异，则说明设定正确，破译成功。
- 为了达成这一目标，我们需要知道下列事实：
  - 所有字母只需在同一位置上出现一次：这是因为同一位置时转子设定相同，使用同一个字母加密得到的另一个字母是不会变的；
  - 同时，也不会出现两个不同字母在同一设置下加密后变为相同字母的情况，这是由于配置了反射器所致的；
- 综上，我们只需把 A 至 Z 依次放入密码中的每一位进行模拟即可，一种简单的配置是从 "AAAAAA", "BBBBBB", 测试到 "ZZZZZZ"，这样做可以得到三个包含所有字母的字母链组。
- 每一次测试的过程如下：
  - 根据德军的加密方式，首先利用待测状态的 enigma 加密密码两次；
  - 例如 "AAAAAA" 加密后得到 "BDZGOW"；
  - 此时 1/4 字母链增加 B-G 的连线，2/5 字母链增加 D-O 的连线，如此类推；
  - 完成所有连线后，统计字母链的条数及每组大小，与密文字母链比较，若相同则为可行配置；
- 经试验，题给的密文是由下列配置加密所得的。
  - 转子顺序：II- III-I
  - Ring Setting：D-E-S
  - Initial Position: A-A-A



(2)

- 首先，我们需要找到密 / 明文对组成的环，例如
  - $P_4S(B) = S(J), P_7S(J) = S(X), P_{11}S(X) = S(B)$；
  - 上述的加密过程可以表述为一个以 $S(B)$ 为开端的环：$S(B) = P_{11}P_7P_4S(B)$；
  - 注意每个环的起始点应相同（例如另一个环可以为 $S(B) = P_{11}P_8^{-1}P_1^{-1}P_{16}^{-1}P_6^{-1}P_4S(B)$ ），这样做可以快速减少密钥空间的规模；
- 原因在于，这样的环可以消除插线板对结果的影响；



- 找到尽可能多的环后，我们可以尝试所有可能的配置（包括转子顺序, Ring Setting 和初始位置），配置确定后，我们也就得到了上式中的 P_i；
- 在不知道插线板的配置下，我们可以暴力枚举 $S(B)$ 的取值为 26 个字母中的一个；
- 在使用上述配置的情况下，若能找到一个满足条件的 $S(B)$ ，则保留此配置，否则排除之；
- 经过 17 个环检验后，仅剩余一个可行配置，此配置就是我们的目标：

```
Testing ring #16.
100%|| 2/2 [00:00<00:00, 207.54it/s]
Testing ring #17.
100%|| 1/1 [00:00<00:00, 206.71it/s]
Testing ring #18.
100%|| 1/1 [00:00<00:00, 206.42it/s]
Testing ring #19.
100%|| 1/1 [00:00<00:00, 207.16it/s]
Testing ring #20.
100%|| 1/1 [00:00<00:00, 186.93it/s]
Testing ring #21.
100%|| 1/1 [00:00<00:00, 186.68it/s]
Testing ring #22.
100%|| 1/1 [00:00<00:00, 186.87it/s]
Testing ring #23.
100%|| 1/1 [00:00<00:00, 186.80it/s]
==================================
Time used: 3047.7s
Result:
 [['DES', ['', '', '', 'A', '', '']]]
==================================
```

- 经试验，题给的密文是由下列配置加密所得的。
  - 转子顺序：II- III-I
  - Ring Setting：D-E-S
  - Initial Position: A-A-A

Part2

1.5

- 利用 $d_K(y) = (y - K) \mod 26$ 并枚举 K 的取值，得到如下解密结果：

K	明文 d_K
0	BEEAKFYDJXUQYHYJIQRYHTYJIQFBQDUYJIIKFUHCQD
1	ADDZJEXCIWTPXGXIHPQXGSXIHPEAPCTXIHHJETGBPC
2	ZCCYIDWBHVSOWFWHGOPWFRWHGODZOBSWHGGIDSFAOB
	...
16	LOOKUPINTHEAIRITSABIRDITSAPLANEITSSUPERMAN
	...

- 从上表可以发现 K=16 时得到了有意义的明文:

```
Look up in the air. It's a bird, it's a plane, it's superman!
```

1.16

(a)

- 逆置换为:

y	1	2	3	4	5	6	7	8
$\pi^{-1}(y)$	2	4	6	1	8	3	5	7

(b)

- 稍加整理，得到:

$\pi^{-1}(y)$	1	2	3	4	5	6	7	8
y	4	1	6	2	7	3	8	5

- 此时先将密文为为八个一组:
  - ETEGENLM / DNTNEOOR / DAHATECO / ESAHLRMI
- 再用逆置换 π^{-1} 解密得到:
  - gentleme/ndonotre/adeachot/hersmail
- 对应的有意义明文为:

```
Gentlemen do not read each other's mail.
```

1.21

(a)

- 首先统计文中出现字母的频率：

字母	频率/%	字母	频率/%	字母	频率/%
C	14.453	Z	5.0781	H	2.3438
G	8.9844	E	4.6875	P	2.3438
S	7.8125	O	3.9062	A	1.9531
K	7.0312	F	3.5156	M	1.9531
I	5.8594	D	3.1250	W	1.9531
Y	5.8594	J	2.7344	Q	0.3906
U	5.4688	L	2.7344	B/R	0.0000
N	5.0781	X	2.7344	T/V	0.0000

- 首先，我们知道密文 $d_K(F) = w$，而从上表中可以猜测 $d_K(C) = e$；
- 又因为密文词组 ZC 的出现的次数较多，猜想 $d_K(Z) = h$，此时原文变为：

```
EMGLOSUDCGDNCUSWYSFHNSFCYKDPUMLWGYICOXYSIPJCKQPKUGKMGOLICHINCGACKSNISA
        E   E     W   WE           E       E           E   E  E
CYKZSCKXECJCKSHYSXCGOIDPKZCNKSHICGIWYGKKGKGOLDSILKGOIUSIGLEDSPWZUGFZCC
E  H E   E E      E      HE     E                                  H  WHEE
NDGYYSFUSZCNXEOJNCGYEOWEUPXEZGACGNFGLKNSACIGOIYCKXCJUCIUZCFZCCNDGYYSFE
      W  HE      E          H  E  W      E     E  E  E  HEWHEE      W
UEKUZCSOCFZCCNCIACZEJNCSHFZEJZEGMXCYHCJUMGKUCY
    HE  EWHEE E  EH   E  WH  H    E  E      E
```

- 从上图可见，明文中出现了三个 whee，进一步观察，发现他们对应密文的后一个字母都是 N，因此猜想 $d_K(N) = l$，因为这样可以组成词组 wheel；
- 接下来查找和明文 he （密文 ZC）匹配的项，密文 UZC 出现了两次，KZC 和 SZC 各一次，故猜想 $d_K(U) = t$，此时密文 UZCFZCCN 对应 the wheel，相对合理；
- 在 $d_K(U) = t$ 对应的前提下，查找密文 U 的相邻词，得到密文 US 出现三次，UM, UC, UZ 出现两次，猜测 $d_K(S) = o$；（因为明文 h 已经出现，而 to 是仅次于 th 的常见字）

```
EMGLOSUDCGDNCUSWYSFHNSFCYKDPUMLWGYICOXYSIPJCKQPKUGKMGOLICHINCGACKSNISA
     OT E  LETO  OW LOWE    T       E   O   E    T       E  LE  E OL O
CYKZSCKXECJCKSHYSXCGOIDPKZCNKSHICGIWYGKKGKGOLDSILKGOIUSIGLEDSPWZUGFZCC
E  HOE   E E O  O E      HEL O  E            O      TO     O  HT WHEE
NDGYYSFUSZCNXEOJNCGYEOWEUPXEZGACGNFGLKNSACIGOIYCKXCJUCIUZCFZCCNDGYYSFE
L    OWTOHEL    LE      T   H  E LW   LO E     E  E TE THEWHEEL    OW
UEKUZCSOCFZCCNCIACZEJNCSHFZEJZEGMXCYHCJUMGKUCY
T  THEO EWHEELE  EH  LEO WH  H    E  E T   TE
```

- 注意到最后一行出现了明文 the o_e wheel，故猜想 $d_K(O) = n$；
- 接下来讨论频率大于 5% 的字母：
  - 没有对应的字母包括 G, K, I, Y；
  - 查找常用字母表，发现 A, I, S, R 出现的频率较高
  - 经暴力查找，发现 $d_K(G) = a$，$d_K(K) = s$ 及 $d_K(Y) = r$ 可以生成较合理的文本

```
EMGLOSUDCGDNCUSWYSFHNSFCYKDPUMLWGYICOXYSIPJCKQPKUGKMGOLICHINCGACKSNISA
  A NOT EA LETO ROW LOWERS  T   AR EN RO   ES  STAS AN  E  LEA ESOL O
CYKZSCKXECJCKSHYSXCGOIDPKZCNKSHICGIWYGKKGKGOLDSILKGOIUSIGLEDSPWZUGFZCC
ERSHOES  E ESO RO EAN   SHELSO  EA  RASSASAN  O  SAN TO A   O  HTAWHEE
NDGYYSFUSZCNXEOJNCGYEOWEUPXEZGACGNFGLKNSACIGOIYCKXCJUCIUZCFZCCNDGYYSFE
L ARROWTOHEL  N LEAR N  T   HA EALWA SLO E AN RES E TE THEWHEEL ARROW
UEKUZCSOCFZCCNCIACZEJNCSHFZEJZEGMXCYHCJUMGKUCY
T STHEONEWHEELE  EH  LEO WH  H A  ER E T ASTER
```

- 观察上文，发现明文 (W)rassasan，猜想断句为 (W)rass as an，此处 $d_K(W) = g$；
- 填入 G 后，第一行出现了 to grow (H)lowers 和 gar(I)en，结合上下文知 $d_K(H) = f$，$d_K(I) = d$；

```
EMGLOSUDCGDNCUSWYSFHNSFCYKDPUMLWGYICOXYSIPJCKQPKUGKMGOLICHINCGACKSNISA
  A NOT EA LETOGROWFLOWERS  T  GARDEN ROD  ES  STAS AN DEFDLEA ESOLDO
CYKZSCKXECJCKSHYSXCGOIDPKZCNKSHICGIWYGKKGKGOLDSILKGOIUSIGLEDSPWZUGFZCC
ERSHOES  E ESOFRO EAND  SHELSOFDEADGRASSASAN  OD SANDTODA   O GHTAWHEE
NDGYYSFUSZCNXEOJNCGYEOWEUPXEZGACGNFGLKNSACIGOIYCKXCJUCIUZCFZCCNDGYYSFE
L ARROWTOHEL  N LEAR NG T   HA EALWA SLO EDANDRES E TEDTHEWHEEL ARROW
UEKUZCSOCFZCCNCIACZEJNCSHFZEJZEGMXCYHCJUMGKUCY
T STHEONEWHEELED EH  LEOFWH  H A  ERFE T ASTER
```

- 继续分析第一行 Not (D)ea(D)le to grow flowers，此处令 $d_K(D) = b$，可得 not be able to grow flowers；
- flowers 后出现了 b(P)t，此处语气应与前文 not 相对，表转折，故 $d_K(P) = u$；
- 此时还剩下 E 的出现频率较高，尝试与常用字母表中剩余的 I 匹配；
- 第一行还剩下 I (M)a(L) not be able to ...，利用剩余的十个字母作暴力匹配，认为 may 是较合理的选项，故选取密文 $d_K(M) = m$，以及密文 $d_K(L) = y$；
- 最后查找出现频率达 2% 的 J 及 X，出现频率相近的剩余明文有 C, P，以 $d_K(J) = c$，以及使 $d_K(X) = y$，得到：

```
EMGLOSUDCGDNCUSWYSFHNSFCYKDPUMLWGYICOXYSIPJCKQPKUGKMGOLICHINCGACKSNISA
IMAYNOTBEABLETOGROWFLOWERSBUTMYGARDENPRODUCES USTASMANYDEFDLEA ESOLDO
CYKZSCKXECJCKSHYSXCGOIDPKZCNKSHICGIWYGKKGKGOLDSILKGOIUSIGLEDSPWZUGFZCC
ERSHOESPIECESOFROPEANDBUSHELSOFDEADGRASSASANYBODYSANDTODAYIBOUGHTAWHEE
NDGYYSFUSZCNXEOJNCGYEOWEUPXEZGACGNFGLKNSACIGOIYCKXCJUCIUZCFZCCNDGYYSFE
LBARROWTOHELPINCLEARINGITUPIHA EALWAYSLO EDANDRESPECTEDTHEWHEELBARROWI
UEKUZCSOCFZCCNCIACZEJNCSHFZEJZEGMXCYHCJUMGKUCY
TISTHEONEWHEELED EHICLEOFWHICHIAMPERFECTMASTER
```

- 从 I ha(A)e always 中可知密文 $d_K(A) = v$，(Q)ust as many defd ... 可知 $d_K(Q) = j$；
- 至此，我们得到了原文：

```
I may not be able to grow flowers but my garden produces just as many
defd leaves, old overshoes, pieces of rope, and bushels of dead grass
as anybody's. And today I bought a wheelbarrow to help in clearing it
up. I have always loved and respected the wheel barrow. It is the one-
wheeled vehicle of which I am perfect master.
```

注：第二行中 defd 应为 dead，可能是由于密文有误造成的。

(b)

- 首先找到长度为 3 且重复出现的密文段，可以发现密文 HJV 出现了 5 次，次数最多；
- 统计 HJV 出现在密文的位置，分别为 107, 125, 263, 317, 329，对应距离为 0, 18, 156, 210, 222，距离的最大公约数为 6，猜想密钥的长度为 6；
- 接下来进行频数统计：

i	$M_g(y_i)$	K
0	0.032, 0.036, **0.065**, 0.039, 0.034, 0.042, 0.037, 0.031, 0.042, 0.046, 0.025, 0.034, 0.038, 0.042, 0.038, 0.046, 0.036, 0.040, 0.042, 0.033, 0.03, 0.039, 0.043, 0.034, 0.042, 0.034	2
1	0.038, 0.039, 0.049, 0.042, 0.040, 0.036, 0.045, 0.030, 0.027, 0.036, 0.045, 0.031, 0.035, 0.048, 0.040, 0.033, 0.036, **0.071**, 0.037, 0.030, 0.029, 0.036, 0.030, 0.038, 0.046, 0.037	17
2	0.035, 0.036, 0.034, 0.038, 0.036, 0.041, 0.028, 0.038, 0.034, 0.042, 0.041, 0.046, 0.040, 0.043, 0.036, 0.032, 0.035, 0.039, 0.042, 0.031, 0.039, 0.033, 0.035, 0.044, **0.059**, 0.045	24
3	0.045, 0.038, 0.044, 0.037, 0.037, 0.038, 0.031, 0.033, 0.039, 0.037, 0.037, 0.051, 0.041, 0.031, 0.035, **0.066**, 0.037, 0.030, 0.039, 0.041, 0.025, 0.035, 0.041, 0.033, 0.035, 0.044	15
4	0.040, 0.033, 0.034, 0.040, 0.045, 0.034, 0.043, 0.046, 0.047, 0.034, 0.034, 0.036, 0.034, 0.035, 0.034, 0.044, 0.034, 0.036, 0.035, **0.056**, 0.041, 0.035, 0.043, 0.044, 0.031, 0.032	19
5	0.042, 0.038, 0.037, 0.042, 0.039, 0.027, 0.033, 0.039, 0.037, 0.034, 0.048, 0.035, 0.025, 0.037, **0.070**, 0.042, 0.032, 0.032, 0.039, 0.033, 0.04, 0.041, 0.035, 0.037, 0.039, 0.048	14

- 由此可知，密钥很可能为 K = (2, 17, 24, 15, 19, 14)，利用 K 解密后的明文为：

```
I learned how to calculate the amount of paper needed for a room when I
was at school. You multiply the square footage of the walls by the
cubic contents of the floor and ceiling combined and double it. You
then allow half the total for openings such as windows and doors. Then
you allow the other half for matching the pattern. Then you double the
whole thing again to give a margin of error, and then you order the
paper.
```

(c)

- 首先统计文中出现字母的次数：

字母	次数	字母	次数	字母	次数
C	32	F	10	X	2
B	21	D	9	H	1
K	20	J	6	N	1
P	20	U	6	S	1
I	16	Q	4	Y	1
A	13	V	4	G/L	0
E	13	Z	4	M/T	0
R	12	O	2	W	0

- 由于 C 的出现次数远高于其他字母，因此猜测 $d_K(C)=e$，以数字表达即为 $e_K(4)=2$，又因为 $e_K(x)=ax+b$，由此可得：

$$4a+b=2$$

- 又因为 $gcd(a,26)=1$，故可以枚举 a 的取值：

a	b	a^{-1}	$d_K(y)$
1	24	1	$d_k(y)=y-24$
3	16	9	$d_k(y)=9(y-16)$
5	8	21	$d_k(y)=21(y-8)$
7	0	15	$d_k(y)=15y$
9	18	3	$d_k(y)=3(y-18)$
11	10	19	$d_k(y)=19(y-10)$
15	20	7	$d_k(y)=7(y-20)$
17	12	23	$d_k(y)=23(y-12)$
19	4	11	$d_k(y)=11(y-4)$
21	22	5	$d_k(y)=5(y-22)$
23	14	17	$d_k(y)=17(y-14)$
25	6	25	$d_k(y)=25(y-6)$

- 当 $a=19,b=4$ 时，利用解密函数 $d_k(y)=11(y-4)$ 可以得到明文：

```
Ô Canada!
Terre de nos aïeux,
Ton front est ceint de fleurons glorieux!
Car ton bras sait porter l'épée,
Il sait porter la croix!
Ton histoire est une épopée
Des plus brillants exploits.
Et ta valeur, de foi trempée,
Protégera nos foyers et nos droits.
```

注：明文是加拿大国歌（法语）。

(d)

- 首先统计字母出现的频率：

字母	频率/%	字母	频率/%	字母	频率/%
L	6.1662	F	4.2895	D	2.4129
S	6.1662	G	4.2895	X	2.4129
Y	5.8981	I	4.2895	P	2.1448
E	5.6300	R	4.0214	O	1.8767
M	5.6300	K	3.4853	Z	1.8767
V	4.8257	J	3.2172	Q	1.6086
C	4.5576	T	3.2172	N	1.0724
A	4.5576	U	3.2172		
H	4.5576	W	2.9491		

- 从表中可见，字母出现的频率相对集中，因此猜测密码为维吉尼亚密码，统计密文段，发现长度为 3 且重复出现的密文段的重复次数只有 2，故采用暴力枚举的方式；
  - 若密钥长度为 1，则重合指数为 0.044；
  - 密钥长为 2，重合指数分别为 0.049 和 0.051；
  - 密钥长为 3，重合指数分别为 0.052, 0.056 和 0.056；
  - 密钥长为 4，重合指数分别为 0.053, 0.068, 0.057 和 0.058；
  - 密钥长为 5，重合指数分别为 0.058, 0.055, 0.054, 0.057 和 0.050；
  - 密钥长为 6，重合指数分别为 0.066, 0.078, 0.070, 0.084, 0.071 和 0.084；
- 经测试，当密钥长为 6 时可以解出密钥 $K = (19, 7, 4, 14, 17, 24)$，对应密文为：

```
I grew up among slow talkers. Men in particular who dropped words a few
at a time like beans in a hill. And when I got to Minneapolis, where
people took a Lake Wobegon comma to mean the end of a story, I couldn't
speak a whole sentence in company and was considered not too briaht, so
I enrolled in a speech couqse taught by Orville Sand the founder of
reflexive relaxology, a self hypnotic technique that enabled a person
to speak up to three hundred words per minute.
```

注：上文中 briant 应为 bright，couqse 应为 course，书上的密文可能印错了。

1.25

- 首先统计密文中不同二元加密组出现的次数，发现 TH 出现 4 次，LM 出现 3 次。
- 对 TH 和 LM 这两个字母组，可以利用常见两字母表的字母组暴力枚举，
  - 利用 8 个元素组成的四条方程可以唯一确定一个二阶矩阵；
  - 得到矩阵后，求矩阵的逆，我们只考虑那些可以求逆的矩阵；
- 经试验，当 $d_K(TH)=in$ 和 $d_K(LM)=th$ 时， 明文有意义，对应的明文为：

```
The king was in his counting house, counting out his money.
The queen was in the parlour, eating bread and honeyz.
```

注：此时 $K=\begin{bmatrix}4 & 11\\13 & 9\end{bmatrix}, K^{-1}=\begin{bmatrix}23 & 21\\13 & 16\end{bmatrix}$.