代換加密法簡介

許介彥
大葉大學　通訊與計算機工程學系

前言

在本刊第239期「密碼學初步」一文中，筆者提及設計加密法的兩種基本運算－換位（transposition）與代換（substitution），並在該文中對前者有比較詳細的說明；本文將繼續為讀者介紹代換運算。

以換位運算做成的加密法常統稱為「換位加密法」（transposition ciphers），而以代換運算做成的加密法則稱為「代換加密法」（substitution ciphers）。代換加密法依不同的做法又可細分為四個類型：

- 簡單型代換（simple substitution）
- 同音型代換（homophonic substitution）
- 多對應代換（polyalphabetic substitution）
- 多區塊代換（polygram substitution）

本文限於篇幅，將針對前面兩種類型作概略性的介紹。

資料的編碼方式

西方國家的文字是由總數不多的字母組成，以英文為例，字母總共有26個，而且這26個字母有一個公認的字母順序（由A到Z）。如果我們將這26個字母由數字0開始依序編號：

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

然後將英文資料中的每個字母用上表中該字母所對應的編號取代，在解讀上並不會造成太大的困擾，因為數字很容易可以被還原為字母。舉例來說：

Y E S T E R D A Y

⇕

24 4 18 19 4 17 3 0 24

這個做法其實已經可以說是一種簡單的代換加密法，不過它太容易被破解了，因此並不實用，然而「將明文看成是由數字組成」的概念卻很有用，因為人類對數字（尤其是整數）已經研究了好幾千年，非常習慣於數字的計算，也累積了許多關於數字的知識；這樣的做法可以讓加密的方法得以和長久以來人類對數字的研究相結合，甚至可以利用數學上眾所公認的難題來設計出不易被破解的加密法。

對電腦來說，將英文字母看成數字當然不是什麼新鮮事，因為電腦內部所有各種類型的資料本來就都是以數字的方式來儲存，一般的字元在電腦內部也都有固定的與數字對應的方式（如ASCII或EBCDIC等）。

本文探討的加密法中，有時候我們會將明文看成是由數字組成的；除非特別聲明，否則就假設是採用將字母A當成0，B當成1，C當成2……的編碼方式；當然，這只是字母與數字無窮多種可能的對應方式之一。

求餘數的運算：「模」

「數論」(Number Theory) 是數學的一個分支，專門研究整數的性質，密碼學的許多理論與數論有著密切的關係；我們要先看數論中關於整數的一個基本運算－「模」(mod)。

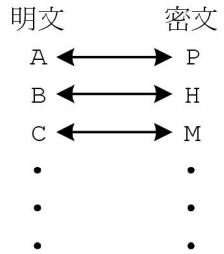
簡單地說，如果 a 是一個整數且 n 是一個正整數，$a \bmod n$ 的值就是 a 除以 n 的餘數，其值必須大於等於 0 而且小於 n。

舉例來說，$40 \bmod 17 = 6$，因為 $40 = 2 \times 17 + 6$；而 $-40 \bmod 17 = 11$，因為 $-40 = -3 \times 17 + 11$。

有了以上基本概念，我們開始來介紹代換加密法。

壹、簡單型代換加密法（Simple Substitution Ciphers）

利用簡單型代換加密法加密時，明文中的每個字母 (或數字) 在密文中被另一個字母 (或數字) 取代，例如：將明文中的字母 A 全用 P 取代，將字母 B 全用 H 取代等，亦即明文中由 A 至 Z 這 26 個字母在密文中分別對應到被打亂順序的另外 26 個字母，明文與密文中的字母有一對一的對應關係，如下圖所示：



關鍵字加密法（Keyword Cipher）

加密與解密雙方約定好以某個英文單字做為關鍵字。以人工進行加密時，先將由 A 到 Z 這 26 個字母由左而右依序寫在同一列上，然後在此列的正下方由左而右依序寫出關鍵字的各個字母（字母若重覆出現的話只寫第一次），緊接著再依序寫出由 A 到 Z 這 26 個字母中尚未出現在關鍵字中的字母。舉例來說，如果以 WEDNESDAY 做為關鍵字，依照上述做法寫出以下兩列：

A	B	C	D	E	F	G	H	I	J	K	L	M
W	E	D	N	S	A	Y	B	C	F	G	H	I

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
J	K	L	M	O	P	Q	R	T	U	V	X	Z

其中，陰影部分為關鍵字寫入後的結果；接著就可以進行加密了，要為明文中的某個字母加密時，先在第一列找到這個字母出現的位置，然後就用它正下方的字母來取代原來的字母。例如，以 WEDNESDAY 做為關鍵字，則

明文：S C I E N C E E D U C A T I O N

密文：P D C S J D S S N R D W Q C K J

要解密時，只要知道加密時所用的關鍵字，即可輕易地將密文還原為明文。

移位加密法（Shift Cipher）

加密時，將明文中的每個英文字母都以其在由 A 至 Z 的字母順序中的位置往後移 k 個位置所得的新字母取代，其中 k 是一個介於 0 與 25 間的整數。舉例來說，如果 k 等於 2，每個字母向後移兩位，則 A 被 C 取代，B 被 D 取代，C 被 E 取代……等，因此

明文：W A L K S L O W L Y I N Z O O

密文：Y C N M U N Q Y N A K P B Q Q

注意：在字母順序中，Z 的下一個字母又繞回到 A。

羅馬帝國的凱撒大帝（Julius Caesar，100-44 B.C.）曾經實際使用過移位加密法，他當時所用的 k 是 3，因此，移位加密法在 k 等於 3 時又稱為凱撒加密法（Caesar Cipher）。

讀者也許已經察覺，移位加密法是非常容易被破解的加密法，因為可能的鑰匙（也就是 k）總共才 26 個（0 至 25），如果密文被截獲，即使不知道加密時所用的 k 是多少，也不難將所有可能的明文列出而從中挑出正確的明文。

仿射加密法（Affine Cipher）

如果將英文字母用前面所說的方式由 A 至 Z 依序編號，那麼移位加密法中新字母與舊字母的關係其實可以簡單地用一個函數來表示：

$$f(a) = (a + k) \bmod 26$$

其中，a 與 $f(a)$ 分別為明文與密文母的編號且 $0 \le k \le 25$。由相加我們很自然地會聯想到相乘，由相乘也可以發展出類似的加密法：

$$f(a) = ak \bmod 26$$

為了讓函數 f 是一對一函數，這裡的 k 必須和 26 互質，否則可能會使得明文中不同的字母對應到密文中相同的字母，而在解密時無法準確地將密文還原為明文。舉例來說，如果 k 為 13（不與 26 互質），則

$f(A)= f(C)= f(E)= \cdots = f(Y)$=A(也就是 0)

$f(B)= f(D)= f(F)= \cdots = f(Z)$=N(也就是 13)

密文將只由字母 A 與 N 組成，使得接收端在解密時遭遇到困難。

相加與相乘又可以組合出如下的加密法：

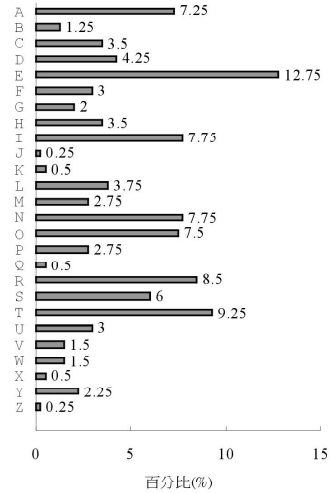
$$f(a) = (ak_1 + k_0) \bmod 26$$

其中 $0 \le k_0, k_1 \le 25$ 且 k_1 與 26 互質。這種類型的函數稱做仿射函數（affine function），因此以這種函數做成的加密法稱為仿射加密法（affine cipher）。

簡單型代換加密法之破解

簡單型加密法在加密過程中都是將明文中的每個字母以固定的另一個字母取代，例如：在凱撒加密法中，字母 A 固定以 D 取代，字母 B 固定以 E 取代等；不同的加密法或鑰匙只是讓用以取代的字母不同而已，這類加密法對專業人士而言，通常很容易破解。

要破解這類加密法常須借助一些統計資料。如果我們隨便找來許多英文的書報雜誌，針對由 A 到 Z 的每個字母在這些文件中出現的次數做統計，將會發現字母 E 是所有英文字母中出現次數最多的字母，比率上而言，約佔了全部字母的 13%，也就是說，隨意截取一段英文資料，此資料的每 100 個字母中，E 大約會出現 13 次。依此類推，如果我們將統計後的各個字母依其在全部字母中所占的百分比畫成一個圖表，所得結果大致上將如下所示（數字代表百分比）：

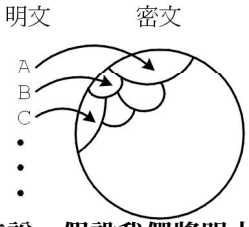


假設某份由簡單型加密法加密的密文被第三者截獲，他可以統計密文中各個字母出現的次數與比率，然後與上表對照；如果字母 K 在密文中出現的次數比其他字母都要多，所佔的百分比又大概是 13%，密文中的字母 K 很可能就對應到明文中的字母 E。依此類推，第三者有可能透過密文中各字母出現的比率與上表比對來將密文中的字母一個一個「還原」，或者經由已知的幾個密文與明文字母間的對應關係而猜出所用的加密法及鑰匙。

上面的圖表是由一般的英文書報資料統計所得的結果，可以用於破解一般英文資料經簡單型加密法加密的密文。對較特殊的英文資料（如：法律條文、武器項目、電腦程式等），文件中所用或常用的英文單字有可能迥異於一般的英文資料，如果要破解這類資料，當然也可以事先收集許多此類型的文件並計算出各個字母出現的比率以得知該類型資料的特性。另外，許多的文件在書寫時常須遵循一定的格式，例如 PASCAL 語言所寫的程式在一個區塊（block）的前後會有 begin 與 end 這兩個保留字，一封電子郵件的開頭一定會有寄件者及收件者的位址等，第三者如果已經知道資料的某些格式，經由對密文中特殊位置的觀察也常常可以得到許多有利於破解的資訊。

貳、同音代換加密法（Homophonic Substitution Ciphers）

我們已經看到，加密時將明文的每個字母固定對應到密文中的另一個字母或數字並不是很好的方式，因為經由密文中各個字母或數字出現比率的統計，不難找出明文與密文中字母與字母間的對應關係。同音型代換加密法的密文通常是由數字組成，為了提高破解的困難，明文中的每個字母是對應到一群數字（而非單一一個數字）。在加密時，明文中的每個字母在密文中還是被一個數字取代，不過這個數字是由某一群候選的數字中隨機選出，這樣一來，密文中各個數字出現的比率將不像明文中的字母那麼明顯，使得破解的困難度大增。此類加密法中，明文中的字母與密文中的數字為一對多的對應關係，且明文中不同的字母所對應到的密文中的數字一定不同，如下圖所示：



舉例來說，假設我們將明文字母由 A 至 Z 等 26 個字母對應到由 0 到 99 等 100 個數字，而且每個字母對應到的數字個數都正比於該字母在全部字母中所占的比率。以下為一種可能的對應關係（為了節省篇幅，只列出 26 個字母中的 8 個）：

字母	%	對應到的數字
A	7.25	16 29 42 57 66 84 90
E	12.75	02 11 17 26 30 39 48 55 63 72 79 85 93
M	2.75	18 52 89
N	7.75	08 25 38 50 69 73 82 95
O	7.5	01 13 23 35 47 60 75 98
T	9.25	06 14 22 37 45 59 67 77 86
U	3	03 44 76
Y	2.25	33 91

明文若為 MEETYOUATTEN，加密後一種可能的結果為：

明文：M E E T Y O U A T T E N

密文：52 79 11 37 33 98 03 84 22 59 48 69

同音型加密法名稱的由來是因為密文中不同的數字卻可能對應到明文中相同的字母，就好像英文單字中同音異義的概念一樣（如meet與meat，單字不同但是發音相同）。破解此類加密法的困難度要比破解簡單型加密法高出許多，特別是當明文中每個字母所對應到的數字個數正比於該字母在全部字母中出現的比率時（如上述做法），因為此時密文中每個數字出現的次數（及所占的比率）幾乎都一樣，明文中的每個字母在整份文件中所占的比率在密文中被好幾個數字平均分攤掉了。

讀者不難想見，密文中能夠使用的數字或符號越多，做出來的密文將越不容易被破解；如果推到極致，一篇密文中所有的數字或符號都只在密文中出現一次，這樣的密文就可能是沒有任何方法可以破解得了的。

比歐加密法（Beale Cipher）

這個加密法是因探險家 Thomas Jefferson Beale 而得名，他在十九世紀初曾利用這個方法為一段描述某批不為人知的寶藏埋藏地點的文字進行加密，所用的鑰匙是一篇文章－美國的獨立宣言（Declaration of Independence）。以下限於篇幅，只列出獨立宣言的開頭一小部份。加密之前先由頭至尾將整篇文章的每個單字由數字1開始依序編號（因此每個單字都對應到某個數字），下面的每一列最前面的數字是該列第一個單字的編號：

(1)When, in the course of human events, it
(9)becomes necessary for one people to
(15)dissolve the political bands which have
(21) connected them with another, and to
(27) assume among the Powers of the earth the
(35)separate and equal station to which the
(42)Laws of Nature and of Nature's God
(50)entitle them, a decent respect to the
(58)opinions of mankind requires that they
(64)should declare the causes which impel
(71)them to the separation. We hold these
(78)truths to be self-evident; that all men are
(86)created equal, that they are endowed by
(93)their Creator with certain unalienable
(99)rights; that among these are Life, Liberty,
(107)and the pursuit of Happiness.

要加密時，將明文中的每個字母以宣言中某個以此字母開頭的單字所對應到的數字取代（因此密文是由數字組成）。舉例來說，明文中的字母 W 可被數字 1, 19, 40, 66, 72, 290, 459等數字中的任意一個取代，因為在獨立宣言中，這些數字所對應到的單字的開頭第一個字母都是 W（宣言的第 1 個單字為 When，第 19 個單字為 which 等）。因此，如果密文是：115, 73, 24, 818, 37, 52, 49, 17, 31, 62, 657, 22, 7, 15, K，所對應的明文將為 "I have deposited … "。

(下轉第 21 頁)