數論在密碼上的應用(上)

楊重駿 楊照崑

一、前　言

數論，顧名思意，是一門研究數字性質的學問。一般所謂的數論，特指正整數(即自然數)的許多性質，像質數的分佈，方程式的正整數解，韓信點兵，及進位法都包括在數論裏面，我們在小學時候學的分解因數，最大公約數也是數論的一部分，可惜因爲數論在日常生活中沒有什麼直接的用處，在中學數學裏很少提到數論，一般被認爲是一種「純數學」，深而無用。可是「無用之用，眞乃大用」，終於在一九七〇年代後期，幾個電機工程師用數論的一些基本定理，製成了一種新的密碼。這種由數論所作成的密碼與以前人們所用的密碼，有着根本性質上的不同，可說是密碼史上一個空前的革新。

密碼通訊在軍事上的用途是大家都知道的，但由於交通的發達，在分秒必爭的工商業社會裏，商業上的情報也已成爲商業盈利的主要依靠。譬如說，有人早幾個小時知道什麼公司有了一種新的發明，或某兩個公司計劃合作，或某地區有物價的大波動，就可以在股票上上

下其手，轉瞬之間收進大筆的財富。因此公司本身內部及公司與公司之間的通訊都希望能對外嚴守機密。但由於現在通訊無論是有線或無線都很容易被敵方竊聽，因此公司必須對情報加鎖，即所謂密碼通訊。

以往人們在軍事上所用的密碼其基本的形式在於「代換」與「置換」。譬如說，我要發出下面一個消息給你，

"我有一個秘密對你說"

我就先把這幾個字換成數字，即一般電碼本上的代碼，假定「我」字的代碼是 3314 ，「有」字的代碼是 1432 ，「一」字代碼是 0001 ，等等，則上面那句話就成了

331414320001……

代換密碼是把 0 ，1 ，2 ，…，9 十個數字互換，譬如我們可以把 0 換成 2 ，1 換成 3 ，等等，若用群論的符號表示，上面的代換可寫成

$$G=\begin{pmatrix}0&1&2&3&4&5&6&7&8&9\\2&3&5&7&6&4&9&0&8&1\end{pmatrix}$$

這個表示法是上行爲 0 ，1 ，2 ，…，9 ，而下行是他們代換成的新數字，即 0 → 2 ，1 → 3 ，2 → 5 ，…。因此剛才的電碼若用 G 法代換，則成了

773636752223……

這時一個不知道這個代換規則的人看到了上面的信號，他就不能從電碼本子裏找出它的原意了。置換法在於把密碼排成一種雙方都知道的形式，如下圖

```
7       5       5
 7     7 2     4
  3   3   2   1
   6 6     2 3
```

則發出的信號爲 7557724331662 3，同樣的，不知道這種特定圖案的人，很難解開原來的信息。

以代換法爲例，像 G 這類的傳換可以有 $10!=3,628,800$ 種不同的變化，假定我們可以在一分鐘內試一種代換，又假定我們的運氣中等，在試到一半時即 $10!/2=1,814,400$ 時可以成功，則在不吃、不睡、不錯的情形下，我們要試 3 年零 165 天，等迷解出來的時候，伙早已打完了。但這只是最基本的密碼而已，在生死關頭，更難解的密碼必然出籠，例如在代換法中，若兩位兩位的代換（即 00 → 79 ，01 → 85 ，…）則其變化可達 $100!=9.32\times10^{57}$ 種之多，如果我們再用硬試的方法，則一百萬人同心協力也得用 6×10^{48} 年才能試出謎底，可是地球的年齡不過只有 5×10^{9} 年而已。

因此解密碼，都不用硬試的方法去解。一般可用統計的方法根據名字（或字母）出現的頻率及發生的事件加以分析，例如在英語中，各字母出現的頻率按多少排列是

$$e,a,o,i,d,h,n,r,s,t,$$
$$u,y,c,f,g,l,m,w,b,\cdots,$$

因此一個出現次數最多的符號就很可能代表 e ，出現次多的符號就很可能代表 a ，並以此類推，但道高一尺，魔高一丈，如果不斷地變化代換的方法，譬如說前一百個字用代換法 G_1 ，第二個一百個字用代換法 G_2 ，則頻率方法亦失去功效。但是魔高一丈，惡魔又高一丈，重賞之下，天下尚沒有絕對解不開的密碼。二次大戰時，美國密碼兼統計學家弗立得門（W. F. Friedman），在一年半的時間完全破獲了日軍的密碼。在中途島一戰，美國海軍以劣勢的軍力，大勝日本皇家海軍（讀過參加該戰役的日本軍官淵田美津雄及奧宮正武所著的「中途島」一書嗎？）。另外英國也幾乎在同時解開了德軍的密碼，做到了知己知彼的優勢地位（知己知彼並不一定百戰百勝，但總比不知彼要好得多。）

過去這類密碼的特質（也可以說是缺點）在於它們是關閉式的製解法，即收發雙方都必須同時知道這種密碼的構造。因此如果在一通訊系統中有一個聯絡站被間諜滲入或被敵人佔領，則密碼的機密可能全盤暴露。而現在用數論的密碼却是公開式的（Pablic-Key Cryptography），即是只有收方知道密碼的解法

，發方只需要知道做法而已，而且這種製法可以公開。因此即使發方被捕，敵人仍搾不出解密碼的機密來，其程序是這樣的：

1. 收方先告訴發方如何用把情報做成密碼（敵人也聽到了這個做法）
2. 發方依法發出情報的密碼（敵人也聽到了這個信號）
3. 收方解開此密碼爲原信息（但敵人却解不開此密碼）

當然把收發互換，就可以互通信息了。剛才說過這種方法最大的好處就是只有一方知道解碼的秘訣，比以前收發雙方都知道秘訣的保密性高多了，自從這類的密碼法發表之後，在美國軍事界、教育界、工商業界引起了一個蔚然大波。對大多數的數論而言則是一則以喜，一則以懼。喜的是，皇天不負苦心人，一向不容易找到飯吃的數論家突然成爲許多經費充裕的軍工商界所爭取的對象。費馬、尤拉以及那些一生窮困而已作古了的數論家可以含笑九泉了。但懼的是，由於這些理論在軍事通訊上的價值，有關這方面的新發現已被視爲國防機密而列爲管制了。以往各國數論家無憂無慮發表論文自由交換意見的日子也許是一去不返了，前程固然似錦，往事却是如煙，純數學最後一塊淨土也終於被實用所"污染了。然而這次因密碼而推動大家對數論的研究，將在數學史上寫下了有趣的一頁。

這種密碼所用的數論並不深，我們可以全部介紹出來，當然在實際用的時候，數字會大得多，但在大小型電子計算機如此普遍的今日，是不會成問題的。

在我們介紹兩種主要的數論密碼之前，我們先將介紹一點數論。

二、因子、質數、同餘式與費馬、尤拉定理

若m，n爲兩整數，且$m>0$，則以m除n可得二整數α與r，使得

$$n=\alpha m+r$$

其中α稱爲商，r稱爲餘數，且$0\leqslant r<m$。若$r=0$，則我們說n可被m所整除，m爲n之**因子**，n爲m之**倍數**，若一數除1與本身之外無其他因子，則此數稱爲**質數**，例如2，3，5，7，11，都是質數，4，6，9，12，却不是質數。我們定義n與s對m有同餘數；

$$m=s\ (\mathrm{mod}\ m)$$

如果n與s被m除時有相同的餘數。例如

$$12\equiv 2\ (\mathrm{mod}\ 10),$$

$$8\equiv 5\ (\mathrm{mod}\ 3)。$$

有一個關於同餘式的簡單定理，我們把它們列出來，讀者很容易證出來。

定理2.1 若$p\equiv q\ (\mathrm{mod}\ m)$，

$a\equiv b\ (\mathrm{mod}\ m)$，

則$ap=bq\ (\mathrm{mod}\ m)$

若兩正整數p，q的最大公因子（約數）是1，則我們稱p，q**互質**，以

$$(p,q)=1$$

表示之。現在我們要證一個有關兩個互質數的一個基本定理。

定理2.2 若兩正整數p，q互質，則可以找到二整數（不一定正）a，b，使得

$$ap+bq=1$$

證明 令A爲含所有$x=ap+bq>0$，a，b爲整數之集合，此集合顯然不是空集合，因可取$a=b=1$，$p+q>0$。令d爲此集合中之最小者，若$d=1$，則本定理得證，若$d>1$，令$ap+bq=d>1$，則任取此集中之另一數$a'p+b'q$，則我們若以d除$a'p$

$+b'q$ 則我們若以 d 除 $a'p+b'q$ ，則得

$$a'p+b'q=\alpha d+r \quad 0\leqslant r<d$$

代入 $d=ap+bq$

則得

$$(a'-\alpha a)p+(b'-\alpha b)q=r$$

此 r 必爲 0 ，否則 r 爲 A 集中一小於 d 之數，與假設 d 爲最小數相矛盾，因 $r=0$ 故 d 爲 A 集中任何一數之因子。因

$$p\in A\ (a=1,\ b=0)$$

$$q\in A\ (a=0,\ b=1)$$

故 d 爲 p ，q 之公因子。但 p ，q 之最大公因子爲 1 ，故 $d=1$ ，定理證畢。

這是一個極有用的定理，讀者也許要問，我們如何找到 a 與 b 使 $ap+bq=1$ 呢？一般可用**輾轉相除法**。

例 找整數 a ，b ，使得 $5a+9b=1$ 。因

$$9=5+4,\quad 5=4+1,$$

故 $1=5-4=5-(9-5)$

$=2\times 5-9\times 1$

故 $a=2,\quad b=-1$

系 2.2 若 w 與 m 爲二互質的正整數且 $m>w$ ，則可找到一正整數 θ 使得

$$w\theta=1\ (\mathrm{mod}\ m)$$

證明 由定理知，有 a 、b 二整數使得 $aw+bm=1$ ，因 bm 爲 m 之倍數，故

$$aw\equiv 1\ (\mathrm{mod}\ m)$$

令 $a=\phi m+\theta \quad 0\leqslant\theta<m$

則得 $\theta w\equiv 1\ (\mathrm{mod}\ m)$ 且 $\theta\geqslant 0$

因 θ 不可能爲 0 ，故本系得證。

最後我們要用到一個不容易證明的“費馬、尤拉（Fermat-Euler ）定理”。但因爲我們只用到這個定理比較容易證明的特殊形式，我們就只證明簡單的部分。

定理 2.3 若 m 爲質數，w 爲任一與 m 互質之整數，則

$$w^{m-1}\equiv 1\ (\mathrm{mod}\ m)$$

證明 先把 w 寫成 w 個 1 的和，則由多項式定理知

$$(1+1+1+\cdots+1)^m$$

之展開式中除 w 個 1 之外，都含有 m 之因子，（m 爲質數 $m!$ 中之 m 不可能消去），故

$$w^m\equiv w\ (\mathrm{mod}\ m)$$

兩邊乘以系 2.2 中之 a ，即

$$aw\equiv 1\ (\mathrm{mod}\ m)$$

得 $w^{m-1}\equiv 1\ (\mathrm{mod}\ m)$

本定理證畢。

系 2.3 若 m 爲二質數 p 、q 之積，w 爲任一與 m（即同時與 p 與 q ）互質之整數，則

$$w^{(p-1)(q-1)}\equiv 1\ (\mathrm{mod}\ m)$$

證明 先用定理之證明法得

$$w^{p-1}\equiv 1\ (\mathrm{mod}\ m)$$

$$w^{q-1}\equiv 1\ (\mathrm{mod}\ m)$$

由定理 2.1 可得

$$(w^{p-1})^q\equiv 1\ (\mathrm{mod}\ m)$$

即

$$w^{pq-q}\equiv 1\ (\mathrm{mod}\ m)$$

即

$$w^{pq-q}\equiv 1\equiv w^{p-1}\ (\mathrm{mod}\ m)$$

因 w^{p-1} 與 m 互質，由系 2.2 之證明可知存在一 a 使 $aw^{p-1}\equiv 1\ (\mathrm{mod}\ m)$，上式乘以 a 得

$$aw^{pq-q}\equiv 1\ (\mathrm{mod}\ m)$$

即

$$aw^{p-1}\,w^{pq-p-q+1}\equiv w^{(p-1)(q-1)}$$

$$\equiv 1\ (\mathrm{mod}\ m)$$

本系證畢。

我們還要利用到另一個簡單的定理，我們也在這節裏把它證完。

定理 2.4 令 $(a_1, a_2, \cdots, a_n)$ 爲一含 n 個正整數的數列，並滿足

$$\begin{cases} a_1 \geqslant 1 \\ a_2 > a_1 \\ a_3 > a_1 + a_2 \\ \quad \vdots \\ a_i > a_1 + a_2 + \cdots + a_{i-1} \\ \quad \vdots \\ a_n > a_1 + a_2 + \cdots + a_{n-1} \end{cases}$$

又令（$x_1, x_2, \cdots, x_n$）爲一由0與1組成的數列，即所有的x_i不是0就是1，現設a_1，$a_2, \cdots, a_n$爲已知，$x_1, x_2, \cdots, x_n$爲未知，c爲一正整數，則方程式

$$c = \sum_{i=1}^{n} a_i x_i \qquad (2.2)$$

只有一解或無解。

證明 設此方程式有二解（$x_1, x_2, \cdots, x_n$）及（$x_1', x_2', \cdots, x_n'$）則消去c之後可得

$$\sum_{i=1}^{n} (x_i - x_i') a_i = 0$$

因 $$|x_i - x_i'| \leqslant 1$$

且 $$\sum_{i=1}^{n-1} a_i < a_n$$

故a_n之係數必爲0，即$x_n = x_n'$，因此

$$\sum_{i=1}^{n-1} (x_i - x_i') a_i = 0$$

同理可得

$$x_{n-1} = x_{n-1}', \cdots, x_1 = x_1'$$

此兩解原爲一。本定理得證。

而要解(2.2)是非常容易的事，因爲若

$$c > a_1 + a_2 + \cdots + a_{n-1}$$

則x_n必爲1，否則必爲0，同理若

$$c - x_n a_n > a_1 + a_2 + \cdots + a_{n-2}$$

則x_{n-1}必爲1，否則必爲0，以此類推，一下子就解出來了。

這幾個定理就足夠瞭解新的密碼法了。

三、狄飛、赫爾曼、麥克兒(Diffie, Hellmon, Merkle)法

此法是由上列三位電機工程師兼數學家（科學本是一家，觸類旁通，稱他們是什麼家並不正確，也不重要），有1976及1978年所發表的方法。在此方法中，所有的信號必須寫成二進位的形式，例如前面的我字代碼3314，若以二進位表示則爲$2^{11}+2^{10}+0\cdot 2^9+0\cdot 2^8+2^7+2^6+2^5+2^4+0\cdot 2^3+0\cdot 2^2+2^1+0\cdot 2^0$，即

110011110010

我們將任何一位信號分成許多段，每段含n個0與1的數，即（$x_1, x_2, \cdots, x_n$），以下是本密碼法的收發程序（參看圖1）

1. 由收報者密秘的取一組滿足定理2.4的正整數列（$a_1, a_2, \cdots, a_n$）及任一大於$\sum_{i=1}^{n} a_i$的質數m，及另一數w，令
   $$a_i^0 \equiv a_i w \pmod{m}$$
2. （公開）發出（$a_1^0, a_2^0, \cdots, a_n^0$）給發報者（敵方可以知道）
3. 發報者用$a_1^0, a_2^0, \cdots, a_n^0$與要發的0，1信號$x_1, x_2, \cdots, x_n$求出
   $$c^0 = \sum_{i=1}^{n} x_i a_i^0$$
   並將c^0之值告訴發報者（敵方可以知道c^0）
4. 收報者收到c^0之後，可以解出原信號$x_1, x_2, \cdots, x_n$（而敵方却不容易用已知的$a_1^0, a_2^0, \cdots, a_n^0$及$c^0$解出信號原因馬上會談到）。

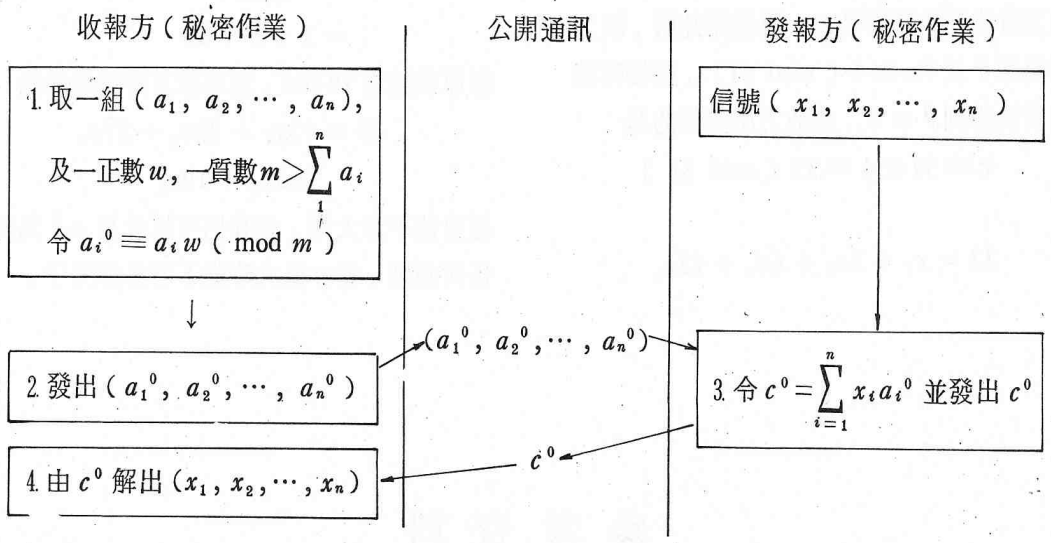

圖1　在此種操作中，敵方僅知（a_1^0，……，a_n^0）及c^0，而發方除了自己的信號外並不比敵方多知道什麼密碼的機密；只有發方完全掌握了解碼的方法。

我們先看收方如何解出x_i，由定理2.2之系可知存在$-\theta > 0$且$w\theta \equiv 1 \pmod m$若將收到之信號方程式兩邊乘以θ可使

$$c^0 = x_1 a_1^0 + x_2 a_2^0 + \cdots + x_n a_n^0 \quad (3.1)$$

變成

$$c^0\theta = x_1 a_1^0 \theta + x_2 a_2^0 \theta + \cdots\cdots + x_n a_n^0 \theta \quad (3.2)$$

但　$a_i^0\theta \equiv a_i w\theta \equiv a_i \pmod m$

故令

$$c^0\theta \equiv c \pmod m$$

(3.2)即成爲

$$c \equiv x_1 a_1 + x_2 a_2 + \cdots + x_n a_n \pmod m$$

因 $m > \sum_{i=1}^n a_i$，上式與$c = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$相同，根據定理2.4之說明，一下子就可以解出$x_1, x_2, \cdots, x_n$可是敵方在整個的過程中知道(3.1)之關係，因$a_i^0$並不見得是一個有$a_i$那樣規則的數列，到目前爲止只有硬試一途，即

$$(x_1, \cdots, x_n) = (1, 0, 0, \cdots, 0), (0, 1, 0, \cdots,)$$

等一個一個的試，對小的n這並不難，但對大的n而言，譬如說$n = 1000$（這並不是一個很長的信號），則平均要試$2^{1000-1} = 10^{301}$次才可以解開，這是一個天文數字，若電腦在一秒鐘內可以做一百萬次檢驗（目前尚辦不到），則一年只可以做3.15×10^{13}檢驗．那麼要解開(3.1)需要10^{288}年，前面談過地球的年齡不過5×10^9年而已，我們且舉一個簡單的例子。

例　令 $n = 5$，$(a_1, a_2, a_3, a_4, a_5) = (1, 3, 6, 12, 25)$，

$w = 13$，　$m = 51$

則收方發出的密碼法$(a_1^0, a_2^0, a_3^0, a_4^0, a_5^0)$分別爲

$$a_1^0 \equiv 13 \times 1 \equiv 13 \pmod{51}$$
$$a_2^0 \equiv 13 \times 3 \equiv 39 \pmod{51}$$
$$a_3^0 \equiv 13 \times 6 \equiv 27 \pmod{51}$$
$$a_4^0 \equiv 13 \times 12 \equiv 3 \pmod{51}$$
$$a_5^0 \equiv 13 \times 25 \equiv 19 \pmod{51}$$

這即是發報者收到的作密碼的a^0，假定發方所要發的情報是10101，則因

$$c^0 = 13 + 27 + 19 = 59$$

故收方所收到的密碼是59，要解開此碼，收方先得找到θ使$\theta w \equiv 1 \pmod{51}$。用輾轉相除法很快得到$\theta = 4$，故收方所要解的是

$$c \equiv 59 \times 4 \equiv 32 \pmod{51}$$

即

$$32 = x_1 + 3x_3 + 6x_3 + 12x_4 + 25x_5$$
$$= 1 + 6 + 25$$

即原信號爲10101，至於敵方所要解的是

$$59 = 13x_1 + 39x_2 + 27x_3 + 3x_4 + 19x_5$$

這自然不算太難，但我們可以看見a_i^0失去了任何規則，當n很大時就不容易解開了。

參考資料

1. Rivest, R. L; Shcmir, A.; Adleman, L. "A method for obtaining digital signature and public-key cryptasystem" *Communication of ACM*, 1978, pp. 120-126.

2. Simmons, G. "Cryptology : The mathematics of sewre communication" *The Mathematical Intelligencer*, 1978, pp. 233-246.

3. Hellman, M. E. "The mathematics of public-key cryptography" *Scientific American*, August, 1979, pp. 146-157.

4. Pomerance, C. "The search for prime numbers" *Scientific American*, December, 1982, pp. 136-147.

本文作者：

楊重駿：現任職於美國海軍研究實驗所

楊照崑：現任教於美國佛羅里達大學統計系