



# 中央研究院 週報

中央研究院 發行 73 年 11 月 01 日創刊 96 年 5 月 3 日出版 院內刊物 / 非賣品

第 1118 期

## 本院要聞

### 人事動態

天文及天文物理研究所籌備處李德昭奉核定為研究副技師，聘期自 96 年 4 月 25 日起。

應用科學研究中心湯朝暉奉核定為研究員，聘期自 96 年 4 月 26 日起。

### 人事獎懲

本院 95 年度選拔優秀技工、工友，業經評選小組於 95 年 11 月 13 日上午 9 時 30 分以票選方式選出 6 名優秀技工、工友，茲臚列獲選人員之優良事蹟如下：

- 一、應用科學研究中心工友王存仁。平日主動配合公務至深夜，並於假日主動加班，另負責有害廢物分類及資源回收，成效卓著，並負責該中心無塵室新建工程，認真監工，致進度超前，工作態度負責盡職，足資獎勵。
- 二、歷史語言研究所工友廖文勝。平日負責學術研究大樓外部環境之維護，提供研究員、行政人員乾淨舒適的空間，並將清掃之樹葉、草根等有機物製成植物營養所需之堆肥，使垃圾減量，工作態度任勞任怨，值得嘉勉。
- 三、歐美研究所技工黃啟祥。94 年度「研究大樓冰水主機汰換工程」委託規劃時，負責廠商進駐之各項監督和安全防護，使工程順利竣工，功績卓著；平日維護所屬大樓機電等相關設備，負責盡職，足資獎勵。

### 四、總務組技工李春村。

平日維護院區水電設備、協助國際研究生宿舍新建工程驗收，提升本院員工辦公及研究生住宿品



質；颱風季節，預先將發電機等救災設備，起動試運及備足油料，留守搶救機電設備，將災害減至最低。並於 95 年 9 月 23 日台電中研變電站發生設備爆炸火警意外事故，第一時間協助院內人員辦理應變措施。工作負責盡職，值得嘉勉。

### 五、總務組技工黃高宗。91 年及 94 年期間分別擔任專用駕駛勤務，未曾延誤首長及副首長之行程，提供安全舒適車況，細心負責，堅守崗位，值得嘉勉。

## 學術活動

### 學術交流

生物多樣性研究中心研究員兼代理主任邵廣昭，於 96 年 5 月 9 日至 13 日赴美出席國際會議。出國期間，中心業務由副主任吳俊宗代理。

### 小啟

本刊第 1117 期公布欄「歷史文物陳列館重新開館」，96 年 5 月 2 日(週三)重新對外開放，誤植為「5 月 5 日」，特此更正並致歉意。

## 本期要目

- |        |        |
|--------|--------|
| 1 本院要聞 | 1 學術活動 |
| 2 公布欄  | 3 知識天地 |
| 6 學術演講 |        |

編輯委員：紀元文 楊大衍 楊淑美 廖弘源 羅紀璋  
排版：林曉真 袁于婷 德仲文化事業股份有限公司  
<http://www.sinica.edu.tw/as/weekly/index.html>, <http://newsletter.sinica.edu.tw/en>  
E-mail: wknews@gate.sinica.edu.tw  
地址：台北市 11529 南港區研究院路 2 段 128 號  
電話：2789-9872 · 2789-9488；傳真：2789-8708

《週報》為同仁溝通橋樑，如有意見或文章，歡迎惠賜中、英文稿。本報於每週四出刊，前一週的週三下午 5:00 為投稿截止時間，逾期稿件由本刊視版面彈性處理。投稿請儘可能使用 E-mail，或送總辦事處秘書組綜合科 3111 室。

## 泰國學術研討會

時間：96 年 5 月 11 日（週五）

地點：本院民族所大樓 2319 會議室

主辦單位：本院人文社會科學研究中心亞太區域研究專題中心

合辦單位：國立暨南國際大學東南亞研究所

相關網站：<http://www.sinica.edu.tw/~capas>

## 天文科普演講活動：「專家談天」系列

日期：96 年 5 月 12 日（週六）下午 2 時至 4 時

地點：台北市天文科學教育館 3 樓第 2 演講室

講題：X 行星—談海王星外的太陽系

講員：金光光博士／本院天文及天文物理研究所籌備處博士後研究

對象：一般民眾

報名網址：<http://www.tam.gov.tw/asp/signup/signup.asp> (限額 200 人)

## 公布欄

### 資訊所許鈞南實驗室榮獲第 1 屆 BioCreative 分子生物文獻探勘評比最佳成績

本院資訊科學研究所許鈞南實驗室與陽明大學生物資訊研究所鍾翊方實驗室研究生共同組成的研究團隊，在西班牙馬德里舉行的第 2 屆 BioCreative 分子生物文獻探勘評比挑戰中大放異采。他們以兩種不同的方法，擊敗來自美國賓州大學，國家生物科技資訊中心(NCBI)，英國劍橋大學，荷蘭、西班牙、德國、韓國、中國等世界名校的研究團隊，囊括了世界第 2 名與第 3 名。雖然第 1 名由美國 IBM Watson 研究中心獲得，但大會表示，前 3 名方法的差距在統計誤差範圍內，因此前 3 名的方法可以並列為最佳成績；另外，在調整抽樣偏離後，中研院—陽明團隊所發表的第一種方法，就由原來的第 2 名躍居為世界第一。

本次評比項目，是由大量的生物醫學論文中，以電腦系統自動擷取其中的基因與基因產物（如蛋白質）的名稱。此系統除了有助於生物醫學研究者搜尋相關的論文，也可以讓電腦由論文中大量擷取特定的分子生物現象（如蛋白質交互作用，生物反應路徑等），節省實際操作實驗驗證所需的大量人力物力與時間，因此各國均投入大量資源進行研發。這項評比的挑戰性在於基因名稱在論文中沒有統一的命名方式，而且容易與其他生物醫學名詞混淆。中研院—陽明團隊使用機器學習演算法，由一萬五千句標注好的訓練樣本，歸納出準確的統計模型。同時，許鈞南實驗室在提升機器學習演算法的效能還有許多重要的成果，將在近期內發表。

這項研究是由國科會基因體醫學國家型計畫中的「陽明—中研院進階生物資訊核心設施」所補助，歡迎全國各大生物醫學實驗室進行合作，以擴大成果效益。

### 本院附設幼稚園 96 學年度招生公告

報名時間：96 年 5 月 1 日至 15 日，每日上午 9 時至 12 時

報名地點：本院附設幼稚園辦公室

幼稚園電話：2783-5325，張桂蘭園長

詳情請見幼稚園網站：<http://www.sinica.edu.tw/~garden> 或各所網站。

# 知識天地

## 一些破密與加密最近的鬥爭

楊柏因（本院資訊科學研究所副研究員）

一般來說，資訊的安全防護被破解或是被侵入系統，很少是因為密碼學的理论本身造成的，絕大多數是人為操作或是程式、系統設計的錯誤被利用。即使如此，相信多數人仍有興趣知道最近學理上種種關於破解密碼的理论。

### 破密的理論與實際

所謂密碼系統，也就是密碼學所稱的基本操作[primitives]，都有設計者希望達成的安全性。這就是用純暴力的方法算出攻擊者不該有的東西—沒加密的原文，或是和真品具有相同效果的數位簽章等—其理論上需要的時間。例如說 DES 加密系統的密碼有 56 個位元，所以用蠻力硬試當然可以在加密  $2^{56}$  次的時間以內算完。以此為準，密碼學家所謂的破解成功，就只表示找到方法可以把所需要的時間壓低到設計者要求的時間以下。換句話說當設計者希望的安全性是  $2^{256}$  那所謂破解可能是  $2^{200}$ ， $2^{100}$ ，甚至  $2^{20}$ 。當然其實用性也差距很大。在密碼學研討會的論文發表，經常是這樣做結：「…所以可以用  $2^{70}$  個位元的記憶體[這或許需要把地殼中所有的矽拿來做 RAM]和  $2^{110}$  時間算出 [128 位元]密碼，也就是破解完成」—而此刻大家會鼓掌並承認他的確破解成功了。

### 最近的新型攻擊：代數破密(Algebraic Cryptanalysis)

密碼系統可分為對稱式(秘密金鑰)和非對稱式(公開金鑰)兩類。後者的特質是：可以在被壞人全程監聽時使用；顯然這很有用，但是同時它也極緩慢，不適合持續操作。因此，公開金鑰密碼系統應用於加解密的主要用途是傳送對稱式系統的當次金鑰(session key)，再以對稱式系統加解密。

對稱式密碼系統分為兩種：區塊式如 DES、AES，和串流式如 RC4。串流式的一般特徵是比較快，因此初期也廣為採用。串流式的加解密簡單的來說就是偽亂數生成器(pseudo-random number generator, PRNG)！我們知道，一個亂數生成器，例如大家熟悉的 rand()函數的理想狀態如下：給定不同的種子(seed)，即可輸出一個旁人無從預測，甚至無法和真正純隨機的數列區隔，但只需要擁有種子即可重現的「亂數」列。

假設我們有一串無限長的而純隨機的位元列[由 0 和 1 組成的序列]，而接收者和發送者已經各預先備妥一份，那麼理論上只需要把要交換的明文也表達成一串位元，並且和這串位元列—我們稱之為單次使用密碼本(one-time pad)—來作一個位元一個位元的邏輯上 XOR 操作，即成為理論上無懈可擊的密碼，所以如果我們設想偽亂數生成器的輸出串列—我們稱之為密碼串列—真的和純隨機的位元串列無法區隔，那麼一切就很完美了。

很不幸，事實上不可能如此，這類的密碼系統必須從種子—密碼學家通稱為金鑰(key)—開始，造出一個初始狀態(initial state)，此後藉由維持一些狀態資料(state)而產生一個偽亂數串，以代替真正的單次使用密碼本。

在程式語言中的 rand()以至於統計學家實作模擬的偽亂數產生器，基本上都根據某個線性遞迴式加上一個輸出函數(output filter) 而來，例如 GNU libc 程式庫中的線性 rand()的輸出，是先從種子  $r_0 = s$  製造出初始狀態：

$$r_i = (16807 * (\text{signed int}) r_{i-1}) \bmod (2^{31} - 1) \quad [i = 1 \dots 30], \quad r_i = r_{i-31} \quad [i = 31 \dots 33]$$

然後進行線性遞迴式  $r_i = (r_{i-3} + r_{i-31}) \bmod 2^{32}$  (for  $i \geq 34$ )，捨棄其前 343 個數字，最後使用簡單的輸出函數  $o_i = r_{i+344} \gg 1$  ( $\gg 1$  代表向右移動，即除以二)。所有您可能在寫 C/C++時簡單呼叫到的亂數程式都具有同一形狀。

作為統計上的工具時，偽亂數列有很多學理上必須滿足的條件，各位可以參考 Don Knuth 聖經(*The Art of Computer Programming*)的第二冊；事實上做模擬，有不少簡單的偽亂數產生器[例如有名的 Mersenne Twister，松本真和西村拓士所提出，[http://en.wikipedia.org/wiki/Mersenne\\_twister](http://en.wikipedia.org/wiki/Mersenne_twister)]就足夠好了，但作為加解密系統的偽亂數列還需要更好的性質。早在數十年前，數學家就發現了所謂 Berlekamp-Massey 演算法：任何長度為  $n$  的線性遞迴式可以

由他的前  $2n$  項唯一且迅速的決定，因此輸出函數必須是非線性的；在 1976 年 Rotenhaus 的引理指出根據線性遞迴系統與某些型式的輸出函數—所謂彎曲的(bent)的函數—所造出來的偽亂數列不容易用線性的方法和真正隨機的數列區隔。有很長的時間，密碼學家認為只要選擇適當的線性遞迴式—硬體上稱為線性回饋的遞移暫存器(LFSR)—和輸出函數，就保證安全。

事情事實上並不這麼簡單，從 2002 年開始，幾乎所有這一類型的密碼系統被破得乾乾淨淨，我們舉一個有點誇張但是事實上發生的例子來說明為什麼會發生。一個稱之為 Toyocrypt 的加解密系統幾乎成為日本的國家標準，以取代老牌的 RC4 系統(前仍用於某些無線網路的加密)。Toyocrypt 使用一個多達 128 個單位元變數的 LFSR，加如下的輸出過濾函數(filter function)：

$$f(s_0, \dots, s_{127}) = s_{127} + \sum_{i=0}^{62} s_i s_{\alpha_i} + s_{10} s_{23} s_{32} s_{42} + s_{1} s_{2} s_{9} s_{12} s_{18} s_{20} s_{23} s_{25} s_{26} s_{28} s_{33} s_{38} s_{41} s_{42} s_{51} s_{53} s_{59} + \prod_{i=0}^{62} s_i,$$

其中  $\alpha_0, \dots, \alpha_{62}$  是從  $\{63, \dots, 125\}$  選出來的排列。Toyocrypt 每被呼叫一次，LFSR 就變換一次由 128 個一位元的變量  $(s_0, \dots, s_{127})$  構成的狀態，同時算出  $f$  作為輸出串列的該位元。剛剛提到的引理說：這樣的密碼很難用傳統攻擊去破密。很不幸的，問題同時也出在這個輸出函數上，因為我們可以把式子的下半段全部刪掉，成立的機率應該還是很接近 1；事實上，差不多是  $1 - 2^{-17}$ 。

有了近似成立的方程式，就可以求解了，我們只要蒐集足夠多的資料，就可以把 key 和 state 當成變數，把方程組一次解出來。在 Toyocrypt 的例子，因為方程式不成立的機率  $2^{-17}$  太小了，以至於我們可以蒐集  $10^5$  筆資料，仍然可以很有信心地說方程組會成立。如果運氣不好，只需要多算一兩次。

使用代數手段來破解密碼系統的方法，相對於兩千年以前的密碼學家以統計作為主要手段的古典攻擊，被統稱為代數攻擊，其中常常會以解方程式作為最後一步。數百個變數和方程式，這樣的方程組還能解嘛？真是不幸，目前的解方程式技巧—特別是當變數僅為 0 或 1 時—有長足的進步。有興趣的讀者可以看三篇有關 XL 演算法的論文 (<http://precision.moscito.org/by-publ>)。

因此，即使製作人認為這個密碼的安全性是  $>2^{80}$  時間[單位是進行一區塊 3DES 密碼系統加密所需的時間，這是密碼學家常用的一個標準]，審查的專家也同意，但是卻可以用  $2^{60}$  以下時間完成破密。

這告訴我們什麼？很明顯我們應該避免採用這一類型的密碼，可惜有一個已經被標準化了，就是藍芽無線傳輸設備中所使用的密碼系統 E0，所以如果您使用藍芽鍵盤或是耳機，一個持有特製天線的壞人可以在百公尺以外截聽到所有被傳遞的資訊，然後很快的解碼出所有傳遞的資料。如耳機拿來聽聽歌當然沒關係，但是如果可能用來輸送機密，例如鍵盤，那就有點糟糕了。

我們也必須小心的設計將來的系統以免重蹈覆轍。今天的系統設計人很可能會採用區塊式密碼系統：美國國家標準局從 1997 年開始徵求新一代的加密標準，並在 2001 年底開始正式採用應徵者之一為先進加密系統[AES，可見 [http://en.wikipedia.org/wiki/Advanced\\_Encryption\\_Standard\\_process](http://en.wikipedia.org/wiki/Advanced_Encryption_Standard_process)]。在十年的努力研究之後，大家目前仍然相信它在可預見的將來是可靠的。

### 我們的研究：XL 解方程組的分析

XL 可以說是 Gröbner 基底法的簡化版—原理比較單純，程式的撰寫也比較簡單。假設在基體  $K$  上，我們有  $n$  個變數  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  的  $m$  個二次（更高次也類似）方程式  $l_1(\mathbf{x}) = l_2(\mathbf{x}) = \dots = l_m(\mathbf{x}) = 0$  且  $n \leq m$ ，那麼  $D$  次的 XL 操作如下：

- i. 第一步驟是 X, X 表示乘法, 或是延伸(eXtended)。我們將每個  $D - 2$  或更低次的單項  $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$  (通常記作  $\mathbf{x}^a$ , 其中  $\sum_i \alpha_i \leq D - 2$ ) 乘以所有方程式  $l_j(\mathbf{x}) = 0$ 。
- ii. 第二步驟是 L, L 表示線性化(linearize)。我們把每個  $D$  次或更低的單項當成獨立的變數, 並把這樣求出的  $m \binom{n+D-2}{n}$  條方程式(當  $q = |K| \leq D$  時會變少)做高斯消去法。而變數的排序, 也就是各單項式被消去的順序, 採用相反字典式(reverse lexicographical order)。那麼最後被消去的, 是第一個變數  $x_1$  的單項。
- iii. 如果獨立的方程式至少有  $\binom{n+D}{n} - D$  條(這個下限在  $|K| \leq D$  時也會變少), 我們應該已經得到僅含  $x_1$  單項式的一條以上方程式, 消去  $x_1$  以外的任何變數。接著可以採用 Berlekamp 演算法求出  $x_1$ 。如果還有沒做完的變數, 那麼重覆上面的動作, 得到每一個變數。

這裡理論上的分析非常複雜, 需要應用到組合學和複變函數的技巧, 我們不多提它, 只簡單告訴各位讀者, XL 在理論上不如更複雜的演算法, 但是使用它, 我們可以應用疏落[sparse]矩陣的技巧—這樣做出來的係數矩陣總是絕大多數為零—進而以更少的時間和資源求出解。我們發現, 解方程式的時間仍是指數成長, 但是成長率比原來想像中的要慢許多。當數字  $n$  大概幾十的時候,  $n$  個方程式解  $n$  個 GF(256)中的變數只需要大約一個  $n$  的多項式乘上  $2^{2 \cdot 4n}$  的時間!

### 「理論上安全」與 QUAD(256, 20, 20)的分析和破解

由於這些破密的進展, 最近的密碼學界開始探討所謂理論上的安全性, 最近的學界有很多文章如此: 假設 ZZ 問題不可解, 如果密碼系統 AA 可以破解, 那麼把這個破解的過程當成黑盒子來使用, 我們就可以證明一連串定理說我們可以做到 BB, CC...最後做到 ZZ-足見 AA 是安全的!

我們最近在這方面的研究指出, 使用這一類的定理和前面第三節的爭端一樣, 需要注意細節, 因為沒有什麼問題真的是不可解的, 頂多只是很難解而已; 如果 ZZ 問題可以經過兩次的破解 AA 系統的過程而解出, 同時我們認為要解決 ZZ 問題至少需要  $2^{100}$  時間, 那麼破解 AA 系統就至少需要  $2^{99}$  時間。這樣的證明大致沒有問題, 但是如果定理的證明中 ZZ 問題需要經過  $2^{40}$  次破解 AA 系統呢? 那我們只證明了這個破解 AA 系統需要  $2^{60}$  時間。這對於現代密碼學家來說是不夠的!

在去年的歐洲密碼學大會中, 知名的法國密碼學家 Gilbert 和他的兩位學生提出如下的串流式密碼系統: 給定初始狀態  $\mathbf{x}_0$  是  $n$  個  $K=GF(q)$  的元素所組成, 更新規則  $\mathbf{x}_{i+1}=\mathbf{Q}(\mathbf{x}_i)$ , 每次的輸出  $\mathbf{y}_i=\mathbf{P}(\mathbf{x}_i)$ ,  $\mathbf{P}, \mathbf{Q}$  都是隨機選出的  $K^n$  到  $K^n$  的函數, 他們提出一些定理認為解多元二次方程式組的困難度可以保證這個密碼系統的安全性。而  $q^n = 2^{160}$  時, 可以得到相當快而且安全的效果。很不幸, 由於前面提到的盲點, 我們發現, 只要解

$$\mathbf{y}_0=\mathbf{Q}(\mathbf{x}_0), \mathbf{y}_1=\mathbf{Q}(\mathbf{P}(\mathbf{x}_0))$$

的方程組那麼在  $q=256, n=20$  時, 這二次和四次的混合方程組可以用 XL 在  $2^{60}$  乘法所需的時間以內計算出來, 這個研究發表在最近於盧森堡舉行的快速軟體加密研討會(FSE 2007), 並且獲得很好的迴響。目前對於尋求更安全的參數組合和更好的破密方式的後續研究仍在各地進行當中。

### 小結

這裡所提供的或許只適合在茶餘飯後提供一些樂趣。但是或許足以提醒您密碼學的理论真的是日新月異, 隨時都可能會有新的發展出現。希望聰明如您們能夠領會到資訊安全需要時時掛在心上的重要性。

## 學術演講

日期	時間	地點	主持人	講員	講題
<b>數 理 科 學 組</b>					
5/4 (五)	14:00	天文所籌備處 會議室		Dr. Ananda Hota (天文所籌備處博士後)	Radio Continuum and HI Study of Gas-Loss Processes in Nearby Galaxies
5/7 (一)	10:30	統計所 2 樓交誼廳	蔡恆修博士	胡毓彬教授 (暨南國際大學)	Identifying the Conditionally Heteroskedastic Factor Model
	11:00	資訊所 新館 106 演講廳		劉必治院士 (Princeton Univ. USA)	Multimedia Watermarking
5/8 (二)	14:00	物理所 1 樓演講廳	王子敬博士	倪維斗教授 (清華大學)	精密測量與真空雙色性
		人文館南棟 應科中心 1101 演講室	張亞中主任 張建成博士	蔡振水教授 (U.C. Irvine, USA)	高頻超音波噴嘴及其生醫領域之應用
5/9 (三)	14:00	人文館南棟 11 樓 環變中心演講廳	黃天福博士	Prof. Maria Lourdes San Diego-McGlone (Univ. of the Philippines)	Marine Scientific Research in the Philippines
	16:00			Prof. Gil Jacinto (Univ. of the Philippines)	Nearshore and Offshore Oceanography in the Philippines
5/10 (四)	14:00	地球所 3 樓演講廳	江博明所長	張儒媛教授 (Stanford Univ., UAS)	Ultrahigh-Pressure Metamorphic Belts in China: Major Progresses in the Last Several Years
	15:30	化學所 A108 會議室	許昭萍博士	陳平教授 (The State Univ. of New York at Buffalo, USA)	The Effect of Highly Fluorinated Amino Acids on Protein Secondary Structure Stability
			原分所 浦大邦講堂		羅履維博士 (國家衛生研究院)
5/11 (五)	10:30	物理所 舊大樓 332 會議室	蘇維彬博士	Prof. Philip LeDuc (Carnegie Mellon Univ., USA)	Nano-/Micro-Technology for Linking Mechanics to Biochemistry in Molecular and Cellular Behavior
<b>生 命 科 學 組</b>					
5/8 (二)	11:00	生醫所 地下室 B1B 演講廳	杜邦憲博士	黃慶裕博士後 (Washington Univ., USA)	Functions of Ataxia Telangiectasia Mutated (ATM) Protein in V(D) J Rearrangement and Thymocyte Development
5/9 (三)	10:30	生化所 114 室		Dr. Yun-Jin Jiang (Agency for Science, Technology and Research, Singapore)	Functions of Jagged-Notch Signaling in Zebrafish Pronephros Development
<b>人 文 及 社 會 科 學 組</b>					
5/7 (一)	10:00	史語所 文物陳列館 5 樓會議室		王汎森特聘研究員 兼所長(史語所)	清代儒者的全神堂—《國史儒林傳》與 嘉、道年間的顧炎武崇拜
5/8 (二)	15:00	經濟所 B 棟 110 室		張瑞娟教授 (暨南國際大學)	Investigating the Credit Supply-versus-Demand Puzzle in a Small Open Economy
5/10 (四)	10:00	近史所檔案館 1 樓 中型會議室		陳儀深副研究員 (近史所)	1970 年刺蔣事件 —以外交部檔案為主要的研究