加密文件之搜尋系統

鄭景榮
南台科技大學資訊工程系
碩士生
e-mail：
m94g0207@webmail.stut.edu.tw

鄭錦楸
南台科技大學資訊工程系
講師
e-mail：
chiouu@mail.stut.edu.tw

郭文中
虎尾科技大學資訊工程系
助理教授
e-mail：
simonkuo@sunws.nfu.edu.tw

摘要

由於網路之普及，以及儲存裝置之越來越低廉，為了某些商業利益，有為數不少之網站伺服器提供免費儲存空間給使用者利用，因此將資料上傳至伺服器端儲存的機會也越來越多。但是如果欲搜尋檔案為已加密檔案時，則搜尋時往往將浪費很多時間在這些檔案之解密上。目前已發表之文獻中，有關搜尋已加密之文件系統－通常其加密演算法皆為區塊加密法(Block Cipher)，此種加密法並不適用於密文搜尋，因為欲搜尋之關鍵詞有可能於原明文文件被加密時，被切割破壞，造成搜尋不到。本論文提出當檔案在加密之狀態下如何進行搜尋之方法，以期達到快速搜尋之目的。文中提出適用於密文搜尋之加密文件系統－串流加密法(Stream Cipher)。有關串流加密之安全性，在本文中做了補強與改進，並給予安全性之證明。

關鍵詞：區塊加密法、串流加密法、加密搜尋

Abstract

Due to the quick development of the internet and low of prices in storage, there are numerous websites to provide users with free space of storage for some commercial purpose. Most users upload files with encryption type. This will induce an interest issue: how to search for a keyword on the encrypted file.

Up to now, there are no practicable schemes to perform searching on the encrypted file. They all meet the same difficult problem that the keywords searched have been usually destroyed while the files are encrypted. The troublesome problem comes from the employ of the block cipher. In this paper, to overcome the difficulty, we adopt the stream cipher to implement such a searching system. Some strategy would be raised for efficiency of search. We also enhance the security of the stream cipher by introducing the permutation. Finally, we show the proof of security about our proposed stream cipher by means of probability.

Keywords: Block Cipher, Stream Cipher, Search Encrypted

1. 前言

目前網際網路已經是十分普及化，網路業者大多提出許許多多免費或是付費之網站空間，而且這些空間越來越大，使得使用者可以將一些可用之資料存放於網路業者端，但是由於將資料存放於伺服器端且未加密之狀態是

十分不安全的，所以使用者將資料上傳至伺服器端時候可先經由加密後再將檔案上傳至伺服器端，然而由於上傳時間越久，上傳資料越多，使用者極有可能忘記當初上傳之資料是放置於哪個文件裡，如果用以往的方法，則必須先將上傳所有資料下載後，全部資料都進行解密後，再做一次搜尋，這樣將會浪費太多時間。另外，現在行動裝置（如手機，PDA…等）越來越普及，也已經與網路連接上軌道，如果使用者個人電腦上傳文件，而欲在行動裝置上下載資料，由於行動裝置儲存空間太小，所以會發生資料無法全部下載的結果。本論文提出可以適用於加密文件上之搜尋系統，來快速找出欲搜尋之文件。

為了解決以上所描述之問題，許多文獻提到其相關之研究[1][2][3][4][5]。在 Song 等人所提出之方法中[6]，使用區塊加密方法進行加密，而在搜尋時將會遇到如果搜尋之加密文字被切塊或者長度不一情況下，將無法正確搜尋出正確文件。而在 Brinkman 等人[7]中，將 Song 等人引用加以改良，但還是局限於區塊加密法，而無法突破搜尋之加密文字被切割分離後所產生之情況。而還有 Chang, Mitzenmacher[8]提出另兩種搜尋法，但還是無法實做其方法，所以本論文提出使用串流加密法來進行加密，以達到可以實現加密資料快速搜尋之目的。

下列為本篇論文之結構，在第二段提出串流加密法的加密方法，也提出我們改善其搜尋相關之安全性。在第三段提出本搜尋法與使用暴力搜尋法之情況，也提出暴力搜尋法相關的使用問題。而第四段提出本論文所強化之安全性。第五段提出本論文其實驗結果，最後一段為結論。

2. 串流加密法

串流加密法[9][10][11] (Stream Cipher) 主要是將明文字串跟由金匙當 seed 經過亂數產生器產生的擬亂序列做 XOR 以產生出密文，而解密反之。

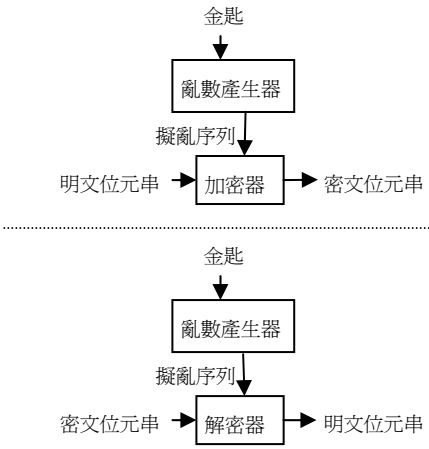

圖 1：串流加密系統方塊圖

2.1 改善之加密法

在本論文中，我們針對串流加密法 (Stream Cipher) 之加密器進行了一些細微之改變。我們將明文M輸入後與其 BBS 產生之 key 做 XOR，得到 C_1，再由 C_1 將其打散得 C_2，而打散之順序則為使用者自行設定。然而最後再由 C_2 經過第二次與 BBS 產生之 key 進行 XOR 得其 C_3。這樣來強化其被明文攻擊後，將馬上破解之可能性，如下圖 2。

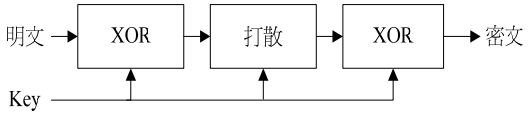

圖 2：強化後加密器流程圖

2.2 亂數產生器

串流加密法其安全性則建立在亂數產生器所產生的擬亂序列要夠亂，因此亂數產生器之安全性也將列入考慮，所以本論文將使用 BBS(Blum Blum Shub)亂數產生器，以提高擬亂序列之亂數與安全性。

選擇 BBS 當亂數產生器，主要是由於 BBS

所產生的亂數週期十分大，可以大大提高其安全性。

本論文將使用 BBS 之初始值作為輸入之 key。

BBS 的演算法：

a. 選定兩個 p 和 q 為質數 ，這兩個質數必須滿足除以 4 後都要餘 3。

b. 選一個亂數 s，s 必須與 n 互值，也就是說 p 和 q 都不是 s 的因數。

c. 根據上面的條件加上下面的演算法來產生一連串的亂數位元 B_i：

$$X_0 = s^2 \bmod n$$
$$for \quad i = 1 \quad to \quad \infty$$
$$X_i = \left(X_{i-1}\right)^2 \bmod n$$
$$B_i = X_i \bmod 2$$

每一次執行的結果都會產生一個最低位元(LSB)，所執行的次數不得超過記憶體(Memory)容量。亂數 S 即為金匙。

3. 搜尋法

假設明文為 abcdefgh，而 key 當 seed 輸入至 BBS 後，產生之亂數輸出為 123，兩者經由串流加密後結果假設為 a1，b2，c3，d1，e2，f3，g1，h2。如圖 3 所示。

我們由此來分析暴力搜尋法與我們之搜尋法最大差別處。

3.1 暴力搜尋法

暴力搜尋法必須以每一 byte 之 keyword 與每一 byte 之 key 做 xor 來產生出每一組有可能會產生之密文，利用如此來針對每一種可能性做搜尋。

例如欲搜尋圖 3 中 b 此 keyword，暴力搜尋法必須將 b 與所有 key 的每一種 byte 做 xor 來產生出每種可能之狀況，再針對每種可能性來對所有之密文來做搜尋，也就是將 b 先對 key 之第一個 byte－1 做 xor 產生出 b1，再以 b1 對所有密文做搜尋，而產生搜尋失敗之結果，再以 b 對 key 之第二個 byte－2 做 xor 產生出 b2，再以 b2 對所有密文做搜尋，而因為密文中有 b2 此 byte，所以產生搜尋成功之狀態。

也因為必須將每一種所有可能發生之狀況做搜尋，所以也導致浪費太多時間在搜尋上，所以針對這情況提出本方法來改善。

3.2 本搜尋法

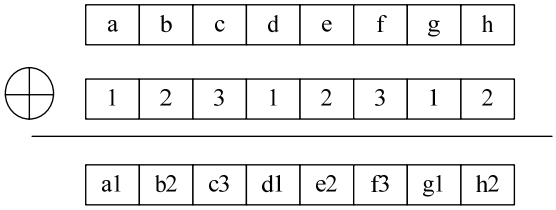

圖 3：經由串流加密法產生之結果

以串流加密法做為加密方法，而使用暴力搜尋法當做搜尋依據，則會浪費太多時間於搜尋上，將喪失此加密搜尋系統之原意(加快搜尋所需檔案之時間)，所以本搜尋法將取代暴力搜尋法來進行搜尋，來促使搜尋速度之增加。

本搜尋法是採用對應式搜尋，也就是依照 key 所對應的位置來做相關位置之搜尋，這樣可以大幅度減少搜尋其他無相關之位置，所以能加快搜尋速度，也可以減少誤判之效果。

例如欲搜尋圖 3 中 b 此 keyword，本搜尋法首先將 b 與所有 key 之每一個 byte 做 xor 產生出每種可能發生之狀況，再針對每種可能性會發生於哪些 byte 所在位置來進行搜尋，也就是將 b 先對 key 之第一個 byte－1 做 xor 產生出 b1，再以 b1 對所有 key 之第一個 byte 會產生加密之位置做搜尋。也就是指針對 a1、d1、g1 此三個 byte 位置做搜尋，而搜尋失敗後，再以 b 對 key 之第二個 byte－2 做 xor 產生出 b2，再針對 key 第二個 byte 產生加密現

之位置 b2、e2、h2 此三處做搜尋，產生出搜尋成功。這樣可以節省其他不可能發生之位置卻做無必要之搜尋的缺點，也節省非常多之無謂時間，因此達到快速搜尋的效果。

3.3 暴力搜尋法可能會產生錯誤之搜尋

因為暴力搜尋法是以逐一比對之方法做搜尋，所以也導致出無法避免之錯誤。因為當搜尋文字與 key 做 xor 時候會產生出一組新的數值，也因為這組新數值是由 keyword 與 key 做 xor 產生，所以會發生有一定機率會與明文與 key 之另外一組數值發生一模一樣之狀況，所以如發生此狀況，則會產生搜尋錯誤之結果。

例如 b 之 ASCII 碼為 98，2 之 ASCII 碼為 50，而將 b 與 2 做 XOR 會產生出 128 之結果。而 c 之 ASCII 碼為 99，3 之 ASCII 碼為 51，而將 C 與 3 做 XOR 會產生出 128 之結果，可是 b 與 2 做 XOR 會產生 c 與 3 做 XOR 一模一樣之結果，所以如使用暴力搜尋法會產生誤判之問題。

例如明文為 accdefg，而 key 做亂數產生器後結果為 123，兩者經由串流加密後結果假設為 a1，c2，c3，d1，e2，f3，g1，h2。如圖 4 所示。

a	c	c	d	e	f	g	h
⊕							
1	2	3	1	2	3	1	2
a1	c2	c3	d1	e2	f3	g1	h2

圖 4：經由串流加密法產生之結果－2

我們欲搜尋 b 此 keyword，而我們以暴力搜尋法來搜尋時，我們在做 b2 搜尋時候，會搜尋到 c3 之位置，而因為 b2 與 c3 數值相同，所以就會產生誤判之結果。

而因為我們所提出之搜尋演算法是針對 key 之 byte 所對應之位置做搜尋而已，就會跳過所有可能會一樣數值之狀況，也就是只針對 c2、e2、h2 做搜尋，所以提升產生搜尋正確率與速度。

4. 安全性

由於本篇方法是採用串流加密法(Stream Cipher)為主要加密方法，搜尋時候是將欲搜尋之 keyword 加密後傳至伺服器端來進行搜尋，因為是加密狀態，所以安全性高。

但是為了抵擋選擇明文攻擊 (chosen plaintext attack)，在我們的串流加密系統中，額外加入一 permutation，此 permutation 的方式亦為受到金鑰的控制，亦即不同金鑰輸入會有不同的 permutation 方式，藉此以提升串流加密系統的安全性。下列即為本加密文件搜尋系統採用的加密方法：

$$c = \{[Perm_{key}(m \oplus key)] \oplus key\}$$

其中 $|m| = l, |key| = l$，且 $Perm_{key} : X \to Y$,

$X \in \{0,1\}^l, Y \in \{0,1\}^l$ 為一受金鑰 key 控制的 permutation。密文 c 的產生，是先與由 BBS 輸出的 key 進行互斥或的計算，接著進行由 key 控制之 permutation，最後與 BBS 輸出的 key 再進行一次互斥或計算，其結果即為密文 c。

因為 $|m| = l, |key| = l$，故 $Perm(m \oplus key)$ 的可能結果共有 $2^l!$，亦即可控制此 $Perm(\cdot)$ 的 key 的可能數目為 $2^l!$。現假設選定一固定之金鑰 key，亦即 $Perm(\cdot)$ 之 permutation 方式固定，我們可得以下兩項結論

(1) 對於任意固定之 $X, Y \in \{0,1\}^l$ 而言，則 $\Pr[Perm_{key}(X) = Y] = 1/2^l$。

(2) 對於任意固定 $X_1, X_2, Y_1, Y_2 \in \{0,1\}^l$ 且 $X_1 \neq X_2$ 而言，則

$$\Pr[Perm_{key}(X_1) = Y_1 \mid Perm_{key}(X_2) = Y_2] = \begin{cases} 1/(2^l - 1), & \text{if } Y_1 \neq Y_2 \\ 0, & \text{if } Y_1 = Y_2 \end{cases}$$

若選擇足夠大 l 時，則上式之

$1/(2^l-1) \approx 1/2^l$，故 l 足夠大且 $2^l >>$ 選擇明文數量，則要由利用選擇明文，來猜對 $Perm_{key}(m \oplus key)$ 的機率幾乎等於 $1/2^l$。因此本系統所提之改良式的串流加密器是安全的，也很適合用於密文的搜尋上。

由於我們針對串流加密法之加密器有做其安全性之改良，藉由 $Perm_{key}$ 強化了串流加密之安全性，使得其搜尋系統將更加安全。

5. 實驗數據

由於本論文提出之方法有別於使用暴力搜尋法來進行搜尋，所以我們將與暴力搜尋法做比較，證明本方法可有效的提升搜尋上時間之節省。

我們一開始先針對 key 長度不變之情況做討論，假設使用 20bytes，而改變其明文長度，來探討本方法與暴力搜尋法兩者之時間搜尋的變化，由圖 5 來看，可以發現本搜尋法變化量不大，而暴力搜尋法卻隨著明文長度以倍數成長，故使用暴力搜尋法，如果原文長度越長，搜尋使用時間將十分可觀。

接著我們探討原文長度固定不變之情況，假設其長度為 20Kbytes，而改變其 key 長度，來探討其搜尋時間之變化量。由圖 6 發現，本搜尋法搜尋次數為幾乎固定，未有任何太大變化，故推出本方法是對原文長度有改變時，其搜尋時間只會稍受影響，而如果單純只是 key 長度改變，則搜尋時間未有任何變化，但使用暴力搜尋法之結果，卻發現變化量卻十分之大，遠大於改變明文長度但不改變 key 長度之情況之變化量，所以可以得知暴力搜尋法其搜尋時間的長短，所到 key 之長度的影響甚大。

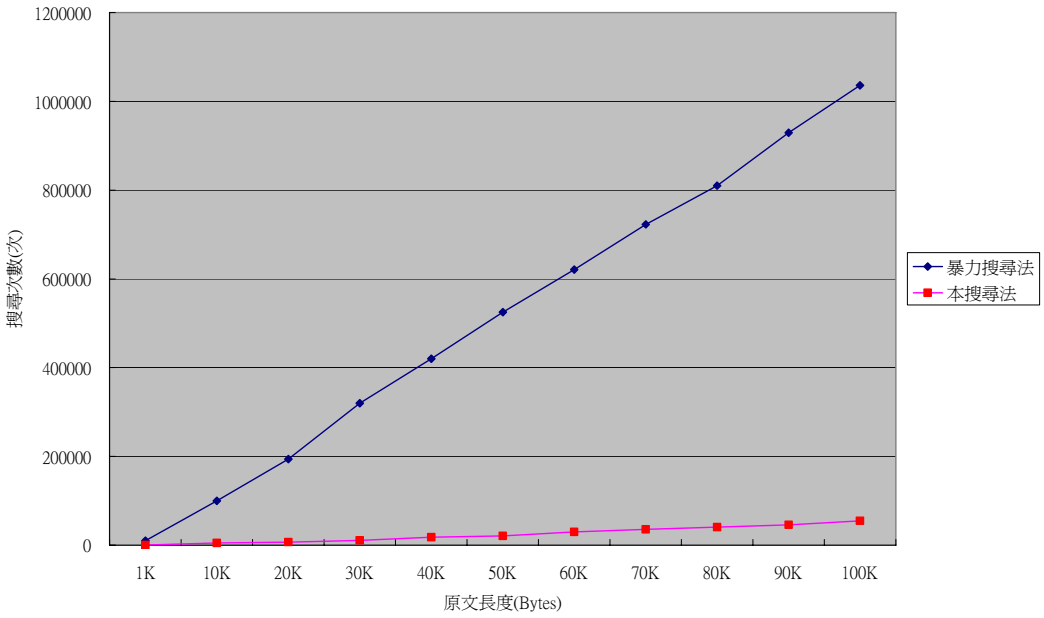

圖 5：key 長度不變，改變原文長度之比較

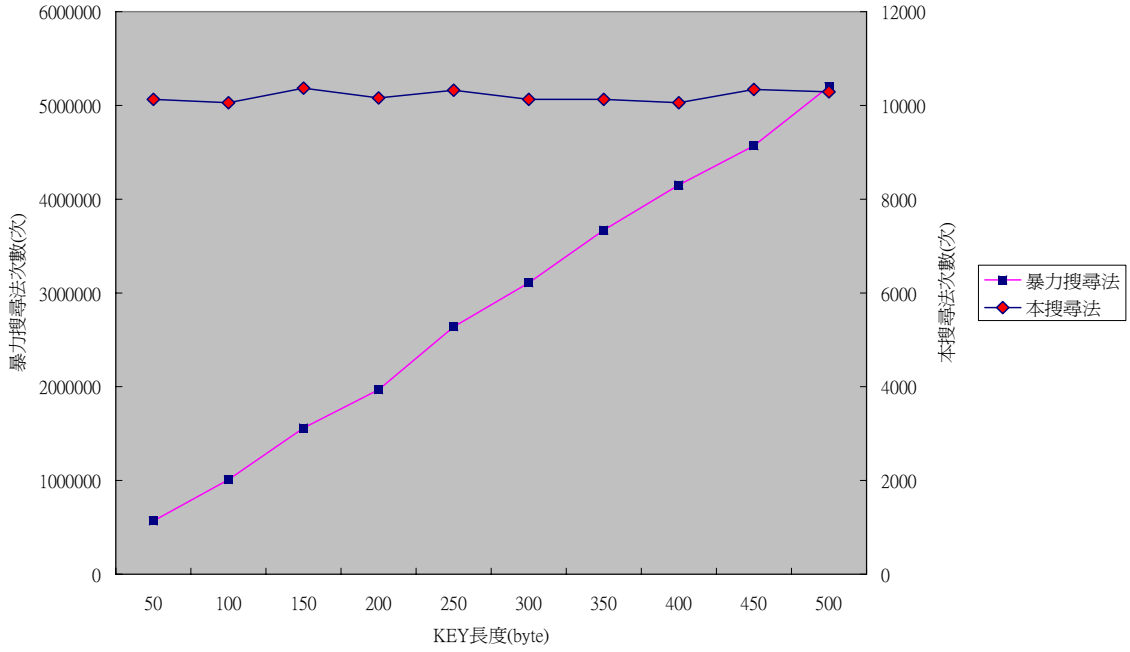

圖 6：原文長度不變，改變 key 長度之比較

6．結論

本篇論文提出與其他論文所未曾提出之方法，主要以串流加密法(Stream Cipher)為基礎，來進行做加密，而使用此方法來進行搜尋加密文件，進而突破在區塊加密上，無法突破之困難處：如果欲搜尋之 keyword 被區塊加密法切割後，進而無法搜尋成功之問題，亦即如果用區塊加密法進行加密文件時，則有可能發生搜尋之 keyword 被切斷之問題，造成無法正常搜尋。另外本方法也加強了明文攻擊之困難度，來增加本篇論文之價值。

參考文獻

[1] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key Encryption with Keyword Search”, ***EUROCRYPT’04***, 2004.

[2] D. Park, K. Kim, and P. Lee, “Public key Encryption with Conjunctive Field Keyword Search”, **WISA’04, LNCS 3325**, pp73-86, 2004.

[3] P. Golle, J. Staddon, and B. Waters, “Secure Conjunctive keyword search over encrytped data”, ***In Proceedings of ACNS ’04, LNCS Vol. 3089***, pp. 31-45, 2004.

[4] H. S. Rhee, I. R. Jeong, J. W. Byun, D. H. Lee ”Difference Set Attacks on Conjunctive Keyword Search Schemes” ***Secure Data Management*** 2006

[5] J. W. Byun, H. S. Rhee, H.A. Park, D. H. Lee “Off-Line Keyword Guessing Attacks on Recent Keyword Search Schemes over Encrypted Data” ***Secure Data Management*** 2006

[6] D. X. Song, D. Wagner, and A. Perrig. “Practical techniques for searches on encrypted data” ***IEEE symposium on Security and Privacy***, pp 44-55, 2000.

[7] R. Brinkman, L. Feng, J. Doumen, P.H Hartel, and W. Jonker “Efficient Tree Search in Encrypted Data” ***Technical Report***

TR-CTIT-04-15 Centre for Telematics and Information Technology, University of Twente, Enschede. pp 1381-3625, 2004.

[8] Y. C. Chang and M. Mitzenmacher. “Privacy preserving keyword searches on remote encrypted data” ***Cryptology ePrint Archive, Report.*** pp 051, 2004.

[9] 賴溪松,韓亮,張真誠 “近代密碼學及其應用” ***旗標*** 2003

[10] 巫坤品,王青青 ”密碼學與網路安全原理與實務” ***碁峰*** 2005

[11] M. Y. Rhee ”Crytography and Secure Communications” ***McGraw-Hill*** 1994