# 不具連結性搜尋樣式之加密資料搜尋

林峻立

樹德科技大學資訊工程研究所

cllin@mail.stu.edu.tw

陳又誠

樹德科技大學資訊工程研究所

s96639116@mail.student.stu.edu.tw

## 摘要

隨著數位化時代的來臨及網際網路的蓬勃發展，越來越多的資料都儲存在電腦上，資訊安全也相對受到重視；而有些機密資料為了不讓其他者得知，通常會將機密資料加密儲存。爲了有效搜尋資料庫中的加密資料，加密資料往往會伴隨相對應的關鍵字加以儲存。以明文關鍵字進行搜尋是最簡單但卻不具隱私性，有許多研究進而提出將關鍵字加密來進行加密資料的搜尋。然而若相同關鍵字的密文每次都是固定的搜尋樣式，卻也容易被統計分析及連結，進而暴露搜尋者的隱私。因此為了滿足搜尋者更高的隱私，本論文將提出一個不具連結性搜尋樣式的加密資料搜尋方法，也就是相同的關鍵字在每次搜尋時都會產生不同的搜尋樣式。我們運用 Blum-Goldwasser 公開金鑰密碼系統來達到不具連結性的搜尋樣式(unlinkable search pattern)，而我們的方法並沒有因為使用公開金鑰密碼系統而加重搜尋者的運算負擔。

**關鍵詞:** 加密資料搜尋、隱私性、不具連結性搜尋樣式、Blum-Goldwasser 公開金鑰密碼系統

## 1.前言

隨著網路科技的崛起，越來越多的企業和個人利用網路儲存資料於資料庫或個人電腦上，然而網路上所傳遞的訊息其安全性日益重要，如一些機密的資料不想隨便的暴露在公眾網路裡，將資料以加密的形式在網路中傳送和儲存，除溝通的雙方外其他人皆無法得知該訊息的內容，這種加密資料的方式是必要的。

隨著時代的演進和利用網際網路溝通次數也越來越頻繁，為了有效的搜尋加密資料，通常加密資料會伴隨著相對應的明文關鍵字加以儲存，然而用明文關鍵字進行搜尋雖然可以達到搜尋加密資料的目的，但是明文關鍵字未經過加密，易於被網路上的第三者利用追蹤、連結的方式蒐集搜尋者的資訊，因此就衍生出安全性的問題如搜尋的隱私性，雖然資料經過加密但是欲搜尋的關鍵字若未加密仍然會洩露部分搜尋者的資訊，可想而知要解決搜尋的隱私性問題必須將搜尋者欲搜尋的關鍵字隱藏，而最簡單直接的方法就是將關鍵字加密，為此早期 Song [1]等人提出將搜尋的關鍵字經過加密的方式來達到加密資料的搜尋。

雖然早期的加密資料搜尋已做到能對加密資料進行搜尋，這樣的搜尋方式存在著一個隱私性的問題，搜尋者欲搜尋的關鍵字雖然經過加密但仍然是固定的字串，這樣的搜尋模式還是易於被網路上的第三者利用統計或追蹤的方式得知搜尋者相關的訊息。

因此本論文提出一個既能達到加密資料搜尋的需求，又可以保護搜尋者隱私的方法來解決上述的問題，我們的方法運用 Blum-Goldwasser [2]公開金鑰密碼系統的特性，使每次欲搜尋的關鍵字都具有隨機性，此方法我們稱為不具連結性的搜尋樣式(unlinkable search pattern)，所謂的 search pattern 是指搜尋者傳送給資料庫據以搜尋的一組資料，search pattern 可能是關鍵字明文或關鍵字密文或其他針對關鍵字的運算結果。由於網路上的第三者無法針對隨機的搜尋樣式進行統計或追蹤，這一方法更提高了搜尋者的隱私性，另外使用公開金鑰密碼演算法往往需要大量的模指數運算，而我們的方法只需在搜尋者端做少量的運算，因此我們的方法在隱私性和效能上都有很好的表現。

本文結構如下，第 2 節我們說明該研究的相關發展，第 3 節將介紹 Blum-Goldwasser 公開金鑰密碼系統的特性並說明我們所提出的解決方法，第 4 節我們將做安全性和效能分析，最後結論在第 5 節。

## 2.相關研究

在資料加密搜尋的研究中，個人隱私性的問題越來越受到重視，如搜尋者欲使用特定關鍵字搜尋時，關鍵字若未加密或者已加密但反覆使用固定的關鍵字搜尋，這樣的搜尋模式易於被網路上的第三者用統計、追蹤或連結的手法得知搜尋者的相關資訊，因此單單只針對資料加密仍然是不夠安全的，所以 Song [1]等人除了將資料加密之外更將關鍵字加密來達到加密資料搜尋的需求，但是這樣的隱私性仍是不足的，然而個人隱私性問題的研究有兩類不同的考量:一類是針對相同存取者，另一類是針對不同存取者。

不同存取者:此類所探討的是資料儲存者和搜尋者(取用者)是不同人，當搜尋者想在資料庫上搜尋一個或一個以上儲存者所儲存的資料，在這裡搜尋者事先不知道有那些資料可以搜尋，然而搜尋者不想透露這些資料讓網路上第三者得知，如在公共的資料庫(e-mail 資料庫)上做加密資料的搜尋，由於儲存在資料庫上的加密資料，可讓任何人作加密資料的搜尋，因此更需要維護個人資料的隱私性，早期 Boneh [3]和 Golle [4]提出利用公開金鑰密碼系統的特性來達到加密資料搜尋的需求，但是相對的伴隨而來的是大量模指數運算，在這之後陸續有其他針對不同存取者所提出的研究如 Cheg [5]和 Abdalla [6]，上述這些研究都無法解決用固定搜尋樣式所帶來的隱私性問題，近幾年 Ryu [7]和 Boneh [8]提出以非固定的搜尋樣式來解決上述個人隱私性的問題，但是其運算量仍是需要大量的模指數運算。

相同存取者:這裡所探討的為資料儲存者和搜尋者是同一人，資料在儲存時伴隨著相對應的關鍵字存放在資料庫上，當搜尋者想在資料庫上搜尋先前所儲存的資料時，必須用當初所設定已知的關鍵字來搜尋，因此這類所探討的是在私人資料庫上用特定的搜尋樣式搜尋加密資料，並希望能夠隱藏個人的搜尋資訊，解決這一問題的相關研究首推 Song [1]等人的作法，但是其作法在搜尋效率上需要長時間的搜尋，因此繼 Song 等人之後 Goh [9]提出 Index 的概念，加快了搜尋的效率但是其缺點是搜尋時易碰撞，之後陸陸續續有其他針對加密資料搜尋的研究如 Joseph [10]、Curtmola [11]、Schwarz [12]、Lee [13]，雖然這些研究有達到加密資料搜尋的需求，但是都無法解決用固定搜尋樣式所帶來的隱私性問題。而本論文所提出的 unlinkable search pattern 加密資料搜尋，能解決上述的問題且在隱私性和效率上有良好的表現。

# 3.不具連結性搜尋樣式之加密資料搜尋方法

在之前的加密資料搜尋法中都存在著一個隱私性的問題，那就是搜尋者使用固定字串的搜尋樣式而導致網路上的第三者，易於用統計、追蹤或連結的方式得知搜尋者相關資訊，針對此問題我們提出一個解決方法，我們運用 Blum-Goldwasser 公開金鑰密碼系統具有機率式加密和低指數運算的特性，來達到我們宣稱的 unlinkable search pattern 的功能，使每次欲搜尋的搜尋樣式都具有隨機性，進而解決上述隱私性的問題並且提高安全性和效率。以下我們將說明 Blum-Goldwasser 公開金鑰密碼系統與我們所提出的加密資料搜尋方法。

## 3.1 Blum-Goldwasser 公開金鑰密碼系統

此系統由 Blum 和 Goldwasser [2]於 1985 年所提出之 RSA 變形公開金鑰密碼系統，其特性為加密時所產生的密文具有隨機性並使用低指數的運算，此系統的金鑰產生、加密步驟和解密步驟詳細說明如下。

### 1.金鑰產生

(1) 任意選取兩個不同的大質數 $p$ 和 $q$ 且滿足 $p \equiv q \equiv 3 \pmod 4$，並計算 $n = p \times q$。
(2) 計算 $a$ 和 $b$ 且滿足條件 $ap+bq = 1$。
(3) 公開金鑰為$(n)$，私密金鑰為$(p, q, a, b)$。

### 2.加密步驟

(1) 計算 $l = \lfloor \log n \rfloor$，$h = \lfloor \log l \rfloor$；
明文字串 $m=m_1m_2...m_t$，每個 $m_i$ 區塊長度為 $h$。
(2) 取一隨機亂數 $r \in Z_n^*$，計算 $x_0 = r^2 \bmod n$。
(3) 計算 For $i = 1$ to $t$
{
$x_i = x_{i-1}^2 \bmod n$
Let $p_i$ be the $h$ least significant bits of $x_i$
$c_i = p_i \oplus m_i$
}
(4) 計算 $x_{t+1} = x_t^2 \bmod n$，
得到密文 $G = \langle c_1, c_2, ..., c_t, x_{t+1} \rangle$。

**3.解密步驟**

(1) 計算

$d_1 = ((p + 1)/4)^{t+1} \bmod (p - 1)$

$d_2 = ((q + 1)/4)^{t+1} \bmod (q - 1)$

$u = x_{t+1}^{d_1} \bmod p$

$v = x_{t+1}^{d_2} \bmod q$

$x_0 = vap + ubq \bmod n$

(2) 計算 For $i = 1$ to $t$

{

$x_i = x_{i-1}^2 \bmod n$

Let $p_i$ be the $h$ least significant bits of $x_i$

$m_i = p_i \oplus c_i$

}

(3) 得到明文 $m = m_1 m_2 ... m_t$。

### 3.2 不具連結性搜尋樣式之加密資料搜尋

我們的方法解決隱私性和運算量上的問題，我們運用 Blum-Goldwasser 公開金鑰密碼系統具有加密隨機性和低指數運算的特性，來達到更高的安全性和更好的效率。首先，將關鍵字和明文加密再經由安全認證的通道存進資料庫，搜尋時利用 Blum-Goldwasser 公開金鑰密碼系統的特性，使用者可產生一個隨機性的 unlinkable search pattern 讓資料庫搜尋。

**1.加密階段**

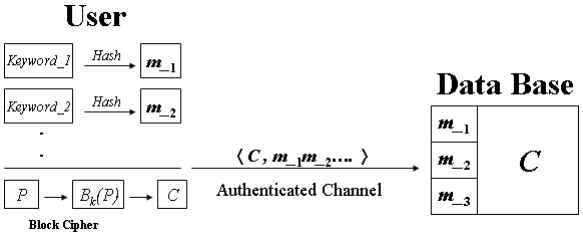


**圖 1、本論文之加密階段**

在圖 1 的加密過程中我們將關鍵字明文(Keyword)和資料明文分別運算，所有關鍵字明文皆經過密碼雜湊函數(cryptographic hash function)運算成為一安全且具隱藏特性的關鍵字雜湊值($m$)，資料明文則使用對稱式區塊加密法(Symmetric Block Cipher)加密而得到密文 $C$，其中 $k$ 為使用者(User)所設定的秘密金鑰，之後再將密文 $C$ 和相對應的關鍵字雜湊值($m$)經由已認證過的安全通道傳送到資料庫(DataBase)儲存，圖 1 為本論文之加密階段。

**2.搜尋階段**

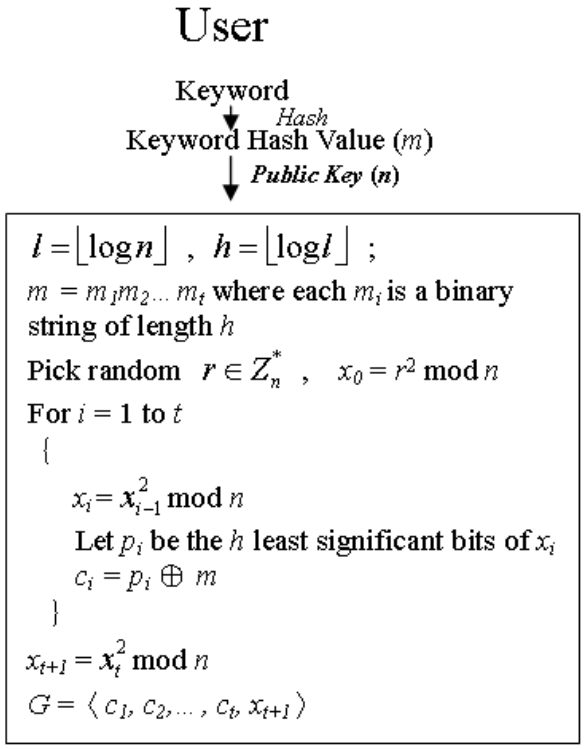


**圖 2、本論文之使用者端搜尋階段**

當搜尋者想讓資料庫幫他搜尋特定的關鍵字，卻又不想洩露任何相關的資訊讓其他人知道，如圖 2 所示我們運用 Blum-Goldwasser 公開金鑰密碼系統的特性，來達到並滿足搜尋者的目的，在此搜尋階段裡我們將分成三個過程來說明：1.取得公開金鑰，2.加密過程，3.解密過程，詳細說明如下。

### 2.1 取得公開金鑰

首先，如圖 2 使用者(User)先將欲搜尋的關鍵字明文(Keyword)經由 Hash 函數運算，成為一個具隱藏特性的關鍵字雜湊值($m$)，並向公正的第三方取得資料庫(DataBase)的公開金鑰($n$)，接著進行加密過程。

### 2.2 加密過程

如圖 2，我們將公開金鑰($n$)和關鍵字雜湊值($m$)代入下列加密演算法進行加密:

(1) 先計算 $l = \lfloor \log n \rfloor$， $h = \lfloor \log l \rfloor$；

$m = m_1 m_2 ... m_t$，每個 $m_i$ 區塊長度為 $h$。

(2) 取一隨機亂數 $r \in Z_n^*$ ，計算 $x_0 = r^2 \bmod n$。

(3) 計算 For $i = 1$ to $t$

{

$x_i = x_{i-1}^2 \bmod n$

Let $p_i$ be the $h$ least significant bits of $x_i$

$c_i = p_i \oplus m_i$

}

(4) 計算 $x_{t+1} = x_t^2 \bmod n$，得到密文 $G = \langle c_1, c_2, \ldots, c_t, x_{t+1} \rangle$。

最後我們得到密文$G$是一個隨機且具隱藏特性的搜尋樣式，我們稱這樣的關鍵字為 unlinkable search pattern，此時再傳給資料庫解密並搜尋，在這演算法中由於亂數 $r$ 為使用者自己隨機產生，因此每次欲搜尋所產生出的搜尋樣式也是隨機的，此特性能防止網路上的第三者藉由統計、追蹤或連結的方式，獲得搜尋者的相關資訊，我們的方法使搜尋者的隱私性達到更高的安全階層。

### 2.3 解密過程

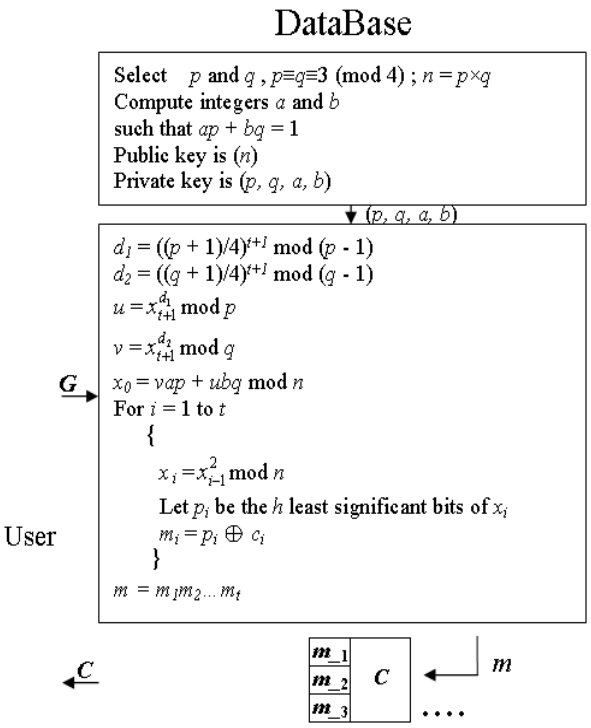


**圖 3、本論文之資料庫端搜尋階段**

如圖 3，資料庫(DataBase)接收到搜尋者欲搜尋的 unlinkable search pattern 也就是密文 $G$，並使用私鑰$(p, q, a, b)$代入下列解密演算法進行解密：

(1) 計算

$d_1 = ((p + 1)/4)^{t+1} \bmod (p - 1)$

$d_2 = ((q + 1)/4)^{t+1} \bmod (q - 1)$

$u = x_{t+1}^{d_1} \bmod p$

$v = x_{t+1}^{d_2} \bmod q$

$x_0 = vap + ubq \bmod n$

(2) 計算 For $i = 1$ to $t$

{

$x_i = x_{i-1}^2 \bmod n$

Let $p_i$ be the $h$ least significant bits of $x_i$

$m_i = p_i \oplus c_i$

}

(3) 得到 $m = m_1 m_2 \ldots m_t$。

最後經過解密得到關鍵字雜湊值$(m)$進行搜尋，資料庫依據關鍵字雜湊值$(m)$找到相對應的加密資料，再將該密文資料回傳給使用者完成整個搜尋的動作。

## 4.安全性和效能分析

一個安全且具隱私性的加密資料搜尋，所需的安全性必須包含儲存資料的機密性和關鍵字隱藏的搜尋。而我們所提出的方法不僅滿足上述的安全需求，更進一步提供 unlinkable search pattern 的搜尋模式，提供更高的隱私性。接下來將說明我們所達到的安全性和效能的分析。

在儲存資料的機密性上，在加密階段中明文資料是先加密後才傳送到資料庫中儲存，其中加密金鑰$(k)$為搜尋者所設置並且是保密的，而伴隨資料密文所儲存的關鍵字也是經過雜湊函數運算過的，因此伺服器並無法解密得知資料和關鍵字的明文。

在關鍵字隱藏的搜尋上，我們的方法中欲搜尋的關鍵字皆經過雜湊函數運算，因此網路上的竊聽者及伺服器皆無法得知關鍵字的內容，而伺服器也僅能以雜湊值搜尋密文的方式為搜尋者進行搜尋。

而進階提供 unlinkable search pattern 的搜尋模式是本論文最重要的特點，在此搜尋的模式中，我們運用 Blum-Goldwasser 這一機率式的 RSA 公開金鑰密碼系統，此模式特殊的地方在於每次欲搜尋的樣式會依使用者所選取的亂數 $r$ 而改變，即使欲搜尋的關鍵字雜湊值$(m)$一樣，也會產生不同的搜尋樣式$(G)$。此方法使得搜尋樣式之間不具連結性，因此提供更高的搜尋者隱私性。

在效能上，雖然我們使用了公開金鑰密碼系統，但是我們利用 Blum-Goldwasser 公開金鑰密碼系統其加密和解密不平衡的指數運算。在搜尋者端的運算只是平方運算，因此並不會造成很大的運算負擔；而在伺服器端雖然有較大的指數運算，但因為關鍵字雜湊值所產生的搜尋樣式其長度有限，且伺服器的配備和運算能力通常是充足且強大的，因此造成的運算負擔也有限。整體而言，我們的方法並沒有因為使用公開金鑰密碼系統而犧牲了效能。

## 5.結論

在本論文中，我們提出 unlinkable search pattern 加密資料搜尋方法，經由每次的關鍵字所產生的隨機搜尋樣式，來防止洩露搜尋者的相關資訊，以保護搜尋者的隱私。不同於以往的加密資料搜尋法，我們的方法不需經過大量的運算，相反地以少量的運算就能達到加密資料兼具隱藏搜尋資訊的目的，由此可見我們的方法是更安全且更具有效率的，尤其在保護搜尋者的隱私上有更好的效果彰顯。

## 參考文獻


[1] D. X. Song, D. Wagner and A. Perrig, “Practical Techniques for Searches on Encrypted Data,” ***In Proc. of IEEE Symposium on Security and Privacy***, pp. 44-55, 2000.

[2] M. Blum and S. Goldwasser, “An efficient probabilistic public key encryption scheme which hides all partial information,” ***Proceedings of Advances in Cryptology - CRYPTO ’84***, pp. 289–299, 1985.

[3] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” ***In Advances in Cryptology - Eurocrypt 2004, volume 3027 of Lecture Notes in Computer Science***, pp. 506–522, Springer-Verlag, 2004.

[4] P. Golle, J. Staddon, and B. Waters, “Secure conjunctive keyword search over encrypted data,” ***In Applied Cryptography and Network Security Conference (ACNS), volume 3089 of Lecture Notes in Computer Science***, pp. 31–45, Springer, 2004.

[5] Y. C. Chang and M. Mitzenmacher, “Privacy Preserving Keyword Searches on Remote Encrypted Data,” ***In Proc, of 3rd Applied Cryptography and Network Security Conference (ACNS)***, pp. 442-455, 2005.

[6] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. M. Lee, G. Neven, P. Paillier and H. Shi, “Searchable Encryption Revisited: Consistency Properties, Relation to Anonymous IBE, and Extensions,” ***In Proc, of CRYPTO***, pp. 205-222, 2005.

[7] E. K. Ryu and T. Takagi, “Efficient Conjunctive Keyword-Searchable Encryption,” ***21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)***, pp. 409-414, 2007.

[8] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. Skeith, ”Public Key Encryption That Allows PIR Queries,” ***Proceedings of CRYPTO***, pp. 1-18, 2007.

[9] E. Goh, “Secure Indexes,” In the Cryptology ePrint Archive, Report 2003/216, March 16, 2004. http://eprint.iacr.org/2003/216/

[10] L. T. A. Joseph, A. Samsudin and B. Belaton, “Efficient search on encrypted data,” ***Networks,Jointly held with the 2005 IEEE 7th Malaysia International Conference on Communication, 13th IEEE International Conference***, pp. 352–357, 2005.

[11] R. Curtmola, J. Garay, S. Kamara and R. Ostrovsky, ”Searchable symmetric encryption: improved definitions and efficient constructions,” ***In: ACM Conference on Computer and Communications Security***, pp. 79–88, 2006.

[12] T. Schwarz, P. Tsui and W. Litwin, “An Encrypted, Content Searchable Scalable Distributed Data Structure,” ***22nd International Conference on Data Engineering Workshops***, pp. 18, 2006.

[13] D. H. Lee, Y. J. Song, S. M. Lee, T. Y. Nam and J. S. Jang, “How to Construct a New Encryption Scheme Supporting Range Queries on Encrypted Database,” ***International Conference on Convergence Information Technology***, pp. 1402-1407, 2007.