多關鍵字可搜尋對稱式加密法之改善

陳昱圻
中興大學資訊科學與工程學系
Email: s9756034@cs.nchu.edu.tw

陳國璋
中興大學資訊科學與工程學系
Email: s9756013@cs.nchu.edu.tw

何宗翰
中興大學資訊科學與工程學系
Email: s9556022@mail.cs.nchu.edu.tw

洪國寶
中興大學資訊科學與工程學系
Email: gbhorng@cs.nchu.edu.tw

***摘要*—機密性資料存放於開放式網路環境的伺服器，而不被資料儲存的伺服器甚至攻擊者獲知存放文件之內容，且資料擁有者又能對資料進行搜尋並取回，這有賴於送出處理過之關鍵字對文件做搜尋。Ballard 等人提出已提出有效率之多關鍵字可搜尋對稱式加密法。本文提出在關鍵字暗門的產生上更有效率之機制。我們的機制在 random oracle model 下安全性等價於 MXDH 假設。**

***關鍵詞*—可搜尋對稱式加密法(searchable symmetric encryption)、多關鍵字搜尋(conjunctive keywords search) 、暗門(trapdoor)**

一、前言

增加資料安全性的方式有許多，簡單又直接的方式是將資料作加密。把加密過的資料儲放在開放空間的伺服器，當使用者需要時，再從伺服器取回加密資料。這種環境下，當使用者要取回一部份所需的資料時，必須先把所有加密資料從伺服器取回，接著解密所有加密資料，再從中挑選所需資料。這方式可以有效地防範惡意伺服器或攻擊者的攻擊，防止非使用者能竊取部份或全部的資料。

然而，隨著資料儲存量增加，取回所有資料，這相當浪費網路頻寬與時間，是個相當不切實際的方式。由其使用者的環境是在運算能力較弱的行動裝置(Personal Digital Assistant, PDA)上、手機(Cell Phone)等，浪費頻寬與時間的方式，是相當不明智。綜合上述原因，只取回必要部份資料，非必要的資料傳輸是重要的。解決的方式，在加密資料上作初步搜尋，在取回想要的部份。

可搜尋式加密法最早是由 Song 等學者提出，利用對稱式金鑰系統[10]，只允許使用者本人建立自己的暗門來搜尋儲存在伺服器上的加密資料；另一方式為使用欄位作為標籤的多關鍵字搜尋[4,5,6,7,9]；還有利用公開金鑰技術[1,3,8]，傳送者利用接收者的公鑰來進行資料加密，接收者利用密鑰建立關鍵字暗門，伺服器利用公鑰進行搜尋。

本文第二部份描述加密文件搜尋之相關研究，第三部份說明我們提出的方法，第四部份為安全性分析與 Ballard 等人方式[2]之比較。最後為本文結論。

二、架構

首先定義架構是一個檔案系統，伺服器是一個誠實且好奇的伺服器，誠實即為它能夠正確的回應所以有的查詢動作，好奇的意思是它會對使用者所上傳之文件感興趣，會想要去探查裡面所含有之內容。

在我們的架構中，與[2]定義相同的 model，使用者上傳的文件如下：

$$[E[M]\,||\,\mathrm{ESSE}(Key,W_1,W_2,...,W_m)]$$

M 為文件之內文，m 為關鍵字的欄位(field)數，key 為使用者金鑰。ESSE 我們提出的可搜尋關鍵字加密演算法。ESSE 不會洩漏任何有關內文的資訊，但卻能夠搜尋不同的關鍵字。此處延用 Golle 等人[7]的假設：

(1) 同一份文件中，相同的關鍵字不會出現在兩個不同的欄位；亦即同一文件內欄位上的關鍵字不會重複出現。
(2) 每份文件都定義每個關鍵字欄位。

因此，可以自行做出定義，舉以出現順序做為欄位的例子，假設我們擁有幾個文件其內文為：D_1: NCHU is in Taichung 與 D_2: I saw a saw, saw a saw，如圖 1 所示。

如果定義的欄位沒有關鍵字，則放入<NULL>，表示欄位內無有效的關鍵字。用一個向量 $D_i=(W_{i,1}, W_{i,2}, \ldots, W_{i,m})$表示第 i 個文件內針對欄位的 m 個關鍵字。在後續的探討中，不會去考慮 $E[M]$這部分，因為其視為這是一個安全的加密法。

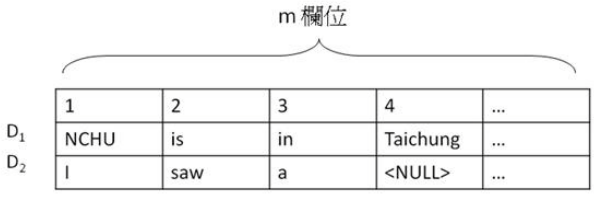

圖 1：自行定義欄位之範例

在搜尋多個關鍵字時，查詢的格式為 $Q=[I_1,I_2,\ldots,I_t,\Omega_1, \Omega_2,\ldots, \Omega_t]$，其中 t 為欲查詢的關鍵字個數，其值域為[1, m]，即意味著只搜尋於所定義的欄位範圍內，而 $I_1,I_2,\ldots,I_t$ 為關鍵字所處在的欄位值，$\Omega_1, \Omega_2,\ldots, \Omega_t$ 為欲搜尋的關鍵字。根據 Q 以及時戳產生對應的暗門，我們稱為 T_Q。伺服器依據 T_Q 去比對文件內是否在欄位內含有符合的關鍵字。

在 Ballard 等人的文章[2]有提到安全性假設 MXDH，訂定 G_1、G_2、G_T，g_1 屬於 G_1 且是 G_1 的生成元、g_2 屬於 G_2 且是 G_2 的生成元。MXDH 問題是分辨兩組數 $L_1=(P_1, aP_1, bP_1, aP_2, bP_2, abP_1)$, $L_2=(P_1, aP_1, bP_1, aP_2, bP_2, cP_1)$，其中 a,b,c 為 Z_p^* 中的隨機值。我們說有一個有多項式時間運算能力的對手 A 具 ε 能力(advantage)去解 MXDH，寫成 $|\Pr[A(L_1)=1]- \Pr[A(L_2)=1]|\geqq\varepsilon$，若 ε 是可忽略的則說 MXDH 是困難的。

Semantic security: 對於公開金鑰密碼系統的 semantic security，意指當只給密文以及對應的加密金鑰，計算能力有限的攻擊者去取得訊息中有意義的資訊是不可行的。

調整式選擇關鍵字攻擊(adaptive chosen keyword attack)，屬於選擇明文攻擊的特例，破密者不只可以預先選擇關鍵字，還可以根據前次的加密結果來調整所欲選擇的關鍵字。同 Golle 等學者提出的 ICLR game[7]，攻擊者可以事先詢問任何關鍵字，並於其後選一挑戰的關鍵字 K_c 給挑戰者，此挑戰的關鍵字於先前並未詢問，挑戰者製造兩個密文 D_0 以及 D_1，其中 D_0 為隨機挑選的關鍵字，D_1 為含有 K_c 以及其他隨機挑選的關鍵字。最後挑戰者給一文件 D，攻擊者要分辨此文件是來自 D_0 還是 D_1。

三、相關研究

Ballard 等人[2]提出了有效率的多關鍵字可搜尋對稱式加密法 (以下簡稱 ESSE)，使用者可以安全的送出加密文件，日後使用者可以針對欲搜尋的多個關鍵字產生暗門來取回他的文件。其中加密訊息格式與本文的建構是相同的，格式為$[E\,[M]\,||\,ESSE(Key, D)]$，ESSE 方法首先定義三個 size 為 p 的群，G_1 及 G_2 為加法群，G_T 為乘法群，P_1 屬於 G_1 且是 G_1 的生成元，P_2 屬於 G_2 且是 G_2 的生成元。另外需要一個雜湊函數 H:{0,1}*$\rightarrow Z_p^*$，以及一個雙線性映射(bilinear map) e: $G_1\times G_2\rightarrow G_T$，以下簡單介紹雙線性映射

的性質：

(1) 可計算性(computable): 任意兩個點 g,h，g 屬於 G_1，h 屬於 G_2，有存在一有效率之演算法來計算 $e(g,h)$屬於 G_T。

(2) 雙線性(bilinear): 對於 1 到 p 的任意整數 x,y 則 $e(xP_1,yP_2)=e(P_1,P_2)^{xy}$。

(3) 不對退化性(non-degenerate): 如果 P_1 為 G_1 的生成元，如果 P_2 為 G_2 的生成元，$e(P_1,P_2)$ 必為 G_T 生成元。

ESSE 機制主要由四個多項式時間隨機演算法組成。

(1) KeyGen(1^k): 產生收件者的隨機私密金鑰 $K \leftarrow \{0,1\}^k$。訂定擬亂函數 $f(\cdot,\cdot) \rightarrow Z_p^*$ 和一個雙線性配對函數 $e: G_1 \times G_2 \rightarrow G_T$ 其中 P_1 為 G_1 生成元，其中 P_2 為 G_2 生成元，P_2 保持私密。

(2) ESSE(K, D): 使用者金鑰對關鍵字集合 $D=\{W_1,W_2,\ldots,W_m\}$產生可搜尋密文 S。選一隨機數 $r \in Z_p^*$ 並計算 $s_i = f_K(W_i)$，得到：

$$S = [rs_1P_1, rs_2P_1, ..., rs_mP_1, rP_1]$$

(3) Trapdoor(K, Q): 使用者利用欲搜尋的內容 $Q = [I_1,I_2,\ldots,I_t,\Omega_1, \Omega_2,\ldots, \Omega_t]$與金鑰產生暗門 T_Q。選一隨機數 $\rho \in Z_p^*$，得到暗門 T_Q 為 $[T_1,T_2,I_1,I_2,\ldots,I_t]$，其中

$$T_1 = \rho \sum_{j=1}^{t} f_K(\Omega_j) P_2,$$

$$T_2 = \rho P_2$$

而 $I_1,I_2,\ldots,I_t$ 為文件中的欄位值。

(4) Test(K, S, T_Q): $S = [A_1, A_2, ..., A_m, B]$，驗證等式如下:

$$e(\sum_{j=1}^{t} A_{I_j}, T_2) = e(B, T_1)$$

如果全部 $W_{Ii} = \Omega_i (1 \le i \le t)$，則輸出”yes”為有搜尋到符合的關鍵字，否則輸出”no”。

在 ESSE 的方法從實做的觀點來看，雙線性映射的 G_1,G_2 常為橢圓曲線加法群，由 G_1,G_2 來得到 G_T 在計算成本上較高的，但在得到 G_T 的生成元後做指數運算會比 G_1,G_2 上做點的乘法在計算成本上較節省。而處於行動裝置的使用者在產生暗門時須支援的運算。

四、我們的方法

我們的方法由下列四個多項式時間隨機演算法所組成。

(1)KeyGen(1^k): 輸入安全參數，定義 2 個雜湊函數 $H_1: \{0,1\}^* \rightarrow Z_p^*$ 與 $H_2: G_T \rightarrow \{0,1\}^l$；定義雙線性配對函數 $e: G_1 \times G_2 \rightarrow G_T$ (G_1,G_2 為加法群、G_T 為乘法群)，其中 P_1 為 G_1 的生成元，P_2 為 G_2 的生成元，P_2 保持私密；產生亂數金鑰 α；並在產生 trapdoor 前事先運算 $X = e(P_1, P_2)$。

(2)ESSE(K, D): 利用使用者金鑰與關鍵字集合 $D = \{W_1,\ldots,W_m\}$產生可搜尋密文 C。選一隨機亂數 $r \in Z_p^*$，並計算 $V_i = H_1(W_i)$，得到

$$C = [rV_1\alpha P_1, rV_2\alpha P_1, ..., rV_m\alpha P_1, \frac{1}{r}P_2]$$

(3)Trapdoor(K, Q): 利用使用者金鑰與欲搜尋的內容 $Q = [I_1,\ldots,I_t, \Omega_1,\ldots, \Omega_t]$來產生暗門 T_Q。挑選亂數 $s \in Z_p^*$，暗門 $T_Q = [T_1,T_2, I_1,\ldots,I_t]$，其中 $T_1 = H_2(X^{\alpha(\sum_{i=1}^{t} H_1(\Omega_i))-s})$, $T_2 = X^s$，而 $I_1,\ldots,I_t$ 為文件中的欄位值。

(4)Test(K, T_Q, C): $C = [A_1,\ldots, A_m, B]$，驗證等式如下：$T_1 = H_2(\frac{e(\sum_{i=1}^{t} A_{Ii}, B)}{T_2})$，如果全部 $W_{Ii} = \Omega_i (1 \le i \le t)$，則輸出”yes”為有搜尋到符合的關鍵字，否則輸出”no”。

五、安全性分析與比較

本文提出的機制安全性建構在以下的定理

上，證明架構於 random oracle model(RO)上。

Theorem 1. 假設 MXDH 是困難的，我們的機制在 RO model 底下，在調整式的選擇關鍵字攻擊下滿足 semantic security。

Proof：假設 A 是一個擁有 ε 能力，破解我們機制的攻擊演算法，且 A 可以做最多 q_T 後門的詢問(為一次搜尋一個關鍵字的次數，相當於關鍵字字集)。我們建立一個演算法在至少 $\varepsilon' \geq \varepsilon/(mq_T e^m)$的機率下可以破解 MXDH 問題，此處的 e 為自然對數的底數。演算法 B 的執行時間大約與 A 相同，然而假如 MXDH 假設在 G_1,G_2 是成立的，則 ε'是一個極小甚至可忽略的函數，因此在安全參數(security parameter)下 ε 必須為可忽略的函數。

將 $[P_1,P_2,a^{-1}P_2,bP_1,cP_1]$ 給演算法 B，其目標為如果 $ab=c$ 輸出 1，它模擬挑戰者的行為以及跟偽造者 A 的互動如下:

H&Enc-queries. A 可以在任何時間對 random oracle H 做詢問動作，為了去回應 H 的詢問，B 建立一個四個欄位<W_j, h_j, x_j, y_j>的列表，此處我們稱為 H-list。此列表初始是空的，當 A 詢問 H 一個 $W_i \in \{0,1\}^*$，B 回應如下:

(1) 如果 W_i 已經存在於H-list 中的一組<W_j, h_j, x_j, y_j>，B 回傳 h_i。

(2) 否則 B 產生一個隨機位元值 $y_i \in \{0,1\}$，Pr[y_i=0]=1/q_T。

(3) 演算法 B 選一隨機 $x_i \in Z_p$，執行下列判斷：

如果 y_i=0，B 計算 $h_i = x_i(bP_1) \in G_1$

如果 y_i=1，B 計算 $h_i = x_i P_1 \in G_1$

(4) B 在 H-list 中增加<W_i, h_i, a_i, y_i>，然後將 h_i 回應給 A。h_i 在 G_1 中是均勻分布而且對於 A 明確要求是相互獨立的。

Trapdoor queries. 當 A 對暗門做詢問符合的文字 W_j(假設 A 每次只會做一個關鍵字之詢問)，B 的回應如下：

(1) 演算法 B 執行上述演算法對於回應 H queries 去取得 $x_{i,j}$ ，且讓<$W_{i,j}$, $h_{i,j}$, $x_{i,j}$, $y_{i,j}$>為對應的 H-list 如果有詢問到 $y_{i,j}=0$，則 B 會失敗。

(2) 我們知道 $y_{i,j}$=1 則定義 $T = H'(X^{x_j - s}), X^s$，對於欲查詢的 W_j 在 α 下是正確的暗門。

Challenge. 最後演算法 A 產生產生一個關鍵字 W 以及位置 z 給 B，而 B 產生挑戰如下:

(1) B 執行先前的演算法去回應 A 對 H 的詢問取得 h 且<W,h,x,y>符合 H-list，如果 c=1 則 B 失敗。

(2) 演算法 B 選多個隨機 $W_{i,j}$，i 為 0,1，j 介於 1~m，除了 $W_{0,z}=W$。B 產生 $D_i=(W_{i,1},\ldots,W_{i,m})$ 兩隨機文件，限制先前的 Trapdoor 不能分辨此兩，給 $W_{0,j}$ 讓<$W_{0,j}$,$h_{0,j}$,$x_{0,j}$,$y_{0,j}$>為符合的 H-list。

(3) B 回應 challenge[$A_1,\ldots A_m,B$]以及 D_0,D_1，計算 challenge value 如下:

如果 $y_{0,j}=0$， $A_j = x_{0,j} cP_1$ 否則 $A_j = x_{0,j} aP_1$ ，

$B = a^{-1}P_2$。若 $c=ab$ ，此 challenge 等同於

$[x_{0.1}aP_1, x_{0.2}aP_1, \ldots, x_{0.j}cP_1, \ldots, x_{0.m}aP_1, a^{-1}P_1]$ 根據此定義，對於被要求的 D_0，是有效的 ESSE 之變數。

More queries. A 可以繼續做詢問，唯一限制就是不能詢問 D_0 或 D_1。演算法 B 則一如往常的回應。

Output. 最後，演算法 A 要輸出 b'為 0 或 1 去分辨 challenge 是 D_0 還是 D_1。如果 b'=0 則 B 輸出"yes" 表示 $c=ab$，否則輸出"no"。

經過 B 的這些完整描述，去確保 B 能夠正確輸出的機率至少是 ε'，所以首先分析 B 在模

擬的過程中不會失敗，定義了以下兩個事件:
E_1: 在任何 *A* 針對暗門的詢問下，*B* 不會失敗。
E_2: 在 challenge 的過程中 *B* 不會失敗。

Claim 1: 由於 *B* 能夠回應所有 *A* 的 Trapdoor queries，*B* 不失敗，則 E_1 的機率是 $1/e^m$。

***Proof*:** 在不失一般性，我們假設 *A* 不會詢問兩次
同一個關鍵字的暗門(一個關鍵字最多一次)，一個對暗門詢問造成 *B* 失敗的機率是 $1/q_T$。讓 W_i 為 *A* 的第 *i* 次對暗門做詢問，也讓$<W_i, h_i, x_i, y_i>$為符合的 H-list。優先出來的查詢，y_i 對 *A* 而言是獨立的，*A* 唯一會得到跟 y_i 有關的是 $H(W_i)$，但是不論 y_i 為 0 或 1，$H(W_i)$出現機率是相同的。然而查詢造成 B 失敗的機率最大是 $1/q_T$，因為 *A* 對暗門最多做 q_T 的詢問，由於任何對暗門的詢問 *B* 都不會失敗的機率至少為$(1-1/q_T)^{q_T} \geqq 1/e$。

這是一點簡單的數學概念，在自然對數的底數 *e* 的定理下 $e=\lim_{n\to\infty}(1+1/n)^n$，當 q_T 足夠大的時候 $\frac{1}{e}=\frac{1}{(1+1/q_T)^{q_T}}=(\frac{1}{1+1/q_T})^{q_T} \leqq (1-1/q_T)^{q_T}$。在 *m* 個關鍵字的情況下為 $1/e^m$。

Claim 2: 在 challenge 的過程中 *B* 不會失敗的機率至少 $1/mq_T$。

***Proof*:** 在 challenge 過程中，*B* 會成功的機率之情況為剛好挑中 *z*(機率為 *1/m*)以及選到 $\Pr[y_i=0] = 1/q_T$ 的那個關鍵字，由此可知 *B* 不會失敗的機率至少為 $1/mq_T$。

演算法 *A* 從未對 W_0 以及 W_1 對暗門做詢問動作，故可以推得 E_1 和 E_2 是獨立的，$\Pr[E_1 \cap E_2] \geqq 1/e^m q_T m$。

根據此二項事件 *B* 成功破解的機率是 $\varepsilon/e^m q_T m$，因為機率極小故 *B* 不易破解 DDH 問題；由於 *B* 不易破解，故假設 *A* 能破解我們的方法是不成立的。因此，可以推論在假設 DDH 破解是棘手的情況下，在抵擋調整式選擇關鍵字攻擊(adaptive chosen keyword attack)，我們提出的方法是符合語意上的安全(semantic security)。

表 1：我們的方法與 Ballard 等人之比較

	Ballard *et al.*	我們的方法
Trapdoor 傳送個數	2	2
Enc	(*m*+1) *mul*	(*m*+1) *mul*
Trapdoor	2 *mul*	2 *exp*
Test	2 *pr*	1 *pr*

另外我們將先進[2]提出的方法以及我們所提出的搜尋架構做比較，如表 1，*m* 為一篇文件內的所有關鍵字個數，*mul* 為點的乘法，*exp* 為指數運算，*pr* 為 pairing 運算。在 Trapdoor 執行效率上我們提出的方法使用預先算好的 *X* 來做 G_T 之有限體的指數運算會比點的乘法來的有效率。

六、結論

本文提出在加密文件多關鍵字搜尋的架構，並提升了暗門產生的效率，更適合行動裝置在產生暗門。我們的方法在 MXDH 假設下滿足語意上的安全去抵擋調整式選擇關鍵字攻擊。

參考文獻

[1] 陳昱圻、洪國寶，“多關鍵字可搜尋公開金鑰加密法之效能改善”，全國資訊安全會議，2009.

[2] L. Ballard, S. Kamara, and F. Monrose, “Achieving efficient conjunctive keyword searches over encrypted data.” International Conference on Information and Communication Security, pp. 414-426, 2005.

[3] D. Boneh, G. Crescenzo, R. Ostrovsky and G.

Persiano, “Public Key Encryption with Keyword Search,” IEEE Symposium on Security and Privacy, pp.44-45, 2004.

[4] J. Byun, D. Lee, and J. Lim, “Efficient Conjunctive Keyword Search on Encrypted Data Storage System”, EuroPKI 2006, LNCS 4043, pp. 184–196, 2006.

[5] R. Curtmola, J. Garay, S. Kamara and R. Ostrovsky, “Searchable symmetric encryption: improved definitions and efficient constructions,” ACM conference on Computer and communications security, pp.79-88, 2006.

[6] E. Goh, “Secure Indexes,” The Cryptology ePrint Archive, Report 2003/216, 2004.

[7] P. Golle, J. Staddon, and B. Waters, “Secure Conjunctive Keyword Search over Encrypted Data”, Applied Cryptography and Network Security Conference, LNCS 3089, pp.31-45, 2004.

[8] D. Park, K. Kim, and P. Lee, “Public key encryption with conjunctive field keyword search”, WISA 2004, pp. 73-86, 2004.

[9] E. Ryu and T. Takagi, “Efficient Conjunctive Keyword-Searchable Encryption”, Advanced Information Networking and Applications Workshops, pp. 409-414, 2007.

[10] D. Song, D. Wagner and A. Perrig, “Practical Techniques for Searches on Encrpted Data”, IEEE Symposium on Security and Privacy, pp.44-55, 2000.