# 프라이버시를 보장하는 SQL 쿼리 가능한 데이터베이스

*박현아 *임종인 *이동훈
고려대학교 정보경영공학전문대학원
*kokokzi@cist.korea.ac.kr

# Privacy Preserving SQL queriable database

*Hyun-A Park *Jong In Lim *Dong Hoon Lee
CIST(Center for Information Security Technology), Korea University

## 요약

정보화 사회에서 사용자의 민감한 정보를 보호하는 가장 확실한 방법은 데이터의 암호화이다. 지금까지 암호화된 데이터에 대해 제안되어진 대부분의 기법들은 모든 종류의 SQL 쿼리가 가능하지 않았고 암호화 기법이나 효율성 등 다양한 측면에서 문제점을 가지고 있었다. 본 고에서 제안하는 기법인 PPSQL은 'Perfect Unlinkability'를 보장한다. 이것은 데이터 자체를 암호화하는 것이 아니라 그 데이터의 열의 위치를 치환시킴으로써 데이터와 그 데이터 소유자와의 관계를 끊어버린다. 이렇게 데이터 자체를 암호화시키지 않음으로써 클라이언트가 쿼리를 만드는 방법이나 서버가 프로세스 하는 방법이 일반 DB와 거의 유사하다. 단지 셋업 시 암호화된 DB를 구축할 때와 서버에 의해 리턴된 결과를 클라이언트가 복호화 할 때의 추가적인 연산만이 필요하다. 또한 우리의 기법은 산술 연산이 가능하고 암호화된 데이터의 속성들 간의 교차 연산을 제외하고는 모든 종류의 SQL 쿼리가 가능하다.

## 1. 서론

유비쿼터스 환경의 도래로 프라이버시는 세계적인 큰 이슈가 되고 있고 이로 말미암아 개인의 민감한 정보를 저장하는 데이터베이스의 관리 역시 중요한 문제로 부각되고 있다. 따라서 이와 관련한 많은 연구가 있어왔다. Public DB상에서 데이터마이닝 시 프라이버시 보호(PPDM, Privacy Preserving in Data Mining)에 관한 연구와 암호화된 문서를 암호화된 검색어로 질의하는 키워드 인덱스 검색 시스템에 관한 연구, 그리고 실제 환경에서 암호화된 데이터베이스에 대한 SQL 쿼리의 최적화에 관한 연구 등이다.

PPDM에서는 프라이버시를 보호하기 위해 anonymization, pertubation, blocking, aggregation 등의 기법들이 사용된다. 하지만, 이런 데이터 조작 기법들은 정확한 통계 결과를 얻을 수 없을 뿐만 아니라 완벽한 프라이버시 역시 보장하지 못한다.

사용자들의 프라이버시를 보호하는 가장 확실한 방법은 데이터의 암호화이다. 최근들어 서버 관리자에 의한 정보 누출과 같은 프라이버시 침해 사례가 증가하고 있다. 암호학적인 기법은 완벽한 프라이버시를 보장하지만 비효율성과 스키마 적용 불가능과 같은 부가적인 문제를 수반한다. 특히 증명 가능한 안전한 암호학적인 기법은 많은 계산 비용과 랜덤한 요소들을 요구하기 때문에 DB 엔지니어들이 원래 그 DB 자체가 제공하는 기능들을 가진 DB 엔진 구축을 어렵게 하며 상용화하기 위한 성능으로는 부적합한 면이 있다.

**공헌도.** 본 고에서 제안하는 기법들은 이전 스킴들의 한계점을 극복하는데 중점을 둔다. 지금까지의 연구에서 암호화된 데이터베이스 상에서 모든 종류의 SQL 쿼리와 산술 연산을 가능하게 하는 기법은 거의 찾아볼 수가 없다. [21]의 기법은 모든 종류의 쿼리를 가능하게 하긴 하지만, 그들이 'privacy homomorphism'[12]의 방법으로 사용한 암호화 기법은 'ciphertext only attack'에 안전하지 않음이 밝혀졌다.

우리가 새롭게 제안하는 기법은 데이터 자체를 암호화하는 것이 아니라 각 속성(attribute)내에서 그것의 열의 위치를 치환시킨다. 이렇게 함으로써 데이터와 그것의 주체와의 연관성을 끊어버려 'Unlinkabikity'를 보장할 수 있는 것이다. 결과적으로, 이 기법은 궁극적인 암호화 기법이 '치환' 함수를 사용한 것이기 때문에 암호화를 하였으나 암호화로 인한 제약 사항은 가지지 않는다고 할 수 있다. 따라서 이 기법은 일반 암호화하지 않은 DB와 비슷한 프로세스를 가지는 면이 많다. 단지 셋업 시 암호화된 DB를 구축할 때와 서버에 의해 리턴된 결과를 클라이언트가 복호화 할 때의 추가적인 연산만이 필요할 뿐이다. 또한 우리의 기법은 암호화된 데이터의 산술 연산이 가능할 뿐만 아니라 속성들 간의 교차 연산을 제외하고는 모든 종류의 SQL 쿼리 또한 가능하다.

## 2. 모델

### 가. 프리미티브의 정의

**정의 1. 일방향 트랩도어 치환 함수(OWTP, one way trapdoor permutation).** 만약 다음을 만족한다면 함수들의 집합 $\Pi=(\Pi_k)$를 일방향 트랩도어 치환 함수라고 한다.

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (IITA-2006-(C1090-0603-0025)).

- $I^K$를 입력값으로 하여 한 쌍의 $(\pi, \pi^{-1})$를 출력하는 확률적인 다항식 시간 알고리즘 $I$가 존재한다. 여기서 $\pi$는 $\Pi_k$에 속한 함수이고, $\pi^{-1}$는 '$\pi$에 대한 트랩도어' 이다.
- 각각의 함수 $\pi \in \Pi_k$는 $\{0,1\}^k$에 관한 치환함수이고 다항 시간 내에 계산 가능하다. 즉, $\pi \in \Pi$와 $x \in \{0,1\}^*$가 주어진다면, 다항 시간 $|x|$내에 $\pi(x)$를 계산할 수 있는 알고리즘이 존재한다.
- 트랩도어 $\pi^{-1}$가 주어진다면, 역으로 $\pi$를 구하는 것은 쉽다. 즉, $y \in \{0,1\}^k$와 $\pi^{-1}$가 주어진다면, 유일한 값 $x=\pi^{-1}(y)$를 계산할 수 있는 알고리즘이 존재한다. 그러나 트랩도어 없이 원함수에 대한 값을 구하는 것은 어려운 일이다.
- 모든 확률적 다항식 시간 알고리즘 B와 모든 정수 c, 충분히 큰 수 $K$에 대하여 다음을 만족한다. [22]

$$\Pr_{\pi = I_\pi(1^k), y \in R\{0,1^k}(B(\Pi, y) = \pi^{-1}(y) < \frac{1}{K^c}$$

**정의 2. 슈도 랜덤 함수 (PRF, pseudo random function)** 만약 오라클 알고리즘 $A$가 $t$ 시간을 최대 실행 시간으로 하여 최대 $q$번의 오라클 쿼리를 하고 $Adv_A < \epsilon$ 의 이길 확률을 가질 때 '$F : K_f \times X \rightarrow Y$는 $(t,q,\epsilon)$ 안전한 슈도랜덤 함수다'라고 말한다. $A$의 이길 확률은 다음과 같이 정의되어진다. $Adv_A = |Pr[A^{Fk}=1] - Pr[A^R=1]|$. 여기서 $R$은 $X$에서 $Y$로 가는 모든 함수 중에서 균일하게 선택되어진 랜덤 함수를 말한다.

**정의 3. 유사 신원 식별자 (quasi-identifier).** 테이블에서 민감하지 않은 속성들의 집합이 외부 데이터와 연결되어서 적어도 한 명의 개인을 유일하게 식별할 수 있다면, 이것을 유사 신원 식별자라고 부른다.

### 나. 프라이버시 요구사항

- **익명성 (Anonymity)**; 익명성은 사용자가 자신의 신원(identity)를 드러내지 않고 서비스나 리소스를 사용하는 것을 보장한다.

- **비연결성 (Unlinkability)**; 사용자들이 리소스나 서비스를 사용할 때 다른 사용자들이 이것을 링크하지 않는 것을 말한다.

- **비관찰성 (Unobservability)**; 사용자가 resource나 서비스를 사용할 때 다른 사용자들이 사용된 리소스나 서비스를 관찰할 수 없는 것을 말한다.

- **가익명성 (Pseudonymity)**; 가익명성은 사용자가 그의 행위에 책임을 져야 하여 익명성을 제공받지 못할 때 사용자의 신원(identity)을 보호할 수 있다. 가익명은 하나 이상의 pseudonyms 하에서 사용자가 그의 신원(identity)을 드러내지 않고 서비스나 리소스를 이용할 수 있도록 하는 것을 말한다.

### 다. 알고리즘

- ***GenSysPar($1^K$, $1^{K'}$)*** - 파라메터 생성 알고리즘 GenSysPar($1^K$, $1^{K'}$)는 시큐리티 파라메터 K, K'을 입력값으로 하여 시스템 파라메터 λ를 생성한다.
- ***GenEnDB(λ, OgDB)*** - 시스템 파라메터와 오리지널 DB, OgDB를 입력값으로 하여 암호화된 DB, EnDB를 출력한다.
- ***ImpQuery(EnDB,Qc)*** - EnDB와 클라이언트의 쿼리 $Q_c$를 입력값으로 하여 서버의 응답 Rs를 출력한다.
- ***Decrypt(Rs)*** - 서버의 응답 Rs 가 주어진다면, 클라이언트의 응답 Rc = D(Rs)를 출력한다. .

## 3. PPSQL의 설계

### 가. GenSysPar($1^K$, $1^{K'}$)

시큐리티 파라메터 K, K'가 주어진다면, 알고리즘 GenSysPar($1^K$, $1^{K'}$)는 시스템 파라메터 λ={f(·), π(·), k, k'}를 출력한다. 여기서 $k \leftarrow K \in \{0,1\}^K$와 $k' \leftarrow K' \in \{0,1\}^{K'}$는 클라이언트의 비밀키이다. $f(\cdot) \leftarrow F_{K'}$와 $\pi(\cdot) \leftarrow \Pi_{K'}$는 슈도 랜덤 함수와 일방향성 트랩도어 치환 함수라고 하며, $F_K$와 $\Pi_K$는 슈도 랜덤 함수 패밀리와 일방향성 트랩도어 치환 함수들의 집합이라고 한다.

### 나. GenEnDB(λ, OgDB)

λ를 생성한 후, 알고리즘 *GenEnDB*는 오리지널 DB를 암호화 시켜 암호화된 DB를 생성한다. 각 속성 $A_j$에 대하여, 유일한 일방향 트랩도어 치환 함수가 랜덤하게 선택되어진다. 우선, 비-수량화 데이터를 슈도 랜덤 함수 $f_k$를 가지고 암호화 한다. 여기서 $k'$은 클라이언트의 비밀키이며, 앞으로 간단히 그냥 $f$로 나타내기로 한다. 그 다음으로는 비수량화 데이터와 수량화 데이터 모두를 치환 함수 $\pi_f$를 *사용하여* 각 속성 필드 내에서 치환 시킨다.

**상세 과정**. 오리지널 DB는 표1의 좌측과 같다. 만약 속성의 개수 m이 4로, 레코드들의 개수 n이 4로 주어진다면;

| $A_1$ | $A_2$ | $A_3$ | $A_4$ | | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|---|---|---|---|---|
| $id_1$ | $q_{1,2}$ | $v_{1,3}$ | $v_{1,4}$ | ⇒ | $f_{1,1}(id_1)$ | $f_{1,2}(q_{1,2})$ | $v_{1,3}$ | $v_{1,4}$ |
| $id_2$ | $q_{2,2}$ | $v_{2,3}$ | $v_{2,4}$ | ⇒ | $f_{2,1}(id_2)$ | $f_{2,2}(q_{2,2})$ | $v_{2,3}$ | $v_{2,4}$ |
| $id_3$ | $q_{3,2}$ | $v_{3,3}$ | $v_{3,4}$ | ⇒ | $f_{3,1}(id_3)$ | $f_{3,2}(q_{3,2})$ | $v_{3,3}$ | $v_{3,4}$ |
| $id_4$ | $q_{4,2}$ | $v_{4,3}$ | $v_{4,4}$ | ⇒ | $f_{4,1}(id_4)$ | $f_{4,2}(q_{4,2})$ | $v_{4,3}$ | $v_{4,4}$ |

표 1 오리지널 DB(좌)와 비수량화 데이터의 암호화(우)

- $id_i$ ; $i$번째 개체의 identity
- $q_{i,j}$ ; j번째 속성(필드)에 있는 $i$번째 개체의 유사 신원 식별자
- $v_{i,j}$ ; j번째 속성(필드)에 있는 $i$번째 개체의 수량화 값
- $n$ ; 행(개체)들의 총 개수
- $i \in [1,n]$, $j \in [1,m]$

우선, 사용자의 identity $id_i$와 유사 신원 식별자 와 같은 비수량화 데이터들을 슈도 랜덤 함수 $f_k$를 가지고 암호화 한다. 그리고 나서, 수량화 값 $v_{i,j}$를 표함한 각각의 속성 $A_j$에서 그 속성 내의 모든 값들은

사전에 미리 할당된 치환 함수 $\pi_j$에 따라 치환되어진다. 즉, $v_{i,j}$에 대하여;

$E(v_{i,j}) \Rightarrow \pi_j(i) = y \Rightarrow v'_{y,j} = E_{y,j} = v_{i,j}$

따라서, j번째 칼럼에서 $i$번째 행의 값은 y번째 행의 값으로 바뀐다. 표1의 암호화된 DB는 표2와 같다 ;

| $A_1$ | $A_2$ | $A_3$ | $A_4$ |
|---|---|---|---|
| $E_{1,1} = f_{4,1}(id_4)$ | $E_{1,2} = f_{3,2}(q_{3,2})$ | $E_{1,3} = v_{2,3}$ | $E_{1,4} = v_{2,4}$ |
| $E_{2,1} = f_{3,1}(id_3)$ | $E_{2,2} = f_{4,2}(q_{4,2})$ | $E_{2,3} = v_{4,3}$ | $E_{2,4} = v_{1,4}$ |
| $E_{3,1} = f_{2,1}(id_2)$ | $E_{3,2} = f_{1,2}(q_{1,2})$ | $E_{3,3} = v_{1,3}$ | $E_{3,4} = v_{4,4}$ |
| $E_{4,1} = f_{1,1}(id_1)$ | $E_{4,2} = f_{2,2}(q_{2,2})$ | $E_{4,3} = v_{3,3}$ | $E_{4,4} = v_{3,4}$ |

표2 암호화된 DB

마지막으로, 클라이언트는 단지 id의 암호화된 필드와 슈도 랜덤 함수 $f_k$, 각 속성에 해당하는 치환 함수 $\pi_j$만을 유지하고 있어서 서버에게 암호화된 DB를 전송하고 복화화된 결과를 출력할 수 있다.

### 다. *ImpQuery(EnDB,Qc)*

만약 사용자가 오리지널 DB에 관한 쿼리를 한다면, 클라이언트는 암호화된 DB에 대한 클라이언트의 쿼리 Qc를 다시 써서 서버에게 전송한다. 그러면 서버는 클라이언트의 요청을 수행하고, 클라이언트에게 그 결과를 리턴한다.

**상세 과정.** 만일 사용자가 오리지널 DB에서 속성 $A_j$의 MAX 값을 요청한다면, 클라이언트는 서버에게 암호화된 DB에 대한 SQL 쿼리를 질의한다. 여기서 SQL 쿼리 구문은 일반 오리지널 DB와 비슷하다. 만약 결과가 $E_{i,j}(=V)$ 라면, 서버는 서버의 결과값 $R_S=(y,j,V)$ 를 리턴한다.

### 라. *Decrypt(Rs)*

서버의 결과값 $R_S=(y,j,V)$를 받자마자, 클라이언트는 V 값의 원래 행의 위치를 찾아내야 한다;

$$r_o = \pi_j^{-1}(y) = i$$

다음으로 클라이언트는 $id_i$를 찾기 위해 $\pi_1(i)=z$를 계산한다. 즉, identity 속성에서 $z$번째 행의 값을 찾아내야 한다는 것이다. $E_{z,1} = E_{\pi_1(i),1} = f_{i,1}(id_i)$이기 때문에 우리는 다음과 같은 방법으로 $id_i$를 구할 수 있다;

$$f^{-1}(E_{z,1}) = f^{-1}(f_{i,1}(id_i)) = id_i$$

따라서 클라이언트의 결과는;

$Rc=(id_i, V)$

## 4. 프라이버시 분석

여기서는 제안하는 기법이 앞서 말한 프라이버시 요구사항을 모두 만족하는지에 대해 살펴보기로 한다.

우선, 익명성과 가익명성은 identity 와 유사 신원 식별자를 암호화함으로써 가능하며 이것은 명백하다. 비연결성과 비관찰성은 암호화 방법의 적용과 속성 내의 모든 값들을 치환함으로써 가능하다.

**정리1.** 만약 $f_k$가 슈도 랜덤 함수이고 $\pi_j \in_R \Pi_k$가 일방향성 트랩도어 치환 함수이면, PPSQL은 안전하다.

**<증명>** PPSQL을 깨는 공격자 $A$가 PPSQL을 깨는데 성공할 확률을 $Ad_A$라고 하고, 총 m개의 속성들 중 $p$개의 수량화 정보와 $q$개의 비수량화 정보로 DB가 구성되어 있다고 할 때, 다음과 같은 식을 만족한다.

$$Adv_A \le \left(\frac{1}{n}\right)^p + \left(Q \cdot \frac{1}{n}\right)^q$$

이 식의 성립은 명백하기 때문에 따로 증명하지 않는다. $\left(\frac{1}{n}\right)^p + \left(Q \cdot \frac{1}{n}\right)^q$의 값은 의미없이 negligible한 값이므로, 이 값에 바운드 되어지는 공격자의 성공확률 역시 negligible하므로 PPSQL은 안전하다.

## 5. 논의 및 결론

본 고에서 제안하는 기법의 주 목적은 데이터베이스에 저장된 데이터를 암호화하여 정보 주체들의 프라이버시를 보장하면서 SQL 쿼리 역시 가능하게 하는 것이다. 이를 위한 방법으로 우리는 데이터 자체를 암호화하지 않고 치환시키는 방법을 이용함으로써 정보와 정보 주체 간의 관계성을 끊어버려 완벽한 비연결성을 보장하고, 그와 동시에 산술 연산 및 속성들 간의 교차 연산을 제외한 모든 종류의 SQL 쿼리를 가능하게 하였다. 암화화된 DB의 비효율성의 주 원인인 검색시 각 행마다 테스트 과정을 거쳐야 하는 문제점을 해결하여 일반 DB와 거의 같은 검색 과정이 되게 함으로써 보다 효율적이다. 즉, 치환 함수를 이용하여 암호화하였으나, 데이터 자체를 암호화 한 것이 아니기 때문에 DB 상에서 암호화로 인한 제약 사항에 거의 구애를 받지 않고 단지 클라이언트 단에서 셋업시와 결과 복호화시 추가 계산 비용만이 들 뿐이다.

하지만, 제안하는 기법은 동적인 DBMS(database management system) 상에서 업데이트에 관한 약점을 가지고 있다. 따라서, 향후엔 이런 모든 문제점을 극복하고 privacy homomorphism을 만족하는 암호 알고리즘과 더불어 모든 종류의 SQL 쿼리를 가능하게 하는 기법에 관한 연구가 더 필요할 것으로 보여진다.

## 참고 문헌


1. Michel Abdalla, Mihir Bellare, Dario Catalano, Eike Kiltz, Tadayoshi Kohno, Tanja Lange, John Malone-Lee, Gregory Neven, Pascal Paillier, and Haixia Shi, Searchable encryption revisited: Consistency properties, relation to anonymous IBE, and extensions. In Proceedings of CRYPTO 05, pp. 205-222. LNCS, 2005.

2. Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky, and Giuseppe Persiano, Public key encryption with keyword search, In EUROCRYPT, pages 506-522, 2004.

3. J. Bethencourt, H. Chan, A. Perrig, E. Shi, and D. Song, Anonymous multi-attribute encryption with range query and conditional decryption, Technical report, CMU-CS-06-135, 2006.

4. Krista Bennett, Christian Grothoff, Tzvetan Horozov, Ioana Patrascu, ""Efficient sharing of encrypted data"", in the proceedings of the 7th ACISP, 2002

5. L. Ballard, M. Green, B. de Medeiros, F. Monrose, Correlation-Resistant Storage via Keyword-Searchable Encryption, SPAR Technical Report TR-SP-BGMM-050705.

6. Dan Boneh and Brent Waters, Conjunctive, Subset, and Range

Queries on Encrypted Data, In the Proceedings of TCC 07, 2007.

7. B. Chor, N. Gilboa, M. Naor, Private Information Retrieval by Keywords Technical Report TR CS0917, Department of Computer Science, Technion, 1997.

8. Y.C. Chang and M. Mitzenmacher, Privacy preserving keyword searches on remote encrypted data, Cryptology ePrint Archive, Report 2004/051, Feb 2004.

9. C. Castelluccia and E. Mykletun and G. Tsudik, Efficient Aggregation of encrypted data in Wireless Sensor Networks, Mobile and Ubiquitous Systems: Networking and Services, 2005.

10. Sun S.Chung and Gultekin Ozsoyoglu, Processing Aggregate Queries Over Encrypted Relational Databases, Technical Report, Case Western Reserve University, 2005.

11. Sun S.Chung and Gultekin Ozsoyoglu, Processing Aggregation Queries over Encrypted Databases, ICDE 2006, LNCS, 2006.

12. J. Domingo-Ferrer, A new privacy homomorphism and applications, Information Processing Letters, 1996.

13. J. Domingo-Ferrer, A Provably Secure Additive and Multiplicative Privacy Homomorphism, ISC, LNCS 2433, pp. 471-483, 2002.

14. Wenliang Du and Yunghsiang S. Han, Privacy-Preserving Multivariate Statistical Analysis : Linear Regression and Classification.

15. T. Dalenius and S. Reiss, Data swapping: A technique for disclosure control, Journal of Statistical Planning and Inference, 1982.

16. Eu-Jin Goh, Secure Indexes, Cryptology ePrint Archive, 2004.

17. Philippe Golle, Jessica Staddon, and Brent R. Waters, Secure conjunctive keyword search over encrypted data, In ACNS, pages 31-45, 2004.

18. Hakan Hacigumus, Balakrishna R. Iyer, Li Chen, Sharad Mehrotra, Executing SQL over Encrypted Data in the Database-Service-Provider Model, In the Proceedings of ACM SIGMOD, June, 2002.

19. H. Hacigumus, B. Iyer, and S. Mehrotra, Efficient Execution of Aggregation Queries over Encrypted Relational Databases, DASFAA, LNCS 2793, pp.125-136, 2004.

20. H. Hacigumus, B. Iyer, and S. Mehrotra, Query Optimization in Encrypted Database Systems, DASFAA 2005, LNCS 3453, pp. 43.55, 2005.

21. Geetha Jagannathan, Rebecca N. Wright, Privacy-Preserving Distributed k-Means Clustering over Arbitrarily Partitioned Data ,KDD 2005.

22. Eyal Kushilevitz1 and Rafail Ostrovsky2, One-Way Trapdoor Permutations Are Sufficient for Non-trivial Single-Server Private Information Retrieval, EUROCRYPT 2000, LNCS 1807, pp. 104-121, 2000.

23. Lea Kissner and Dawn Song, Private and Threshold Set-Intersection, CMU TR, 2004.

24. A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam,l-diversity: Privacy beyond k-anonymity, in proceedings of ICDE 2006, LNCS, 2006.

25. Einar Mykletun and Gene Tsudik, Aggregation Queries in the Database-As-a- Service Model, In the proceedings of DBSEC 2006.

26. W. Ogata and K. Kurosawa, Oblivious Keyword Search, Journal of Complexity, Volume 20, Pages 356 - 371, 2004.

27. G.Ozsoyoglu, D.Singer, S.Chung , Anti-Tamper Databases: Querying Encrypted Databases, In the Proceedings of IFIP 11.3 Conference 2003 on Database Security, August 3 - 6.

28. Hyun-A Park, Jin Wook Byun, Dong Hoon Lee, Secure Index Search for Groups, TrustBus 05, LNCS 3592 pp.128-140, 2005.

29. R. L. Rivest, L. Adleman and M. L. Dertouzos, On data banks and privacy homomorphisms, in Foundations of Secure Computation, New-York: Academic Press, pp.169-179, 1978.

30. Radu Sion and Bogdan Carbunar, Conjunctive keyword search on encrypted data with completeness and computational privacy, Cryptology ePrint Archive : Report 2005/172.

31. Dawn Xiaodong Song, David Wagner, and Adrian Perrig, Practical techniques for searches on encrypted data, Proceedings of the 2000 IEEE Symposium on Security and Privacy, page 44, 2000.

32. Brent R.Waters, Dirk Balfanz, Glenn Durfee, and D. K. Smetters, Building an encrypted and searchable audit log, In Proceedings of Network and Distributed System Security Symposium 2004 (NDSS04), February 2004.

33. Zhiqiang Yang, Rebecca N. Wright, Improved privacy-preserving bayesian network parameter learning on vertically partitioned data, PDM 05.

34. Z. Yang, S. Zhong, and R. N. Wright, Privacy-Preserving Queries on Encrypted Data, In Proceedings of the 11th European Symposium On Research In Computer Security (Esorics), 2006.

35. S. Zhong, Z. Yang, and R. N. Wright, Privacy-enhancing kanonymization of customer data, In PODS, 2005.

36. Ontario, Office of the Information and Privacy Commissioner (IPC) and Netherlands Registratiekamer (1995), Privacy-Enhancing Technologies: The Path to Anonymity, Information and Privacy Commissioner and Registratiekamer, at: http: // www. ipc. on ca/ english/ pubpres/ papers/ anon-e.htm