情 報 保 護 學 會 誌
第19卷 第3號, 2009. 6

데이터베이스 아웃소싱을 위한 준동형성 암호기술

송 유 진*, 박 광 용**

요 약

최근 정보보호 패러다임은 네트워크 보안에서 데이터베이스 보안으로 진화되고 있는 가운데 데이터베이스의 관리 및 운용을 외부사업자에게 위탁하는 데이터베이스 아웃소싱 서비스(DAS, Database As a Service)가 활성화되고 있다. 이러한 가운데 포털사이트나 기업내 데이터베이스에 저장된 데이터 유출이나 도난의 위험성이 증가되고 있다. 본 논문에서는 DAS 모델 환경에서 암호화 데이터의 효율적인 연산이 가능한 준동형 암호 방식을 검토한다. 또한 본 논문에서는 Generalized Paillier 암호방식을 암호화된 개인 의료데이터의 연산에 적용한다.

Ⅰ. 서 론

최근 정보보호 패러다임은 네트워크 보안에서 데이터베이스(DB) 보안으로 진화되고 있는 가운데 개인정보의 대량 유출 사건이 끊임없이 발생하고 있어 DB 보안에 대한 요구 사항이 매우 높아지고 있다. 아울러 DB의 관리 및 운용을 외부사업자에게 위탁하는 데이터베이스 아웃소싱 서비스(DAS, Database As a Service)가 활성화되고 있다.

데이터베이스 아웃소싱(outsourcing)이란 데이터베이스 관리를 전문으로 하는 기업이 데이터 소유자 대신에 데이터를 관리하는 서비스이다. 기존의 데이터베이스 관리시스템(DBMS)은 다양한 보안 기능을 갖추고 있지만 원칙적으로 DB 관리자의 신뢰를 전제로 하고 있다. 서비스 이용자는 DB 관리자에게 노출되지 않도록 기밀내용을 외부의 데이터베이스 서비스 제공자(ISP, Information Service Provider)에게 위탁해야 한다. 이러한 배경에서 DB 관리자에 대해 기밀성을 유지할 수 있는 암호화 데이터베이스 시스템이 연구되고 있다[2][3][4][9].

한편, 일반적인 공격자로는 침입자(시스템에 대한 접근 권한을 불법으로 취득해서 정보를 얻으려는 사람), 내부자(신뢰할 수 있는 사용자 그룹에 소속하면서 자신의 접근권한 이외의 정보를 얻으려고 하는 사람)이 고려될 수 있는데 데이터베이스 아웃소싱의 경우, DB 관리자도 공격자가 될 수 있다.

관리자가 공격자가 되는 예로서

① 소홀한 관리에 의해 다른 사용자나 침입자에게 권리가 부정하게 양도되는 경우

② 관리자 자신이 의도적으로 데이터를 제3자에게 매도하는 경우

등을 들 수 있다. 관리자 자신 또는 관리자로부터 부정하게 권리를 얻은 사용자는 다음과 같은 공격이 가능하다.

· 직접적인 공격 : DB의 부정한 열람 · 수정 · 삭제 등

· 간접적인 공격 : 로그 기록이나 시스템 카탈로그, 통계 정보의 부정한 이용 등

· 메모리 공격 : 서버에 직접 접근해서 메모리상 데이터의 부정 이용 등

이러한 공격의 대처 방법으로서

① 열람해도 내용 파악을 할 수 없도록 하는 데이터 암호화,

② 해석으로부터의 정보누설을 방지하는 로그 기록의 관리,

③ 서버의 메모리상에 흔적를 남기지 않도록 하는 데이터 관리의 철저

등이 있다. 이와 같이, DB 관리 소홀에 따른 기밀정보 유출이나 관리자의 내부 범죄가 발생할 경우, DB에

* 동국대학교 정보경영학과 (song@dongguk.ac.kr)
** 동국대학교 전자상거래협동과정 (freemickey@dongguk.ac.kr)

저장된 중요 데이터를 안전하게 관리할 수 없게 된다. 이러한 배경에서 암호화된 DB와 색인을 관리자 서버에 두고 DB 관리자에 대해서도 기밀성을 유지할 수 있는 암호화 데이터베이스 시스템 모델인 DAS 모델이 연구되고 있다[3].

본 논문의 2장에서는 DAS 모델의 필요성과 모델에 대해 알아본다. 3장에서는 DAS 모델 환경에서 암호화된 데이터의 연산을 효율적으로 수행할 수 있는 준동형 암호 기술에 대해 서술한다. 4장에서는 준동형 암호기술을 이용한 u-healthcare에의 응용에 대해 알아보고 5장에서 결론을 맺는다.

Ⅱ. DAS 모델

2.1 DAS 모델의 필요성

최근 각종 포털사이트 및 기업에서 개인정보의 유출 문제로 인해 큰 피해를 입고 있으며, 그에 대한 대책이 시급한 실정이다. 그러나 DB 유출에 따른 대부분의 피해는 외부자의 유출이 아닌 내부자에 기인하고 있다[4]. 또한, 대부분의 데이터 보호는 접근제어에만 의존하거나 인덱스 암호화를 통해 검색시 서버에서 복호화가 이루어지고 있다. 즉, 내부자에 의한 공격은 고려되지 않고 있기 때문에 내부 공격자로부터 DB 를 암호화하는 방법이 주로 사용되고 있다.

DB를 암호화할 경우 DB가 해킹이 되더라도 데이터 분석이 불가능한 암호화된 형태로 접근하기 때문에 기밀 데이터의 외부유출 및 내부자에 의한 유출을 방지할 수 있다. 따라서 이러한 유출문제를 해결하기 위해서 DB를 암호화하여 관리한다. 그러나 DB를 암호화할 경우, 암호화의 장애물로써 높은 구축 비용, 복잡성 및 시스템 성능저하 문제가 발생한다.

이러한 문제를 해결하기 위해 DAS 모델[2]이 제안되었고 이러한 모델상에서 효율적인 질의(Query)처리가 가능한 방법이 필요하게 되었다. 본 논문에서는 효율적인 질의 처리를 통해 암호화된 데이터의 연산이 가능한 준동형 암호화 및 u-healthcare에의 응용에 대해 검토한다.

DAS 모델[2][3]은 데이터베이스를 아웃소싱하는 개념이다. 대부분의 기업과 포털 사이트들은 DBMS를 필요로 한다. 그러나 DBMS는 구축, 설치 및 유지보수 등이 매우 복잡하고, DBA는 높은 비용과 숙달된 기술을 필요로 한다. 이때, DAS 모델을 기업에서 사용하는 경우, 비용 절감과 막대한 인프라 투자 및 관리의 어려움을 줄일 수 있다.

한편, 기업이나 공공기관 등에서 민감한 기밀정보를 DB에 보관하게 될 경우, 기밀성 유지가 필요하다. 또한 의료 서비스를 제공할 경우, 사용자 및 의료 내역 에 대한 프라이버시 보호가 필요하다.

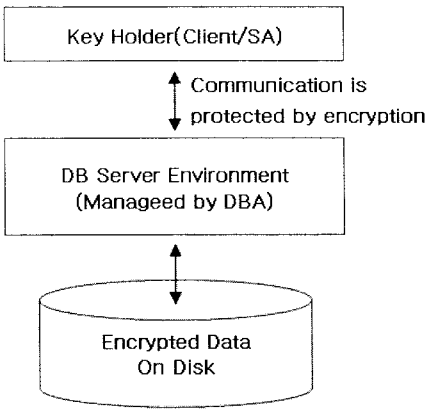
예를 들어, 의료정보, 위치정보 및 신용정보와 같은 개인의 민감한 정보를 보관하고 있는 데이터베이스 시스템의 DAS 운용환경에서는 기밀성 및 프라이버시 보호를 위하여 반드시 데이터를 암호화할 필요가 있다. 이러한 경우, AES 등 기존의 암호화 알고리즘을 DB에 적용할 때, 암호화된 데이터의 순서가 평문과 상이하게 되므로, 색인정보를 구축할 수 없게 되어 효율성 문제가 발생한다. 따라서, DAS 환경에서 DB 서비스를 안전하게 사용할 수 있는 효율적인 프라이버시 보호 방식이 요구된다.

본 논문에서는 프라이버시를 보호하면서 효율적으로 암호화된 데이터에 대한 연산이 가능한 모델에 대해 검토한다.

2.2 DAS 환경에서의 암호화 모델

DAS 모델은 inside-the-box와 outside-the-box 암호화 모델이 있다[4]. inside-the-box 암호화 모델에서는 DBA에 의해 관리, 검사되는 운영환경으로 DB 서버에 의해 암/복호화가 수행된다.

SA(Security Administrator)가 DB 서버 환경 외부에 위치하고 있는 outside-the-box 암호화 모델은 DB 서버



〔그림 1〕 Outside-the-box 암호화 모델

환경 외부에서 암호화가 수행된다.

따라서, 내부자에 의한 데이터 유출을 고려할 경우 outside-the-box 암호화 방식이 보안상 더욱 안전한 모델이다. 이러한 모델에서 클라이언트 어플리케이션 자체에서 암/복호화가 이루어질 경우, 클라이언트가 SA의 역할을 하며, 이때 암/복호화를 수행하는 SA를 Key Holder라고 한다[4][그림 1].

그러나 DAS 모델에서의 주요 이슈는 질의를 수행하는 것인데 질의처리는 DB 서버상에서 처리되는 것이 대부분으로써 암호화된 데이터를 복호화하지 않고 서버에서 질의를 처리하는 것은 매우 어렵다.

이와 같이 DAS 모델에서는 암호문상에서 직접 연산을 수행할 경우, SUM과 AVG와 같은 aggregate 질의 수행이 가능한 준동형 암호방식에 대해 살펴본다.

Ⅲ. 준동형 암호(Homomorphic Encryption)

암호화된 데이터상에서 직접 연산을 수행하는 개념은 최초로 Rivest[1]에 의해 1978년에 소개되었다. Privacy Homomorphism(PH)은 암호화된 데이터에 대하여 복호화 함수에 대한 특별한 정보가 없이 연산이 가능한 방식이다. 엄밀히 말해 대수학적인 환(Ring)상에서 PH암호문 E는 $E(xy)$와 $E(x+y)$에서 x와 y에 대한 복호화 과정이 필요 없이 효과적으로 계산할 수 있는 알고리즘이다. 따라서 PH는 복호화 함수에 대해서 모르더라도 암호화된 데이터만으로도 연산을 가능하게 한다.

PH는 다음과 같이 정의된다[11].

$a_1, a_2, \ldots, a_m$은 암호화되지 않은 값이고 E_k 는 k 를 사용하는 암호화 함수이며, D_k 는 이에 대응하는 복호화 함수이다.

여기서, $A=\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$와 $B=\{\beta_1, \beta_2, \ldots, \beta_n\}$가 Function Families라고 한다. 이때,

$D_k(\beta_i(E_k(a_1), E_k(a_2), \ldots E_k(a_m))) = \alpha_i(a_1, a_2, \ldots a_m)$

를 만족하는 경우, $i \le 0 \le n$인 모든 i에 대해서 (E_k, D_k, A, B)는 PH이다.

RSA 암호의 예를 들면, 다음의 관점에서 승법 준동형 암호임을 알 수 있다.

만약, $C_i = E(M_i) = M_i^e \bmod n$가 주어질 경우,

- $C_1 = M_1^e \bmod n$, $C_2 = M_2^e \bmod n$
- $C_1 * C_2 = M_1^e * M_2^e \bmod n = (M_1 * M_2)^e \bmod n$

이다. 즉, 암호문의 곱은 평문의 곱과 같은 결과를 얻을 수 있다.

3.1 Symmetric 준동형 암호

3.1.1 Domingo-Ferrer[11]

Domingo-Ferrer는 1996년 "A new privacy homomorphism and applications"[11]에서 덧셈, 뺄셈, 곱셈이 모두 가능한 대수적 PH를 제안하였다. 여기서, Domingo-Ferrer의 알고리즘에 대해서 설명한다.

(1) 키 생성 과정

- p, q : 두 큰소수 선택, $n = pq$ 를 계산
- d : 양의 정수 선택
- r_p, r_q 선택

[공개키] : n, d

[비밀키] : p, q, r_q, r_q

(2) 암호화 과정

p와 q를 충분히 큰 소수로 하고 n을 pq로 계산한 후 양의 정수 d를 선택한다. 여기서 공개키는 (d, n)이고 비밀키는 (p, q, r_p, r_q)이다. n을 증가시킴으로 비밀을 보호할 수 있으나, 암호화된 데이터의 효율성이 저하된다는 문제가 있다. $x \in Z_n$을 랜덤하게 분할하여 $(x_1, x_2, \ldots, x_n)$로 한다. 즉,

$x = \sum_{i=1}^{d} x_i \bmod n$이고 $x_i \in Z_n$이다.

$E_k(x) = ([x_1 r_p \bmod p, x_1 r_q \bmod q],$
$[x_2 r_p^2 \bmod p, x_2 r_q^2 \bmod q], \cdots, [x_i r_p^n \bmod p, x_i r_q^n \bmod q])$

(3) 복호화 과정

각각의 i에 대하여 $[x_i \bmod p, x_i \bmod q]$를 복원하기 위하여 $([x_i r_p^{-i} \bmod p, x_i r_q^{-i} \bmod q])$으로 i번째 좌표쌍 $[x_i r_p^i \bmod p, x_i r_q^i \bmod q]$의 스칼라 곱을 계산한다. 이후에 계산한 $[x \bmod p, x \bmod q]$을 더한다. 최종적으로 중국인의 나머지 정리를 사용하여 $x \bmod n$을 구할 수 있다.

3.1.2 Domingo-Ferrer[12]

2002년 Domingo-Ferrer는 "A provably secure additive and multiplicative privacy homomorphism"[12]

에서 안전성의 문제를 보완한 새로운 방식을 제안하였다.

(1) 키 생성 과정

다음을 만족하는 m, m'을 선택한다. 이때, 정수 m은 다음 두 가지 성질을 만족해야 한다.

- 큰 정수로서 많은 수의 약수를 가져야 한다.
- 여러 개의 소수를 반복적으로 곱셈함으로써 많은 수의 약수를 갖도록 한다.

m'는 위에 선택된 m에 대하여 나머지가 없는 m'를 선택하고 r은 위의 m에 대하여 $r\times r^{-1} \bmod m = 1$을 만족하는 r을 선택한다.(이때 큰 정수 m은 많은 수의 r, r^{-1}이 존재하도록 m을 선택) 예를 들어, divisor 9일 때, $2\times 5 = 10 \bmod 9 = 1$이므로 5는 2의 역원이라고 할 수 있다. 평문 a를 d개로 분할시킬 경우 사용되는 d는 3이상의 큰 값을 선택하도록 한다.

[공개키] : m, d

[비밀키] : r, m'

(2) 암호화 과정

$a \in [-m', m']$ 랜덤하게 분할하여

$(a_1, a_2, \ldots, a_n) \in [-m, m]$로 한다. 즉,

$$a = \sum_{i=1}^{n} a_i \bmod m'$$

$$c = E_k(a) = (a_1 r^1 \bmod m, a_2 r^2 \bmod m, \cdots, a_i r^i \bmod m)$$

(3) 복호화 과정

$$(a_1, a_2, \cdots, a_n) = (r^{-1} \bmod m, r^{-2} \bmod m, \cdots, r^{-i} \bmod m)$$

$$a = D_k(c) = \sum_{i=1}^{n} a_i \bmod m'$$

키 생성 및 암복호화 과정에 대한 예를 다음에 나타낸다.

(1) 키 생성 과정

- $m = 416,\ m' = 208$
- $d = 4$
- $r = 7,\ r^{-1} = 119$
- $a = 12,\ b = 14$

평문 a를 d개의 m'로 분할하는 단계

$12 = 2+3+5+2,\ 14 = 1+3+6+4$

(2) 암호화 과정

$$\begin{aligned} E(12) &= ((2\times 7^1) \bmod 416, (3\times 7^2) \bmod 416, \\ &\quad (5\times 7^3) \bmod 416, (2\times 7^4) \bmod 416) \\ &= (14, 147, 51, 226) \end{aligned}$$

$$\begin{aligned} E(14) &= ((1\times 7^1) \bmod 416, (3\times 7^2) \bmod 416, \\ &\quad (6\times 7^3) \bmod 416, (4\times 7^4) \bmod 416) \\ &= (7, 147, 394, 36) \end{aligned}$$

(3) 복호화 과정

$$\begin{aligned} E(12) + E(14) &= (14,147,51,226) + (7,147,394,36) \\ &= (21, 294, 445 \bmod 416, 262) \\ &= (21, 294, 29, 262) \end{aligned}$$

$$\begin{aligned} D(E(12) + E(14)) &= 21\times 119^1 \bmod 416 + \\ &\quad 294\times 119^2 \bmod 416 + \\ &\quad 29\times 119^3 \bmod 416 + \\ &\quad 262\times 119^4 \bmod 416 \\ &= 3+6+11+6 = 26 \end{aligned}$$

그러나, 2002년도에 제안한 방식은 비밀키를 사용자간 미리 공유해야 하며 암호문끼리의 승산횟수에 비례하여 암호문의 크기가 증가하게 되고 알려진 평문 공격에 취약하다는 사실이 2003년 D.Wagner의"Cryptanalysis of an algebraic privacy homomorphism"[14]과 F. Bao의 "Cryptanalysis of an algebraic privacy homomorphism"[15]에서 밝혀졌다. 더욱이, 1996년에 제안한 알고리즘 역시 J.H Cheon[16]에 의해 알려진 평문 공격에 취약하다는 사실이 추가로 밝혀졌다.

3.1.3 NTT Chida 방식[13]

NTT Chida 는 Domingo-Ferrer 암호를 개량한 새로운 대수적 암호이다. 기존은 Domingo 방식은 평문 공간을 비밀로 하는 반면, NTT 방식은 평문공간을 공개하며 기존의 해결하지 못하는 Provably secure에 근거를 두고 있다.

Provably secure는 공격자가 어떠한 것을 행할까를 규정한 공격모델을 가정하고 이 가정하에 어떠한 안전성을 달성할 수 있을까를 증명함으로써 안정성을 논리적으로 보증하는 기법으로 현재 수학적 미해결 문제(소인수분해문제, 이산대수문제)의 곤란성을 적용한 것이라고 할 수 있다.

NTT Chida방식을 구성하기 위해 필요한 Domingo 방식의 개요를 설명하면 다음과 같다.

(1) 적절한 m의 약수 및 m' 및 $r \in Z_m^*$을 선택하여

m', r을 비밀키로 한다

(2) 평문 $a \in Z_{m'}$선택,

$$a = \sum_{i=1}^{d} a_i \bmod m'$$

를 분할하여 d개의 $a_i \in Z_m (i = 1, \ldots, d)$를 생성한다.

(3) 분할된 a_i를 $c_i = a_i r^i \bmod m$로

$$P_{a,r}(X) = \sum_{i=1}^{d} c_i X^i (= E(a))$$

암호화한다.

위의 암호 처리로부터 얻어진 암호문 $P_{a,r}(X)$의 복호처리는 단순히 $P_{a,r}(r^{-1}) \bmod m'$로 하면 된다. 이것에 의해

$$P_{a,r}(r^{-1}) = \sum_{i=1}^{d} a_i = a \bmod m'$$

이 성립한다. 또한 위의 암호처리 절차(2)에 있어서 보안 파라미터(d)는 [12]의 논문에서 $d \geq 3$을 선택하도록 권고한다.

한편, 위의 암호화함수 E에 의해 얻어지는 임의의 암호문 $E(a_1) = P_{a_1,r}(X), E(a_2) = P_{a_2,r}(X)$에 대하여

$$Q^{\pm}_{(a_1,a_2),r}(X) = P_{a_1,r}(X) \pm P_{a_2,r}(X) \bmod Z_m[X],$$
$$R_{(a_1,a_2),r}(X) = P_{a_1,r}(X) P_{a_2,r}(X) \bmod Z_m[X]$$

로 했을때,

$$Q^{\pm}_{(a_1,a_2),r}(r^{-1}) = a_1 \pm a_2 \bmod Z_m[X],$$
$$R_{(a_1,a_2),r}(r^{-1}) = a_1 a_2 \bmod Z_m[X]$$

가 성립하기 때문에 Domingo-Ferrer 암호는 APH (Algebraic PH)이다. 또한 임의의 $e \in Z$ 및 암호문 $E(a)$에 대해서 $eE(a) = E(ea) \bmod Z_m[X]$은 성립하지만 $e + E(a) = E(e+a) \bmod Z_m[X]$은 성립하지 않는다. 즉, 정수배는 계산할 수 있지만 정수합은 분명하지 않다.

NTT Chida 방식은 우선 키를 생성함에 있어 RSA와 같은 방식의 큰 소수 $n = pq$를 선택한 후 Domingo-Ferrer과 같이 r을 선택한다.

(1) $n = pq$를 계산한다.
(2) 보안 파라미터 d를 결정한다.
(3) $a_i \in Z_n$에 대해 공간 $a_i = (a_1, a_2, \ldots, a_n)$을 선택한다.
(4) $a_0 = a - \sum_{i=1}^{d} a_i \bmod n$ 을 계산한다.
(5) $c_i = a_i r^i \bmod n$ 을 계산한다. $(i = 0, \ldots, d)$
(6) $P_{a,r}(X) = \sum_{i=0}^{d} c_i X^i (= E(a))$으로 암호화 한다.

상기 암호처리에 의해 얻어진 암호문 $P_{a,r}(X)$의 복호처리는 Domingo-Ferrer 암호처럼 암호문 $P_{a,r}(r^{-1}) \bmod n$로 하면 된다. 이것에 의해 $P_{a,r}(r^{-1}) = \sum_{i=0}^{d} a_i = a \bmod n$이 성립한다.

여기서 제안방식은 다항식 표현된 암호문이 정수항 c_0를 갖고 c_0는 비밀키 r을 포함하지 않는다 그리고 이것에 의해 $e \in Z$에 대해 $e + E(a) = E(e+a)$ 성립한다. 즉 정수합을 계산할 수 있다.

3.2 Asymmetric 준동형 암호

3.2.1 Generalized Paillier[6]

Generalized Paillier[6] 암호방식은 기본 Paillier[5] 암호방식의 일반화로써 Generalized-Paillier 방식은 $\bmod n^{s+1}$의 계산을 수행한다.

여기서 n은 RSA modulus 이고 s는 자연수이고 기본 Paillier 방식의 경우는 $s = 1$이 되는 경우이다. $Z^*_{n^{s+1}}$에 속하는 위수 $\Phi(n^{s+1})$은 n^s로 나누어진다. 더욱이 $Z^*_{n^{s+1}}$은 $G \times H$의 직접곱으로 표현될 수 있는데, 여기서 G는 위수 n_s의 순회군이고, H는 Z^*_n 의 동형이다.

암호화를 위해 메시지는 G/H에 의해 해당 잉여류로 변환되고 이 방식의 안전성은 H의 다른 잉여류에서 랜덤요소를 구별하는 어려움에 근거한다. 두 개의 랜덤요소가 동일한 잉여류내에 있다면 두요소를 구별하기 어려운 Semantically secure라고 한다.

기본 Paillier처럼 Generalized Paillier의 안전성은 합성 잉여 가정(Decisional Composite Residuosity Assumption)에 의해 증명될 수 있다.

(1) 키 생성 과정
- p, q : 큰 소수 선택
- $n = pq$를 계산, $\lambda = lcm(p-1, q-1)$를 계산
- $g \in Z^*_{n^{s+1}}$를 선택할 때 n과 $x \in H$에 서로 소가되는 $g = (1+n)^j x \bmod n^{s+1}$를 만족하는 g를 선택
- CRT(Chinese Remainder Theorem)을 사용하여 $d \bmod n \in Z^*_n$와 $d = 0 \bmod \lambda$를 만족하는 d를 선택한다.

[공개키] : n, g
[비밀키] : d

(2) 암호화 과정

$m \in Z_{n^s}$은 평문이라 하고, $r \in Z^*_{n^{s+1}}$을 임의로 선택한다.

$c = g^m \cdot r^{n^s} \bmod n^{s+1}$

(3) 복호화 과정

암호문 $c \in Z^*_{n^{s+1}}$ 대해 $c^d \bmod n^{s+1}$를 계산한다.

$c^d = (g^m r^{n^s})^d = ((1+n)^{jm} x^m r^{n^s})^d =$
$(1+n)^{jmd \bmod n^s} (x^m r^{n^s})^{d \bmod \lambda} = (1+n)^{jmd \bmod n^s}$

$m = (jmd) \cdot (jd)^{-1} \bmod n^s$

이때, 복호화에 대한 단순화 과정은 다음과 같다.

- $g = n+1$로 고정되어 있다.
- $d = 1 \bmod n^s$ 와 $d = 0 \bmod \lambda$를 만족하는 d를 계산한다. 이 경우 $c^d = (1+n)^m \bmod n^{s+1}$로 복호화한다.

3.2.2 준동형 공개키 암호화 시스템[18]

준동형 공개키 시스템의 기본 방식은 키 생성(KeyGen), 암호화(Encrypt), 복호화(Decrypt)의 세 가지 알고리즘으로 구성된다.

(1) 키생성(Keygen(τ)) 과정

- 위수 $n = q_1 q_2$인 군 G에서 두 개의 생성자 $g, u \in G$를 랜덤하게 선택한다.
- $h = u^{q_2}$ 계산 (h는 위수가 q_1인 군 G_1의 생성자)

[공개키] $PK = (n, G, G_1, e, g, h)$
[비밀키] $SK = q_1$

(2) 암호화(Encrypt(PK, M)) 과정

- 메시지 m의 메시지 공간을 $\{0, ..., T\}$라고 한다. ($T < q_2$)
- 랜덤값 $r \in Z_n$을 선택하여 메시지 m을 암호화한다. $C = g^m h^r \in G$

(3) 복호화(Decrypt(SK, C)) 과정

- 비밀키 $SK = q_1$로 암호문 C를 다음과 같이 복호화한다.

$C^{q_1} = (g^m h^r)^{q_1} = (g^{q_1})^m$

이 시스템의 안전성은 부분군 결정 문제의 어려움에 기반을 둔다. 이때, 부분군 결정 문제란 합성수 위수 $n = q_1 q_2$을 갖는 군 G의 원소가 주어졌을 경우, 이것이 위수가 q_1인 부분군 G_1에 속하는지를 결정하는 것이다.

- 가법(additive) 준동형성

메시지 $m_1, m_2 \in \{0, 1, ..., T\}$의 암호문 $C_1, C_2 \in G_1$이고 랜덤값 $r \in Z_n$이 주어졌을 때,

$C = c_1 c_2 h^r = m_1 + m_2 \bmod n$

$C = g^{m_1} h^{r_1} \cdot g^{m_2} h^{r_2} = m_1 + m_2 \bmod n$과 같다.

- 승법(multiplicative) 준동형성

g_1은 위수가 n이고 h_1은 위수가 q_1일 때, $g_1 = e(g,g)$, $h_1 = e(g,h)$이고, $\alpha \in Z$ 일때 $h = g^{\alpha q_2}$ 라고 한다. 암호문 C_1, C_2 이고 랜덤값 $r \in Z_n$이 주어졌을 때,

$C = e(C_1, C_2) h_1^{\tilde{r}} = g_1^{m_1 m_2} h_1^{\tilde{r}} \in G_1 = m_1 m_2 \bmod n$과 같다.

$\alpha \in Z$ 일 때 $\tilde{r} = m_1 r_2 + m_2 r_1 + \alpha q_2 r_1 r_2 + r$과 같이 계산되어 진다.

3.2.3 타원곡선 ElGamal 암호(EC-EG)[19]

[17]에서 제안된 Original ElGamal 암호화 방식이 승법 준동형이므로 가법 준동형 암호화를 필요로 하는 DAS모델에서 Aggregation Query 등에 직접적으로 사용될 수 없다. 그러므로 ElGamal 암호화 방식은 additive group으로 변화되어야 한다.

함수 map( )은 평문값을 아래와 같은 식을 만족하는 평문 곡선점 M으로 매핑하기 위해 사용된 결정적 매핑 함수이다. 여기서 $m_1, m_2 \in GF(p)$상의 mod 연산이 필요로 한다.

$$map(m_1 + ... + m_n) = map(m_1) + ... + map(m_n)$$

타원곡선상의 가법연산은 정수의 덧셈을 수행하기 전에 곡선상의 operand를 필요로 하기 때문에 가법연산은 해당 타원곡선 점으로 매핑되어야 한다. 이러한 이유에 의해 매핑함수가 필요하다.

[6]에서 제안된 것처럼 준동형 매핑 함수는 타원곡선의 생성점의 배수 사용에 근거한다. 배수사용의 의미는 매핑함수가 평문 M을 점 mG로 변환하는 것이다. 역매핑함수 rmap( )는 주어진 점 mG로부터 m을 추출한다.

매핑 함수는 다시 말하면,

$map: m \rightarrow mG$

$m \in GF(p)$는 방정식

$$M_1 + M_2 + \ldots + M_n = map(m_1 + m_2 + \ldots + m_n) = (m_1 + m_2 + \ldots + m_n)G = m_1G + m_2G + \ldots + m_nG$$

에 대한 $m_1, m_2, \ldots, m_n \in GF(p)$ 생성점 G와 mod$p$로 부터 성립되는 사실에 의해 요구되는 준동형 함수 특성을 만족한다. 단지, 정수를 타원 곡선점으로 변환시키므로 매핑함수는 보안성과는 관련성이 없다. 이러한 매핑함수는 EC-EG암호화 방식의 보안성을 증가시키거나 감소시키지 않는다.

[비밀키] $x \in GF(p)$

[공개키] E, p, G, Y, 여기서 $Y = xG$고 타원곡선상의 점 $G, Y \in E$를 같은 $GF(p)$상의 타원곡선 E를 선택한다.

(1) 암호화 과정

평문 $m \in [0, p-1]$와 랜덤 $k \in [1, n-1]$가 주워진다. n은 E의 위수이다.

$M = map(m)$

$C = enc(m) = (R, S) = (kG, M + kY)$

(2) 복호화 과정

$M = dec(C) = dec(R, S) = -xR + S$

$m = rmap(M)$

3.3 Stream 기반 준동형 암호[8]

Stream 기반 준동형 암호인 CMT[8]방식은 다음과 같은 과정을 거친다.

먼저 $0 \le a < m, 0 \le k < m$ 만족하는 m(큰 정수) 를 선택한다. 여기서,

k : 대칭 비밀키

a : 암호화 될 메시지

(1) 암호화 과정

$c = Enc_k(a) = a + k \bmod m$

(2) 복호화 과정

$a = Dec_k(c) = Enc_k(a) - k \bmod m$

이 방식은 다음 방정식이 항상 만족하기 때문에 암호화된 값에서 요구되는 가법 준동형 성질을 제공한다.

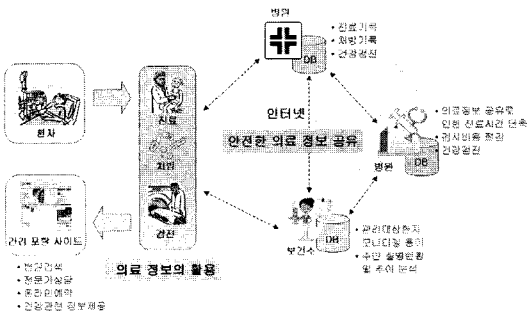
$$Dec_k(c_1 + c_2 + \ldots + c_n) = Dec_k(Enc_{k_1}(a_1) + Enc_{k_2}(a_2) + \ldots Enc_{k_n}(a_2)) = (a_1 + a_2 + \ldots a_n) \bmod n$$

$$(k = k_1 + k_2 + \ldots + k_n) \in [0, m-1] \text{ and } a_1, a_2, \ldots, a_n \in [0, m-1]$$

Ⅳ. u-healthcare에의 응용

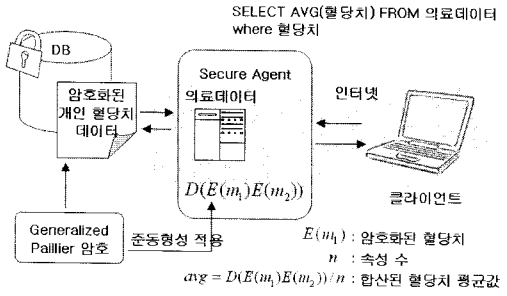
정보기술의 발전으로 다양한 분야에서 융합이 이루어지고 있다. 이러한 분야로 IT와 BT가 결합된 의료분야가 있다. 본 논문에서는 Generalized Paillier 암호화 방식의 의료분야에 대한 적용성에 대해 검토한다.

과거의 의료분야는 병원내에서만 환자의 의료정보화가 이루어졌으며 병원간에는 의료정보가 공유되지 않았다. 그러나 최근 기존 병원 의료정보시스템인 EMR에서 병원간 의료정보화 시스템인 EHR로 발전하게 되어 병원간 환자의 의료정보를 공유하게 된다. 그러나 의료정보를 공유할 경우 개인 의료데이터의 공유를 통해 프라이버시 문제가 발생할 수 있다[7].



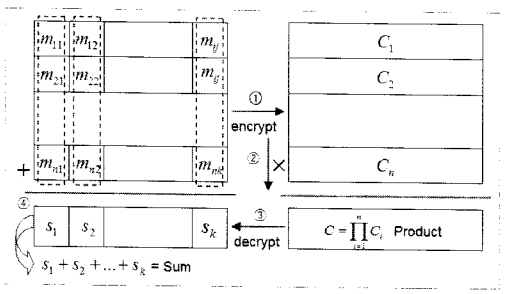
〔그림 2〕 의료정보의 공유와 활용

본 논문에서는 Generalized Paillier 암호의 준동형성을 이용하여 데이터의 기밀성을 보장받을 수 있도록 한다. 예를 들어, 개인의료정보 관련 수치정보인 각종 혈당 통계 수치 및 의료 결과를 추출할 경우, 프라이버시를 침해하지 않토록 암호화된 상태에서 데이터의 연산이 가능한 응용에 대해서 검토한다[그림 3].



〔그림 3〕 의료데이터 평균치계산을 위한
준동형 암호화 적용 모델

[1]의 논문에 의하면 [그림 4]와 같이 평문(왼쪽)의 속성을 나타내는 각 열의 합은 평문을 Generalized Paillier 암호로 암호화(encrypt)한 결과인 각 행$(c_1, c_2, \ldots c_n)$의 곱과 동일하다.



〔그림 4〕 Paillier암호 기반의 평문 암호화

예를 들어 ① 암호화된 3개의 암호화 블록(n=3)이 있을 때, Generalized Paillier의 준동형성을 이용하여 ②와 같이 계산한 후 ③ 복호키를 이용하여 복호화할 경우 ④ 평문의 합과 암호문의 곱 연산 결과가 동일하게 됨을 알 수 있다.

사용되는 기호는 다음과 같다.

m : 평문 블록

m_{ij} : 각 평문의 블록값 i행$(1 \le i \le n)$, j열$(1 \le j \le k)$

C_i : 암호화된 행의 값$(1 \le i \le n)$

K : 복호키

S_i : 같은 비트길이의 암호문$(1 \le i \le k)$

Input : C_i, K

Output : $\sum_{i,j} m_{ij}$

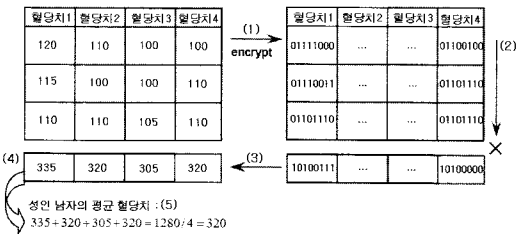
(1) Generalized Paillier 암호방식으로 $C = \prod_{i=1}^{n} C_i$ 계산한다.

(2) 암호문 C를 복호키 K를 사용하여 $S = dec(K, C)$로 복호화 한다.

(3) 복호화된 S는 같은 비트길이로 나눈다. $(s = s_1 \circ s_2 \circ \ldots \circ s_k)$ 이때, S의 개수는 k값을 갖는다.($\circ$: bit string concatenation)

(4) $\sum_{i=1}^{k} s_i$을 출력한다.

(예제) [그림 5]에서 개인의료정보 테이블의 성인 남자에 대한 혈당치 통계를 알고 싶을 경우에 알고리즘 1을 이용한다.



〔그림 5〕 혈당치 통계 예제

(1) 개인의료정보(왼쪽)를 암호화 한다.

(2) 암호화된 암호문(오른쪽)을 Generalized Paillier을 사용하여 혈당치를 계산한다.

(3) (2)에서 계산된 암호문을 복호화 한다.

(4) 복호화된 4개의 블록을 지정된 블록 비트 크기로 분할한 후 혈당치의 평균을 계산한다.

(5) 혈당치를 나타내는 속성의 값이 4개인 경우에는 1280/4=320으로 계산할 수 있다.

따라서, 개인의료정보의 암호화를 통해 환자개인의 프라이버시를 보호할 수 있다.

V. 결 론

유비쿼터스 컴퓨팅 환경이 도래함에 따라 DB 정보 및 관리의 범위는 크게 증가하였다. 따라서 안전한 DB 관리를 위해 DAS 모델에 대한 고려가 필요하게 되어 DB 암호화를 통해 데이터를 관리하게 되었다. 그러나 암호화된 데이터를 관리할 때 효율적 측면에서 많은 검

토가 필요한 실정이다.

따라서 이를 해결하고자 암호화된 상태에서 연산이 가능한 다양한 준동형 암호방식을 검토하였으며, 현재 이슈가 되고 있는 의료분야에 적용가능할 것으로 전망된다. 향후 연구 과제로써 암호화 및 연산과정을 더욱 효율적으로 수행할 수 있는 준동형 암호화 구조에 대한 규명이 이루어져야 할 것이다. 또한, 준동형 암호화 방식의 구현을 통해 의료분야에의 적용가능성을 효율성 및 안전성 측면에서의 검증이 필요하다.

참고문헌

[1] R. Rivest, L. Adleman, M. Dertouzos, "On data banks and privacy homomorphisms, Foundations of Secure Computation," Academic Press, pp. 169-179, 1978.

[2] H. Hacigumus, B. R. Iyer, and S. Mehrotra, "Providing database as a service," In ICDE, March 2002.

[3] H. Haciqumus, B. R. Iyer, C. Li, and S. Mehrotra, "Executing SQL over encrypted data in the database-service-provider model," Proc. of the ACM SIGMOD Conf. on Management of Data, Madison, Wisconsin, June 2002.

[4] Tingjian Ge, Stan Zdonik, "Answering Aggregation Queries in a Secure System Model," VLDB '07, September 23-28, 2007.

[5] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes" In EuROCRYPT'99, 1999.

[6] I. Damgard M. Jurik, "A generalisation, a Simplification and Some applications of Paillier's Probabilistic public-key System," Public Key Cryptography, pp. 119-136, 2001.

[7] E. Mykletun, J. Girao, and D. Westhoff. Public Key Based Cryptoschemes for Data Concealment in Wireless Sensor Networks. In IEEE Int. Conference on Communications ICC, Istanbul, Turkey, June 2006.

[8] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient Aggregation of Encrypted Data in Wireless Sensor Networks," Mobiquitous, 2005.

[9] 三浦志保, 渡辺知惠美, "管理者に對しても機密を保持できる暗号化データベースの索引構成法", DEWS2007 E7-8, 2007.

[10] O. Ugus et al, "Performance of Additive Homomorphic EC-ELGamal Encryption for TinyPEDS," Technischer Bericht der RWTH Aachen ISSN 0935-3232, Germany, July, 2007.

[11] J. Domingo-Ferrer, "A new privacy homomorphism and applications", Information Processing Letters, vol. 60, no. 5, pp. 277282, Dec. 1996.

[12] J. Domingo-Ferrer, A provably secure additive and multipliactive privacy homomorphism. ISC2002, LNCS. vol. 2443. Springer-Verlag, Berlin, pp. 471-483, 2002.

[13] Koji Chida, An Algebraic Privacy Homomorphism based on Factoring and Its Applications (In Japanese), Technical Report of IEICE, ISEC 2003-59, 2003.9.

[14] D. Wagner, Cryptanalysis of an algebraic privacy homomorphism, ISC2003, LNCS, vol. 2851, Springer-Verlag, Berlin, pp. 234-239, 2003.

[15] F. Bao, Cryptanalysis of a provable secure additive and multiplicative privacy homomorphism, International Workshop on Coding and Cryptography (WCC), pp. 43-50, 2003.

[16] J. Cheon, W. Kim, H. Nam, Known-plaintext cryptanalysis of the Domingo-Ferrer algebraic privacy homomorphism scheme, Information Processing Letters, vol. 97, pp. 118-123, 2006.

[17] T. E. Gamal. A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms. In Advances in cryptology Proceedings of CRYPTO 84, volume 196 of Lecture Notes in Computer Science, pages 10-18. Springer-Verlag New York, Inc., 1985.

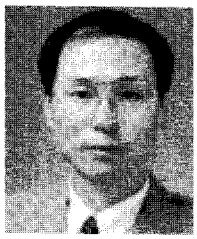
[18] Dan Boneh, Eu-Jin Goh, and Kobbi Nissim, "Evaluating 2-DNF Formulas on Ciphertexts", Springer, TCC, Volume 3378 of LNCS, 2005, pp. 325-341.

[19] O. Ugus, A. Hessler, and D. Westhof, "Performance of additive homomorphic EC-Elgamal encryption for TinyPEDS," 6. Fachgespräch "Drahtlose Sensornetze", Tech. Rep., July 2007.

〈著者紹介〉

박 광 용 (Kwangyong Park)
학생회원

2008년 2월: 동국대학교 전자상거래학과 졸업
2008년 3월~현재: 동국대학교 석사과정(전자상거래 기술전공)
<관심분야> 암호이론, 데이터베이스 보안, 유비쿼터스 프라이버시 보호



송 유 진 (Youjin Song)
정회원

1982년 2월: 한국항공대학교 전자공학과 졸업
1987년 8월: 경북대학교 대학원 정보시스템학과 석사
1995년 3월: 일본 Tokyo Inst-itute of Technology 정보보호학과 박사
1988년~1996년: 한국전자통신 연구원 선임연구원
2003년~2005년: 미국 University of North Carolina at Charlotte 연구교수
2006년 7월~8월: 일본 정보보호대학원대학 객원교수
1996년~현재: 동국대학교 정보 경영학과/대학원 교수
2005년~현재: 동국대학교 부설 전자상거래연구소 소장
1998년~현재: 한국정보보호학회 이사
2006년~현재: 국제e-비즈니스 학회 이사
2006년~현재: 한국사이버테러 정보전학회 이사
2001년: ICISC2001 운영위원장
2003년: 하계CISC2003 프로그램 위원장
2006년: CISC-S2006 공동프로그램 위원장
2007년: 한국정보시스템학회 추계학술발표대회 공동 조직 위원장
<관심분야> Secret Sharing, Privacy Protection, 전자상 거래응용 보안 (Location Privacy, 디지털컨텐츠 보호, SCM/CRM 보안 등), Context Aware Application Security