通信情報保護學會誌
第4卷 第1號, 1994. 3

통계 데이타베이스의 보호에 관한 조사 연구

김 철*

요 약

정보화 사회에서는 가계, 기업, 정보 등의 정보 활동의 주체들이 가진 정보자산은 데이타 베이스(이하 DB)와 소프트웨어(S/W)로 대별할 수 있으며, 이중 DB는 정보화 사회의 기반시설의 하나라고 볼 수 있다. 특별히 통계 DB는 각 주체들에게는 필수적인 정보를 갖고 있다. 금융자산의 정보, 국방에 관련된 병력, 장비, 군수물자 등의 정보, 회계정보 뿐 아니라 인구센서스, 경제계획수립 등등의 다양한 분야에 이 통계 DB는 사용되고 있다. 이러한 통계 DB는 기존의 DB에서의 데이타의 저장, 관리, 추출 기능 외에 통계적인 데이타의 분석기능이 요구되고 있다. 통계처리를 위한 데이타베이스관리 시스템(DBMS, database management system)은 주로 기존의 DBMS에 통계 처리를 위한 기능을 추가하거나 통계를 위한 DBMS를 따로 구축하는 방법을 사용하고 있다. 따라서 일반적인 DB 보호 기술과 더불어 통계 DB의 환경을 이해하는 보호 기술이 요구되고 있다.

일반적으로 DB를 보호하는 방법으로는 물리적인 보안(physical security)과 운영체계 보안(operating system security)이 있으며, 이들과 함께 데이타 암호화(data encryption)의 방법을 사용하고 있다. DB의 보안 방법에 관한 연구 중 George I. Davida 등에 의한 방법은 중국인의 나머지 정리(chinese remainder theorem)를 사용하는 암호화 알고리즘을 이용하여 레코드(record) 단위의 암호화를 하며, Khamis A. Omar 등에 의한 방법은 읽기(read), 쓰기(write), 갱신(update)의 3단계의 사용자 등급을 부여하여 DB접근의 제약을 가하는 기능을 갖고 있다.

본고에서는 특히 그 중요성이 더해가고 있는 통계 DB의 일반적인 개념을 살펴보며, 특성 지향형 질의 모델(characteristic-specified query model)의 보호기술을 살펴본다. 특별히 본고는 통계 DB의 보호에 대한 일반적인 조사 연구로서 잘 알려진 사실들을 많은 참고 문헌과 더불어 소개하는 내용으로 통계 DB의 보호에 관한 새로운 연구 결과는 아니다.

1. 동기 : 통계 DB의 안정성

통계 DB는 주로 다음 두 가지 목적을 갖고 있다. 개개의 항목에 관한 정보를 제공하는 것과 통계적인 분석을 위하여 정보를 취합하는 것이 바로 그것인데 이 목적은 서로 상반된 요소를 포함하고 있다. 즉 허가된 사용자에게만 개개의 정보를 제공하는 것과 동시에 개개의 정보에 접근이 허락되지 않은 사용자가 포함되어 있을지도 모르는 다수에게 통계적인 분석을 위한 정보를 제공한다는 것이다. 이 상반된 면의 틈새에서 바로 통계 DB의 안전성 문제가 나타날 수 있다.

통계 DB의 안정성에 관한 핵심적인 면은 다음과 같은 질문으로 나타낼 수 있다.

"허락된 질의(query)들에 대한 통계적 분석을 담고 있는 응답을 이용하여, DB에 저장되어 있는

* 정회원, 광운대학교 이과대학 수학과 조교수

비밀정보의 값을 알아내는 것이 가능한가?"

일반적으로는 이것은 불가능하다고 여길지도 모른다. 즉 통계 DB에서 개개의 정보에 대한 접근은 차단되기 때문에 허락된 질의를 통하여 개개의 비밀정보를 밝힐 수 없다고 생각할 수 있다. 그러나 다음 예를 통하여 통계 DB가 갖고 있을 수 있는 취약성을 살펴보도록 하자. 다음 표와 같은 가상적인 군사정보 DB를 가정하자.

연령	부대명	육군/해병대	주둔지역	주둔기간(연)	병력(명)
1	D부대	육	강릉	2	850
2	B부대	해	속초	3	800
3	M부대	해	동해	1	700
4	S부대	육	고성	2	900
5	E부대	해	강릉	2	860
6	C부대	육	동해	4	950
7	U부대	육	속초	3	900
8	R부대	해	고성	4	950
9	I부대	해	강릉	1	750
10	T부대	육	속초	5	950
11	Y부대	해	동해	5	850

이 DB의 질의는 다음 세가지 형태만이 허락된다고 가정하자.

1. 일반적인 정량적 질의 : 즉 강원 지역에 주둔하고 있는 부대의 수에 관한 질의등
2. 평균 표준 편차 등의 통계적 파라메터에 관한 질의 : 즉 11개 부대의 평균 병력에 관한 질의등
3. AND, OR, NOT과 같은 불리안 연산(Boolean operation)에 의한 질의의 결합

물론 T 부대의 병력수는 얼마인가와 같은 직접적인 질의는 허락되지 않는다. 위의 보기에 DB에서 가장 중요한 비밀정보는 병력수이기 때문이다. 이제 몇 번의 질의를 통하여 T부대의 병력수를 밝혀낼 수 있음을 보인다.

질의 1 : 육군 부대의 수는?
응답 1 : 5개
질의 2 : 얼마나 많은 육군 부대가 속초지역에 주둔하고 있는가?
응답 2 : 2개
질의 3 : 5년동안 속초지역에 주둔한 육군 부대의 수는?
응답 3 : 1개
질의 4 : 5년동안 속초지역에 주둔한 육군 부대 중 병력 900명 이상의 부대 수는?
응답 4 : 1개

이제 T부대의 병력의 수는 900명 이상이라는 것이 밝혀졌다. 물론 응답 3의 결과 이 일련 질의들의 목표는 T부대로 모아지고 있음을 알 수 있다. 질의 4와 같은 질의를 1050명이상, 또는 930명 이상, 다시 970명 이상이라는 식으로 좁혀 나감으로써 T부대의 병력을 정확히 알아낼 수 있다. 이에 대응하기 위한 단순한 방법은 적은 수의 레코드에 관한 질의는 거부하는 기능을 갖도록 하는 것일 수도 있다. 그러나 이 또한 별로 소용이 없음을 다음과 같은 질의들로 알 수 있다.

질의 1 : 해병대 전체 병력의 수는?
응답 1 : 4910명
질의 2 : 5년 동안 속초 지역에 주둔한 육군의 병력의 수와 해병대 전체 병력의 수의 합은?
응답 2 : 5860명

이 두 응답의 차가 바로 T부대 병력의 수가 된다. 역시 이에 대응하기 위하여 통계 DB가 위와 같이 절대값(absolute value)인 병력수 자체를 알려주지는 못하게 하고 평균등과 같은 통계값만을 알려준다고 하자. 그러나 이 또한 T부대의 병력을 다음과 같이 밝혀낼 수 있다.

질의 1 : 해병대 부대의 평균 병력의 수는?
응답 1 : 818.33명
질의 2 : 5년 동안 속초 지역에 주둔한 육군의 병력의 수와 해병대 전체 병력의 평균은?
응답 2 : 837.143명

질의 3 : 해병대 부대의 수는?

응답 3 : 6개

여기서 마지막 질의는 공개되는 정보이므로 질의가 필요없다. 이 경우 T부대의 병력수를 쉽게 계산할 수 있다.

T부대의 병력수=

(응답2)×(응답3+1)−(응답1)×(응답3)

통계 DB에서 비밀 정보를 찾아내는 일반적인 방법은 호프만(L. J. Hoffman)과 밀러(W. F. Miller)에 의해 1970년 처음 제안 되었는데 그 알고리즘을 살펴보면 다음과 같다.

먼저 P_1, P_2, ..., P_m을 특성(trait)들의 집합이라 하고, $\#(P_1 \& P_2 \& \ldots \& P_m)$는 이 특성들을 모두 갖는 항목들의 수라고 정의하자. $\#(P_i)$에는 응답을 하지만 특성 P_i를 가진 개개의 항목에 관한 정보 즉, 신분이나 이름등은 응답을 하지 않는다고 가정하고, 아울러 $\#(P_1 \& P_2 \& \ldots \& P_m)$과 같은 여러 특성들을 공통적으로 같는 항목의 수에 관하여는 응답을 한다고 가정하자. 이와같은 가정들 하에서 P_1, P_2, ..., P_m 특성을 포함하는 항목이 P_0 특성을 포함하고 있는지의 여부를 결정하는 알고리즘을 살펴보자.

단계 1 : $\#(P_1 \& P_2 \& \ldots \& P_m)=1$인지를 검증한다. 즉, 특성 P_1, P_2, ..., P_m을 모두 포함하는 항목이 1개인지를 검증한다.

단계 2 : $\#(P_1 \& P_2 \& \ldots \& P_m \& P_0)$를 검증하여 그 결과가 1이면, 질의 대상인 항목은 특성 P_0를 갖고 있는 것이며, 그 결과가 0이면 질의 대상인 항목은 특성 P_0를 갖고 있지 않는 것이 된다.

단계 3 : 단계 2의 $\#(P_1 \& P_2 \& \ldots \& P_m \& P_0)$의 결과가 1보다 큰 값을 가진다면 질의 대상인 특별한 항목에 대하여 그것이 특성 P_0를 갖고 있는지를 알 수 없다. 그러나 만약에 $\#(P_1 \& P_2 \& \ldots \& P_m)=\#(P_1 \& P_2 \& \ldots \& P_m \& P_0)$라면 특성 P_0는 질의 대상 항목을 상징한다고 볼 수 있다.

단계 4 : 그러나 만약에 $\#(P_1 \& P_2 \& \ldots \& P_m) > \#(P_1 \& P_2 \& \ldots \& P_m \& P_0)$라면 질의 대상인 특별한 항목에 대하여 그것이 특성 P_0를 갖고 있는지를 알 수 없다.

위 알고리즘은 그 후 계속 향상되어 왔고, 이에 대한 많은 논문들이 발표되고 있다.

2. DB의 일반적인 보호 기술

2.1 접근 단계별 보호 기술

가. 물리적 보호

컴퓨터 시스템을 보유하고 있는 각 사이트는 외부의 침입자로부터 물리적으로 보호하기 위한 수단으로 출입구를 통제하거나 비밀 출구를 갖추어야 한다.

나. 인적 보호

권한이 있는 사용자가 뇌물을 받거나 다른 수단으로 인하여 침입자에게 액세스하도록 하는 가능성을 배제하기 위하여 사용자에 대한 권한부여를 신중하게 검토하여야 한다.

다. 운영체제 보호

DB시스템 내부의 보안에도 불구하고, 운영체제가 지니고 있는 보안상의 약점이 데이타를 권한없이도 액세스하게 할 수 있다. 거의 모든 DB 시스템은 단말기나 통신망을 통하여 원격 액세스(remote access)가 가능하므로 운영체제 내부의 소프트웨어 레벨의 보안 역시 물리적 보안만큼 중요하다. 운영체제 레벨의 보안은 시스템 액세스를 위한 패스워드(password)로부터 동시에 실행되는 프로세스의 고립화에 이르기까지 여러 레벨에서 구현되어진다. 화일 시스템 역시 어느 정도의

보안 기법을 제공한다.

라. DB 시스템 보호

DB 시스템에 대한 부분적인 권한이 주어진 사용자는 데이타베이스의 한정된 부분에 대해서만 액세스할 수 있다. 다른 사용자는 이 데이타에 대한 질의를 할 수 있으나, 이 데이타를 수정할 수는 없다. DB 시스템이 이러한 제약조건을 유지하는 책임이 있다.

2.2 권한부여와 뷰

뷰는 사용자에게 그 사용자가 사용하지 않는 데이타를 감추기 위한 수단이다. 데이타를 감출 수 있는 뷰의 개념이 시스템의 사용을 간단하게 할 뿐만 아니라 보안을 증진시킬 수 있다. 사용자에게는 관심이 있는 데이타만을 주의집중하게 함으로써 사용자는 시스템을 간단하게 이용할 수 있다. 사용자들에게 자기의 뷰만을 사용하게끔 제약을 부과함으로써 보안이 유지되도록 한다.

가. 릴레이션

사용자에게 한 릴레이션에 대한 액세스를 직접 허용할 수도 있으며 거절할 수도 있다.

나. 뷰

사용자는 자기의 뷰에 나타난 데이타를 액세스할 수도 있고, 또 액세스할 수 없는 경우도 있다.

사용자는 한 릴레이션 전체를 직접 액세스할 수는 없더라도 뷰를 통하여 이 릴레이션의 일부분을 액세스할 수는 있다. 그러므로 릴레이션 레벨의 보안과 뷰 레벨의 보안이 결합하여, 어떤 사용자에게 그가 필요한 데이타만을 정확하게 액세스하도록 할 수 있다. 어떤 사용자는 DB의 일부분에 대하여 아래와 같이 여러가지 형태의 권한을 가질 수 있다.

가. 판독(read)권한 : 데이타의 판독은 가능하나 수정할 수는 없다.

나. 삽입(insert)권한 : 새로운 데이타를 삽입할 수 있으나 존재하는 데이타를 수정할 수는 없다.

다. 갱신(update)권한 : 데이타의 수정은 가능하나 삭제할 수는 없다.

라. 삭제(delete)권한 : 데이타를 삭제할 수 있다.

어떤 사용자는 위와 같은 데이타의 액세스 권한뿐만 아니라, 아래와 같은 데이타베이스의 스키마를 수정할 수 있는 권한도 부여받을 수도 있다.

가. 색인에 대한 권한 : 색인을 생성하거나 삭제할 수 있다.

나. 자원(resource)에 대한 권한 : 새로운 릴레이션을 생성할 수 있다.

다. 변경(alteration)할 수 있는 권한 : 어떠한 릴레이션내의 속성을 삽입하거나 삭제 할 수 있다.

라. 포기(drop)할 수 있는 권한 : 릴레이션을 삭제할 수 있다.

삭제(delete)할 수 있는 권한은 단지 튜플(tuple) 삭제만을 허용한다는 점에서 포기할 수 있는 권한과 다르다. 만일 어떤 사용자가 한 릴레이션의 모든 튜플을 삭제하더라도, 이 릴레이션은 여전히 존재하게 되며, 이 때에 이 릴레이션은 단지 아무런 튜플이 없는 상태가 된다. 그러나 한 릴레이션이 포기된다면 더 이상 이 릴레이션은 존재하지 않게 된다. DB 관리자가 이와같은 권한부여를 관장한다. DB 관리자는 새로운 사용자에게 권한을 부여할 뿐만 아니라 DB 구조를 새롭게 할 수 있다. 이러한 형태의 DB 관리자의 권한은 운영체제에서 수퍼 사용자나 오퍼레이터에게 부여되는 권한과 유사하다. 일부 형태에 권한을 부여받은 사용자는 다른 사용자에게 이들 권한을 부여할 수 있다. 그러나 가까운 장래에 권한이 취소될 수 있으므로 사용자간에 부여받은 권한을 적절하게 관리할 필요가 있다.

2.3 무결성 제약조건

무결성 제약조건은 권한이 있는 사용자가 데이타의 일관성을 손상시키지 않고 DB를 변경할 수 있는 수단을 제공하며, 따라서 무결성 제약조건은 DB를 우발적인 위험으로부터 보호해 준다. 무결성 제약조건은 DB에 관련되는 술어 형태로 표현된다. 그러나 무결성 제약조건을 검사하기 위해서는 많은 비용이 요구된다. 그러므로 적은 부가노력으로 검사할 수 있는 무결성 제약조건을 고려한다. 이것은 종속성 유지분해에 의해서 릴레이션 스키마를 분해하는 과정에서 유지되어야 한다.

2.4 영역제약 조건

쉽게 검사할 수 있는 제약조건 중의 하나로 영역 제약조건이 있다. 영역제약 조건을 검사하는 일은 프로그래밍 언어에서 실행시에 데이타의 형을 조사하는 일과 유사하다. 영역제약 조건과 밀접하게 관련되어 있는 제약조건의 한 형태는 빈값(null value)을 삽입하도록 하는 것이다. 특정한 속성에는 빈값을 허용하지 않으며 다른 속성에는 빈값을 허용할 수 있다.

2.5 암호화

DB 시스템에 권한을 부여하는 여러가지 기능들만 가지고서는 아주 민감한 데이타에 대한 보호가 불충분할 수 있다. 이러한 경우, 데이타는 암호화(encryption)되어야 한다. 암호화된 데이타는 이 데이타에 대한 암호해독(decryption)방법을 알지 못하는 사용자에게 읽혀질 수 없다. 데이타를 암호화하는 많은 기법들이 있다. 암호화하는 기법이 너무 간단하면 권한이 없는 사용자도 암호화된 데이타를 쉽게 해독할 수 있기 때문에 충분한 보안을 유지해 줄 수 없다. 비록 데이타를 암호화된 상태로 저장하였을지라도 암호화된 정보시스템의 가장 취약점은 분류된 데이타가 평문 상태로 처리되기 때문에 누출(leak)의 위험이 있으며, 비교나 검색과 같은 특정의 연산은 기존의 알고리즘을 가지고 암호화했을 경우에도 복호화가 필요없이 연산이 가능할 수 있으나 대부분은 복호화를 한후에 연산이 가능하다.

따라서 DB 연산을 높이기 위해서는 암호화된 상태에서 연산이 가능한 암호시스템이 바람직하며, 최근에는 레시듀코드(Residue Code)를 사용하여 암호화된 상태에서 데이타를 복호화하지 않고 간단한 산술연산을 할 수 있는 암호시스템에 대한 방법이 제안되었다.

3. 통계 DB의 보호 기술

3.1 모델 설정

통계 DB를 질의의 형태에 따라 두 종류의 모델로 나누어 보도록 한다. 두 모델 모두 실제의 통계 DB를 단순화 시킨것이지만, 보호 측면의 특징들은 그대로 갖고 있는 모델들이다.

처음 모델을 *DC*라고 표기하자. 이는 특성에 중점을 두는 모델로서 DB를 단순하게 길이가 k인 키(key)로 부터 실수 R로 가는 함수로 표현할 수 있다. 여기서 길이 k는 계속 일정하며, 최소한 2이상이다. 즉 다음과 같이 표현될 수 있다.

$$DC : \{0, 1\}^k \rightarrow R$$

예를 들면, $k=2$인 경우 $\{0, 1\}^k = \{00, 01, 10, 11\}$이다. 그리고 *DC*는 이들 모두를 실수로 대응하거나 그 일부만을 대응시킨다. 따라서 어떤 키 v에 대하여 $DC(v)$는 대응하는 값이 있을 수도 있고, 없을 수도 있다. 바로 이 *DC*의 대응 영역이 우리가 관심있게 보아야 할 비밀 정보이다. 특성 중점 모델의 질의를 f라고 하면, 그 형태는 $\{0, 1, *\}^k$로 구성되어진 일련의 스트링으로 나타낼 수 있다. 물론 여기서 f는 통계값, 중간값, 최대, 최소 등의 많은 임의의 변수의 함수이다. 또한 DB 관리자(즉, 시스템)에게 $\{0, 1, *\}^k$로 구성된 f형태

의 질의 q를 제시하면, q와 대응되는 $\{0, 1\}^k$안에 있는 키 v들을 모두 찾아서 그 대응값 $DC(v)$들에 f를 적용한다.

예를 들어 이 DC 모델을 살펴보자. 먼저 다음과 같은 가상적인 모델을 생각하자.

신청 학과목수	학 년	입학년도	성 적
4	1	1992	0.0
3	박사과정	1989	3.9
5	3	1991	2.3
6	2	1991	3.6
4	석사과정	1992	3.0
5	4	1992	2.5
6	2	1990	3.3
6	3	1988	2.0

이 표는 다음과 같은 형태로 변환되어진다.

신청 학과목수	학 년	입학년도	성 적
0100	001	01110	0.0
0011	110	01011	3.9
0101	011	01101	2.3
0110	010	01101	3.6
0100	101	01110	3.0
0101	100	01110	2.5
0110	010	01100	3.3
0110	011	01010	2.0

따라서 학생들은 15개 학과목까지 신청할 수 있고, 학년등은 8가지 영역으로, 그리고 입학년도는 1979년도-2009년도를 다룰 수 있다. 여기서 1991년에 입학한 학생들의 평균성적을 묻는 질의가 있다면, 이는 *******01101로 나타낼 수 있고, 6개 과목을 신청한 학생들의 평균을 묻는 질의라면 0110*******으로 나타낼 수 있다.

두 번째 모델을 DK라고 표기하자. 이는 키에 중점을 두는 모델로서 DB를 단순하게 $\{1, 2, \ldots, N\}$으로부터 실수 R로 가는 모든 대응관계가 성립하는 함수로 표현할 수 있다. 즉 다음과 같이 표현될 수 있다.

$$DK : \{0, 1, \ldots, N\} \rightarrow R$$

여기서 N은 DB의 항목의 수이다. 질의 i의 결과가 $DK(i)$가 되며, 그것은 바로 비밀 정보가 되는 것이다. 당연히 질의 i가 오직 한 항목만을 목표로 하고 있으면, 그 응답은 거부되는 것이다. 따라서 이 DK 모델에서 형태 f의 k-질의는 k개의 첨자로써 $(i_1, i_2, \ldots, i_k)$로 표현된다. 물론 여기서 f는 k개의 변수를 갖는 함수이고, 형태 f의 질의 $(i_1, i_2, \ldots, i_k)$의 결과는 $f(DK(i_1), DK(i_2), \ldots, DK(i_k))$로 표현된다. 또한 질의의 수열은 j와 m이 다를때, i_j와 i_m이 달라야 질의가 받아들여 진다. 예를 들어 다섯명의 학생의 성적의 평균을 내는 경우 질의는 학생의 이름이라는 키를, 예를 들면 (김, 이, 박, 조)등을 사용하여 평균을 내는 함수에 대입하게 된다.

이제 첫째 모델을 살펴보기 전에 이 두 개의 모델과 관련된 보안성을 먼저 정의하여 보자. DB의 특성상 사용자는 DB의 항목의 몇 가지를 알 수 있다. 예를 들어 어느 학생의 이름, 입학년도등은 알고 있으며, 이를 이용하여 개개인의 성적등을 알려고 할 것이다. 이를 고려하여 보안성을 다음과 같이 정의한다.

먼저 D_0를 사용자가 DB내에서 이미 알고 있는 키들의 집합이라고 하자. 즉, 이 D_0는 DC 모델에서는 $\{0, 1\}^k$부분집합이 될 것이며, DK 모델에서는 $\{0, 1, \ldots, N\}$의 부분 집합이 될 것이다. 그리고 Q_m을 m개의 질의들의 순열이라고 하자.

Q_m에 있는 질의들에 대한 응답으로 전에는 몰랐던 키 x들의 값 $DC(x)$ 또는 $DK(x)$가 알려진다면, 이 DB는 위태롭게 된다고 한다. 모든 m에 대하여 어느 m질의에 대하여서도 DB가 위태롭게 되지 않을 때, 이 DB를 안전하다고 한다.

3.2 특성 지향형 질의 모델의 사례

주어진 통계 DB에 대하여 합을 계산하라는 질의를 가진 모델의 안전성을 나타내는 사례를 들어

본다. D_0가 공집합(empty set) 즉, DB의 어떤 엔트리들도 알려져 있지 않다고 가정하고 합만을 요구하는 질의에 대하여 응답한다고 하자. 즉 형태 f가 여기서는 SUM으로 표현되는 질의라 하자. 즉 $\{0, 1, *\}^k$의 질의에 대한 응답은 $DC(w)$들의 합으로 표현된다. 여기서 w는 질의 q와 일치되는 키이다. 예를 들기전에 먼저 정리를 살펴본다. 우선 DB가 p-질의에 의하여 위태롭게 된다면 $(p-1)$-질의에 대하여서도 위태롭게 된다. 따라서 오직 1-질의에 관하여서만 주의를 기울여도 된다.

❖ 정 리

DB의 어느 한 요소도 알려져 있지 않다고 가정하자. 즉, D_0가 공집합이라고 하자. 그렇다면 SUM 형태의 p-질의에 의하여 침해받을 DB는 없다.

이 정리는 모순을 유도함에 의하여 증명할 수 있으나 생략하기로 한다. 다만 그 증명과정에서 도출되는 $WT(v)$와 관련된 것만을 살펴보자. $WT(v)$는 키 v의 무게로 1의 개수가 된다. 이 $WT(v)$를 이용하여 새로운 DC를 정의할 수 있다. 즉 $WT(v)$가 0일 때는 새로운 $DC(v)$를 원래의 $DC(w)+1$, $WT(v)$가 1일 때는 새로운 $DC(v)$를 원래의 $DC(w)-1$으로 정의한다. 이 경우 새로운 DC에 부여되는 1-질의에 대한 응답은 원래의 DC에 부여되는 1-질의에 대한 응답과 다름이 없다. 이제 하나의 예를 들어 SUM 형태의 질의에 안전함을 보이자.

키 v	$DC(v)$	키 v	$DC(v)$
000	1	100	5
001	2	101	6
010	3	110	7
011	4	111	8

언급한 대로 p-질의들은 1개의 질의로 표현될 수 있다. 예를 들면, 2-질의 *1*은 3+4+7+8=22로 응답되는데 이는 *10(3+7=10을 응답)과 *11(4+8=12를 응답)의 두 개의 1-질의의 결합으로 볼 수 있다. 여기에는 전부 12개의 1-질의가 있다. 각각의 질의 및 이에 대한 응답은 다음과 같다.

*00	(6)	0*0	(4)	00*	(3)
*01	(8)	0*1	(6)	01*	(7)
*10	(10)	1*0	(12)	10*	(11)
*11	(12)	1*1	(14)	11*	(15)

이제 새로운 DB의 1-질의에 대한 응답을 살펴봄으로써 이 특별한 사례의 통계 DB 안전성을 보인다. 위에서 언급한 대로 $WT(v)$를 찾음으로 새로 구성된 DB는 다음과 같다.

키 v	$DC(v)$	키 v	$DC(v)$
000	1+1=2	100	5−1=4
001	2−1=1	101	6+1=7
010	3−1=2	110	7−1=6
011	4+1=5	111	8−1=7

이 경우 12개의 1-질의에 대한 새로운 응답들을 보면, 아래와 같이 원래의 응답들과 일치함을 알 수 있다.

*00	(6)	0*0	(4)	00*	(3)
*01	(8)	0*1	(6)	01*	(7)
*10	(10)	1*0	(12)	10*	(11)
*11	(12)	1*1	(14)	11*	(15)

4. 결 론

보안을 유지하는 또 다른 기법으로는 데이타 오염(data pollution)기법이 있다. 이 기법은 질의에 대한 응답 수치를 불규칙적으로 위조하는 것이다. 단, 이 위조는 응답에 대한 통계적 의미가 손상되지 않는 범위하에서 이루어져야 한다. 이와 유사한 기법으로는 질의 자체를 불규칙적으로 수정하는 것이 있다. 이들 두 기법은 정확성과 보안유지면에서 상반되는 장단점을 가지고 있다. 통계적 데이타를 보안하는 기법에도 불구하고, 비도덕

적인 사용자가 정보를 추론해 낼 수 있다. 그러나, 좋은 기법을 사용함으로써 비도덕적인 사용자가 타인 소유의 비밀을 알려고 하는 경우에 많은 시간과 경비가 들도록 만들어서 고의적 의도 자체를 포기하도록 할 수 있다. DB보안에 관한 연구의 역사적 진행 과정에서 볼때 DB의 완벽한 보안유지란 매우 어려운 문제임을 알 수가 있다. 가장 주된 이유는 DB의 기본기술이 먼저 개발된 이후에 보안 문제를 고려해 왔었기 때문이다. 즉, DB기술이 DB 보안 기술보다 훨씬 앞서 갔었다. 이러한 기술 격차를 줄여나가는 방법으로 새로운 DB 기술을 보안 관련 기술과 병행하여 연구하는 것이 국내의 현실에서 볼 때 가장 바람직할 것이다. 아울러 통계 DB라는 특성화된 DB의 적절한 보호에 관한 연구가 필요하다고 본다.

참 고 문 헌

[1] Davida, G. I., Wells, D. L. Kam, J. B., "A database Encryption system with subkeys", ACM Trans. on Database system, Vol. 6, No. 2, Jun 1981.

[2] Demilo, R.A., Dobkin, D.P. and Liption, R.S. "Even Data Bases That Lie Can Be Compromised", IEEE Transaction on Sofware Eng., Vol SE-4(1) Jan. 1977, pp. 73-75.

[3] Demilo, R.A., and Dobkin, D. "Recent Progress in Secure Computation", IEEE 1978, pp. 209-214.

[4] Denning, D.E. and Denning, P.J. : "Data Security", Computing Surveys, Vol. 11, No. 3, Sep. 1979, pp. 227-248.

[5] Denning, D.E., Denning P.J. and Schwartz, M.D. "The Tracker : A Threat to Statistical Data Bases Security", ACM Transaction on Data Bases Systems, Vol. 4(1), March. 1979, pp. 76-96.

[6] Denning, D.E. and Schlorer, J. "A Fast Way For Finding A Tracker in Statistical Data Bases", ACM Trans. on Data Bases System, Vol.5, No. 1, 1980, pp. 88-102.

[7] Denning, P.J. "Cryptography and Data Security", Addison-Wesley, Reading Massachusetts, 1982.

[8] Denning, D.E. and Schlorer, J. "Inference Control For Statistcal Databases", Computer, July 1983, pp. 69-82.

[9] Denning D.E., Cryptography and Data Security, Addison-Wesley, Reading, MA, 1982.

[10] Dobkin, D., Jones, A.K, and Lipton, R.J. "Secure Data Bases Protection Against User Inference", ACM Trans. on D.B. Systems, Vol. 4(1), Mar. 1979, pp. 97-106.

[11] Garvey, T.D. and Lunt, T.F., "Multilevel Security for knowledge-based systems", Proc. EISS Workshop on Database Security, European Institute for system Security, Karlsruhe, Germany, April 1990.

[12] Graham, G.S. and Denning, P.J., "Protection : Principles and Practice", Proc. Spring Joining Computer Conf., Vol.40, AFIPS Press, Montvale, NJ, 1972.

[13] Hinke, T.H., "DBMS trusted computing base taxonomy", Proc. Third IEIP

Workshop on Database Security, Septmber 1989.

[14] Hinke, T.H., Garvey, C., Jenses N., Wilson, J. and Wu, A. "All secure DBMS design", Proc. 11th National Computer Security Conf., Appendix, October 1988.

[15] Hoffman, L.J. and Miller, W.F. "Getting A Personal Dossier From A Statisitical Data Bank", Datmation, Vol. 16(1), May. 1970 pp. 74.

[16] Jajodia, S. and Kogan, B. "Intergrating an object-oriented data model with multilevel security", Proc. 1990 IEEE Symp. on Security and Privacy, May 1990.

[17] Kam, J.B. and Ullman, J.D. "A Model of Statistical D. B. and Their Security", ACM Trans. on D.B. Systems, Vol.2, No.1 1977, pp. 1-10.

[18] Lunt, T.F."Aggregation and inference : Facts and fallacies," Proc. 1989 IEEE Symp. on Research in Security and Privacy, May 1989.

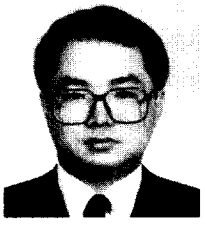
[19] Lunt, T.F. and Hsieh, D. "The Scan View secure database system : A progress report", Proc. 1990 European Symposium on Research in Computer Security, October 1990.

[20] Omar, K.A. and Wells, D.L. "Modified Architecture for subkeys Model", IEEE security and privacy 1983 pp. 79-86.

[21] Schlorer, J. "Disclosure From Statistical D.B. : Quantitative Aspects for Trackers", ACM Trans. on Database Systems, Vol.5(4), Dec. 1980, pp. 467-492.

[22] Schwartz, M.D., Denning, D.E. and Denning, P.J. "Secure Data Bases Under Linear queries", Proc., AFIPS, 1977, pp. 395-398.

□ 著者紹介



김 철(金 鐵) 정회원

연세대학교 이과대학 수학과 졸업(이학사)
미국 North Carolina 주립대 대학원 수학과 졸업(이학석사·박사)
미국 North Carolina 주립대 수학과 시간강사
미국 Shaw University 전임강사
미국 University of South Dakota 수학과 조교수
현재 광운대학교 이과대학 수학과 조교수

※ 연구 관심분야 : 추상 대수학의 응용, 암호학의 수학적 이론, Chaos의 암호학에의 응용등임.