情報保護學會論文誌
第20卷 第3號, 2010. 6

다자간 환경에서 프라이버시를 보호하는 효율적인 DBSCAN 군집화 기법*

김 기 성,† 정 익 래‡
고려대학교 정보경영공학전문대학원

Practical Privacy-Preserving DBSCAN Clustering Over Horizontally Partitioned Data*

Gi Sung Kim,† Ik Rae Jeong‡
Graduate School of Information Management and Security, Korea University

요 약

본 논문은 다자간 환경에서 프라이버시를 보호하는 효율적인 DBSCAN 군집화 기법을 제안한다. 기존 DBSCAN 군집화 기법에 가짜 데이터 삽입을 통한 프라이버시 보호 기법을 적용해 다자간 환경에서 프라이버시를 보호하는 기법으로 확장했다. 기존의 프라이버시를 보호하는 다자간 환경의 군집화 기법들은 비효율적이어서 실제 환경에 적용하기 힘들지만 제안한 기법은 이러한 문제를 해결한 매우 효율적인 기법이다. 본 기법은 다자간 환경뿐만 아니라 비 다자간 환경에도 적용 가능한 효율적인 기법이다.

ABSTRACT

We propose a practical privacy-preserving clustering protocol over horizontally partitioned data. We extend the DBSCAN clustering algorithm into a distributed protocol in which data providers mix real data with fake data to provide privacy. Our privacy-preserving clustering protocol is very efficient whereas the previous privacy-preserving protocols in the distributed environments are not practical to be used in real applications. The efficiency of our privacy-preserving clustering protocol over horizontally partitioned data is comparable with those of privacy-preserving clustering protocols in the non-distributed environments.

Keywords: Privacy, DBSCAN, Clustering

I. 서 론

최근 IT 기술의 급속한 발전으로 인해 엄청난 양의 데이터 축적이 사회 각 전반에 나타나고 있다. 예를 들어, 백화점에서는 하루에도 수만 건의 거래 내역이 자료화 되어 저장되고, 병원에서는 수많은 환자의 진료 기록이 저장된다. 대용량 자료에서 의미있는 정보를 찾아내는 데이터마이닝(data mining)[1]은 이러한 환경에서 그 역할이 점점 더 증대 되고 있다.

데이터마이닝 기법 중 군집화[2]는 서로 유사한 속성을 가지는 데이터들의 군집을 형성하는 알고리즘으로, 군집간의 유사성은 최소화하고, 군집내의 유사성은 최대화 시키는 기법이다. 이러한 군집화 기법을 통해 백화점에서는 구매내역의 유사성을 토대로 고객을 분류하여 관리할 수 있고, 병원에서는 특정 질환에 따라

접수일(2009년 12월 28일), 수정일(2010년 5월 12일), 게재확정일(2010년 6월 11일)
* 본 연구는 지식경제부 및 한국산업기술평가관리원의 산업원천기술개발사업(정보통신)[KI002113, Car-헬스케어 보안 기술개발]과 2008년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음. (KRF-2008-331-D00581)
† 주저자, gisung2137@hanmail.net
‡ 교신저자, irjeong@korea.ac.kr

환자들이 갖는 유사성을 찾아 낼 수도 있다.

하지만 데이터마이닝은 항상 프라이버시 문제를 수반한다. 예를 들어, 고객의 구매 기록을 바탕으로 데이터 마이닝을 실시한다고 하자. 이러한 경우 수많은 고객의 개인 기록이 그대로 유출될 것이다. 이러한 이유로 학계에서는 이미 오래전부터 프라이버시를 보호하는 데이터마이닝을 활발히 연구해오고 있다[3-6].

프라이버시를 보호하는 데이터 마이닝은 환경에 따라 크게 비 다자간 환경과 다자간 환경으로 나뉘어 연구되어 왔다.

비 다자간 환경은 데이터베이스의 소유주가 한명인 환경으로 소유주와 데이터마이닝을 실시하는 마이너(miner)가 다른 환경이다[7-10]. 예를 들어, 국가기관에서 특정 질환의 특징을 파악하기 위한 연구의 목적으로 병원에 환자의 진료 기록을 요청할 수 있다.이러한 환경에서 자료를 원본 그대로 교환하게 되면, 자료의 프라이버시가 그대로 노출될 수 있기 때문에 문제가 된다. 따라서 이러한 환경에서는 데이터를 원본과 다르게 변형시켜 마이너에게 주는 방법을 사용하여 문제를 해결한다. 주로 데이터 소유주가 자신의 데이터에 노이즈 값을 추가하여 원본 데이터의 형태를 변화시켜 프라이버시를 보존하는 기법을 중심으로 연구가 진행되어 왔다. 하지만 이 기법은 연산량이 많고, 검증되지 않은 안전성, 결과의 부적확성 등이 문제로 지적되고 있다.

다자간 환경은 다수의 데이터베이스 소유주들이 각자의 데이터에 대한 프라이버시를 보호하면서 전체 데이터베이스에서 만족하는 군집화 결과를 얻는 환경이다[11-13]. 이러한 기법의 대부분이 이미 안전성이 확보된 안전한 다자간 계산(secure multi-party computation)이나 암호학적인 기법 적용하기 때문에 대체로 만족할 만한 안전성을 가진다. 하지만 사용되는 기법들이 많은 연산량을 요구하고 있어 매우 비효율적이다. 또한, 다자간 환경에 기반한 기법들이기 때문에 비 다자간 환경에 적용할 수 없다.

이에 본 논문에서는 프라이버시를 보호하는 군집화 분야에 한 번도 시도된 적이 없는 가짜 데이터 삽입 기법을 통해 다자간 환경에서 프라이버시를 보호하는 효율적인 군집화 기법을 제안한다. 본 기법은 기존의 다자간 환경 기법들이 사용하는 암호학적 기법이나 안전한 다자간 계산을 사용하지 않아 매우 효율적이고, 비 다자간 환경에도 적용 가능한 매우 현실적인 기법이다.

II. 관련 연구

비 다자간 환경에서의 기법 연구는 다음과 같다.

[7]은 비 다자간 기법 중 처음으로 제안된 논문으로 기하학적 변환 방법인 이동(shift), 확대(scale), 변환(rotation) 등을 통해 원본 데이터를 변형함으로써 데이터의 프라이버시를 보호하는 기법을 제안했다. 하지만 제안된 기법은 알려진 원본 데이터와 변형된 데이터 쌍에 의해 쉽게 공격 가능하다.

[9]에서는 [7]에서 사용한 기법 중 하나인 변환을 특화시켜 새로운 변환 기법인 RBT(Rotation Based Transformation) 기법을 제안했다. RBT 기법은 원본 데이터의 차원을 임의로 2개씩 선택하여 2차원 점으로 생각하여 회전변환 하는 방식이다. 또한, 자료의 민감도에 따라 사용자가 원하는 정도의 프라이버시 조절이 가능한 기법이다. [7]의 안전성을 보완했으나 필요한 연산량이 너무 많다.

[10]에서는 데이터베이스의 주성분 분석을 통해 원본 데이터의 차원을 낮추어 프라이버시를 보존하는 기법을 제안했다. 차원이 낮은 데이터에서 높은 차원의 데이터를 복구하는 것은 힘든 문제라는 것을 기반으로 이전과는 다른 연구 방향을 제시하게 된다. 또한, 차원을 낮추는데 있어 원본 데이터의 성질을 최대한 보존하기 위해 주성분 분석을 사용함으로써 실질적인 거리보존 기법이 되어 높은 정확성을 보장하는 기법이다. 하지만 주성분 분석에 필요한 연산량이 너무 과도해 매우 비효율적이다.

위의 비 다자간 환경과는 다르게 [11-13]에서는 다자간 환경에서 프라이버시를 보호하는 데이터 마이닝 기법을 연구했다. [11]에서는 수직 분할된 데이터베이스를 가정하고, [12]에서는 본 논문과 같은 수평 분할 환경을 가정했다. [11-12]모두 안전한 다자간 계산 기법을 설계하고, 이를 군집화 기법에 적용하여 프라이버시 보호 기법으로 확장했다. 하지만 안전한 다자간 계산에 필요한 연산량이 많아 비효율적이다. [13]에서는 암호학적 기법 중 하나인 Homomorphic 암호화 기법을 적용했다. 안전성이 증명된 암호화 기법을 사용하기 때문에 높은 프라이버시를 보장하지만, 모든 데이터를 각각 암호화하기 때문에 이 기법 역시 비효율성이 문제가 된다. 이처럼 다자간 환경의 기법들은 엄밀한 안전성을 보장하는 암호학적 기법이나 안전한 다자간 계산을 사용하기 때문에 매우 비효율적이며, 비 다자간 환경에는 적용할 수 없는 단점이 있다.

III. DBSCAN 군집화

DBSCAN[14] 알고리즘은 군집화 알고리즘 중에서 대표적인 알고리즘으로 데이터들 사이의 밀도 차이에 기반을 두고 수행되는 알고리즘이다. DBSCAN 알고리즘의 특징은 군집 개수가 초기에 주어지지 않아도 효율적으로 군집화를 수행하고, 같은 데이터 집합에 대해 여러 번 수행 하더라도 항상 같은 결과 값을 내놓는 정확하면서 효율적인 알고리즘이다.

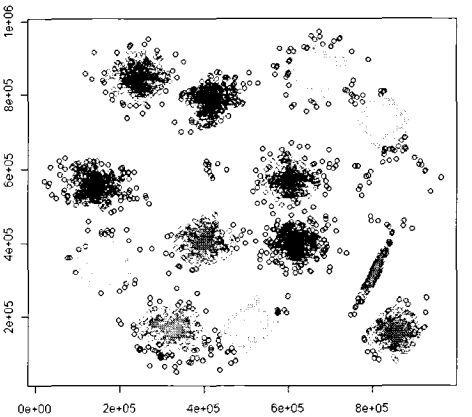
D는 사용자의 데이터베이스를 의미하며, 점 $p,q \in D$는 d차원 벡터이다. $dist(p,q)$는 점 p와 q사이의 거리를 나타내며, 보통 유클리디안 거리를 사용한다. $N_{Eps}(p) = [q \in D | dist(p,q) \leq Eps]$은 점 p를 중심으로 반경 Eps안에 있는 점들의 집합을 의미한다.

DBSCAN 군집화는 D의 모든 점을 3종류로 분류한다. 점p가 $|N_{Eps}(p)| \geq MinPt$을 만족할 때 즉, 점 p의 Eps반경 안에 $MinPt$개 이상의 다른 점이 존재할 때, 점 p를 핵심점으로 분류한다. $MinPt$는 사용자가 임의로 지정하는 한계 값을 나타낸다. 자기 자신은 핵심점이 아니지만 자신의 Eps반경 안에 다른 핵심점이 존재할 때, 이 점을 변두리점으로 분류한다. 마지막으로 핵심점과 변두리점을 제외한 나머지 점을 노이즈점으로 분류한다. Eps와 $MinPt$은 알고리즘 수행에 핵심적인 역할을 하며, 이를 지정하는 방법은 [14]에 소개되어 있다. 기본적인 DBSCAN 알고리즘의 원리는 하나의 핵심점이 정해지면 그 핵심점을 중심으로 Eps 안에 다른 핵심점이 존재 할 때, 이를 같은 군집으로 지정한다. 이러한 군집은 핵심점끼리 연결되면서 그 크기를 확장해가며, 군집의 경계는 변두리점들로 정해진다. 노이즈점들은 알고리즘에 영향을 주지 못하기 때문에 DBSCAN 알고리즘은 노이즈에 강한 알고리즘이다.

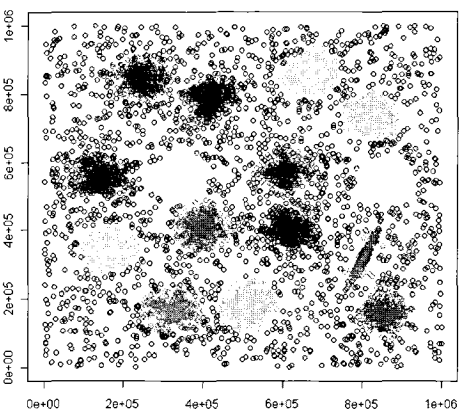
IV. 제안하는 기법

DBSCAN은 기본적으로 데이터간의 밀도 차이를 기반으로 실행되는 알고리즘이다. 밀도 차이에 영향을 주지 않고 데이터를 변형할 수 있다면 같은 군집화 결과를 얻을 수 있다. 따라서 가짜 데이터를 균등하게 삽입한다면 전체적으로 밀도 상승은 일어나지만 밀도 차이는 보존할 수 있다.

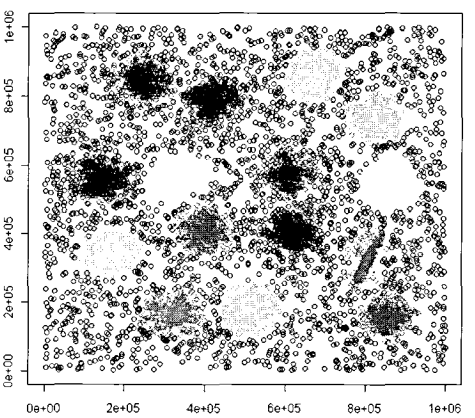
[그림 1]에서 [그림 3]은 원본 데이터베이스에 가짜 데이터를 삽입한 후, DBSCAN 군집화 실험 결과를 보여주고 있다. 원본 데이터에서 생성된 군집이 변형된 데이터에서 그대로 생성되고 있음을 알 수 있다. 또한, 전반적으로 데이터의 밀도 상승이 일어났지만,



[그림 1] 원본 데이터에 대한 DBSCAN 군집화 결과



[그림 2] 원본 데이터의 50% 만큼 가짜 데이터가 삽입된 변형된 데이터에 대한 DBSCAN 군집화 결과



[그림 3] 원본 데이터의 100% 만큼 가짜 데이터가 삽입된 변형된 데이터에 대한 DBSCAN 군집화 결과

원본 데이터의 밀집도가 높은 곳은 변형된 데이터에서 여전히 높은 밀집도를 나타내고 있다.

각 그림의 군집 결과의 정확성을 판단하기 위해서 군집화의 정확성에 대한 정의가 필요하다.

정의 1. 데이터베이스 D와 D'의 군집화 결과가 다음과 같을 때 D와 D'에서 군집화 결과가 같다고 말한다.

1. 데이터 p와 q가 D에서 군집화 후 같은 군집에 속하는 점일 때, 데이터 p와 q는 D'에서도 군집화 후 같은 군집에 속해야 한다.
2. 데이터 p와 q가 D에서 군집화 후 서로 다른 군집에 속하는 점일 때, 데이터 p와 q는 D'에서도 군집화 후 서로 다른 군집에 속해야 한다.
3. D와 D'에서 군집화 후 생성된 군집의 개수는 같아야 한다.

위의 정의를 바탕으로 위 그림들의 군집화 결과를 분석해 보면 모두 같은 군집화 결과를 나타내고 있음을 알 수 있다.

본 논문에서 제안하는 다자간 환경에서 프라이버시를 보호하는 효율적인 DBSCAN 군집화의 자세한 기법은 다음과 같다. 데이터 소유주 P_i는 각각 자신의 데이터베이스 D_i를 가지고 있다. $p=(p_1,...,p_d)$는 데이터베이스에 속하는 d차원의 데이터를 나타내며, $u_i \le p_i \le v_i$의 범위를 갖는다고 가정한다.

데이터 소유주$(P_1,...,P_n)$가 $(D_1 \cup ... \cup D_n)$에서 만족하는 DBSCAN 군집 결과를 얻기 위해서 다음과 같은 프로토콜을 수행한다.

1. P_i는 가짜 데이터 개수 m_i를 설정한다.
2. P_i는 가짜 데이터 셋 $F_i=\{f_1,...,f_{m_i}\}$을 생성한다. $f_j(1 \le j \le m_i)$는 $f_j=(f_{j,1},...,f_{j,d})$을 나타내며, 각각의 $f_{j,k}(1 \le k \le d)$는 $[u_k,v_k]$에서 랜덤(random)하게 선택한다.
3. P_i는 변형된 데이터베이스 $D_i'=D_i \cup F_i$를 생성하여 다른 모든 데이터베이스 소유주에게 보낸다.
4. P_i는 모든 $D_i'(1 \le i \le n)$를 수집하여 $(D_1' \cup ... \cup D_n')$을 생성하여 DBSCAN 군집화를 실행한다.

V. 프라이버시 분석

제안한 기법은 가짜 데이터를 삽입함으로써 공격자가 어느 것이 진짜 데이터인지를 판단하기 힘든 점에 안전성으로 두고 있다. 하지만 공격자가 얻을 수 있는 정보는 존재한다. D'에는 균등하게 가짜 데이터가 삽입되어 있으므로 공격자는 생성된 군집속에 보다 많은 진짜 데이터들이 존재함을 예측할 수 있다. 따라서 공격자가 특정한 군집을 보고 가짜 데이터를 얼마만큼 확률로 판단할 수 있을지를 해당 군집의 프라이버시로 정의하는 게 타당하다. 또한, 이러한 군집 프라이버시 중 가장 작은 값을 전체 데이터베이스의 프라이버시로 정의한다.

변형된 데이터베이스 D'에 m개의 가짜 데이터가 삽입되었고, 군집화 결과로 c개의 군집 $(C_1,...,C_c)$이 생성되었다고 하자. $|D'|$와 $|C_i|$는 각각 D'의 데이터 개수와 군집C_i의 데이터 개수를 나타낸다. 프라이버시 분석을 위해 특정 군집이 전체 데이터베이스에서 차지하는 비중을 알아야 하기 때문에 부피의 정의가 필요하다. $V(D')$는 데이터베이스 D'의 부피를 의미하며, $\prod_{k=1}^{d}|v_k-u_k|$으로 계산된다. 즉, D'에 속하는 점들에 대해 각각의 차원들이 속하는 범위의 곱으로 정의한다. $V(C_i)$는 군집 C_i의 부피를 의미하며, C_i를 하나의 데이터베이스로 보고 $V(D')$와 같은 방법으로 계산한다. 이를 바탕으로 특정 군집 C_i의 프라이버시 $Privacy(C_i)$와 전체 프라이버시 $Privacy(D')$는 다음과 같이 정의한다.

$$Privacy(C_i)=(V(C_i)/V(D)\times m)/|C_i| \quad (1)$$

$$Privacy(D')=Min(Privacy(C_i)) \quad (1 \le i \le c) \quad (2)$$

VI. 실험 및 비교

본 장에서는 여러 데이터베이스를 바탕으로 실제 가짜 데이터 삽입 알고리즘을 실행해보고 그에 따른 분석을 실시한다. 또한, 지금까지 연구된 다른 기법들과의 비교를 해봄으로써 본 기법의 우수성을 말한다.

실시하는 실험은 군집화 알고리즘은 R언어로 시뮬레이션하며, 가짜 데이터 삽입은 C로 실시했다. 사용한 데이터베이스는 http://cs.joensuu.fi/sipu/datasets/을 이용하였다. 실험 컴퓨터는 Intel Pentium4 3.2GHz, RAM 2G이고 운영체제로는 Window XP를 사용하였다.

[표 1] 실험 결과를 분석해 보면, 대부분의 데이터베이스에서 해당 데이터 개수의 100% 이하로 가짜

표 1. 시뮬레이션 결과

데이터베이스	House	A1	A2	S1	S3
데이터 개수	1857	3000	5250	5000	5000
차원	3	2	2	2	2
군집 개수	1	20	31	15	15
30%추가	전체 데이터 개수의 30%만큼 가짜 데이터 삽입시				
정확도	O	O	O	O	O
프라이버시	0.08	0.02	0.07	0.09	0.05
50%추가	전체 데이터 개수의 50%만큼 가짜 데이터 삽입시				
정확도	O	O	O	O	O
프라이버시	0.12	0.05	0.13	0.18	0.12
100%추가	전체 데이터 개수의 100%만큼 가짜 데이터 삽입시				
정확도	O	O	X	O	X
프라이버시	0.23	0.13	-	0.34	-
200%추가	전체 데이터 개수의 200%만큼 가짜 데이터 삽입시				
정확도	O	X	-	X	-
프라이버시	0.31	-	-	-	-

데이터를 삽입했을 경우 정상적인 결과를 나타내었다. 또한, 프라이버시 역시 가짜 데이터를 삽입하는 개수에 따라 증가함을 알 수 있고, 실험에서는 최대 0.3 근방까지 프라이버시를 가질 수 있었다. A1, S1은 군집의 경계가 또렷한 데이터베이스이고, A2와 S3은 그보다 약간 불분명한 데이터베이스였다. 결과에서 알 수 있듯이 군집화가 뚜렷하게 가능한 데이터베이스의 경우에 보다 많은 가짜 데이터삽입이 가능했고 그에 따른 높은 프라이버시를 보장 할 수 있었다. 따라서 본 기법은 군집의 경계가 뚜렷한 환경에서 보다 높은 성능을 발휘함을 알 수 있다.

본 기법은 DBSCAN의 입력 변수인 *Eps*와 *MinPt* 값을 조절함으로써 프라이버시 값을 향상 시킬 수 있다. *Eps*값은 작게, *MinPt*값은 크게 입력할 경우 보다 높은 프라이버시를 얻어낼 수 있다. 아래의 [표 2]는 A1에 대해 *MinPt*값의 변화에 따른 최대 프라이버시를 나타낸 표이다.

표 2. A1의 시뮬레이션 결과

Min	50	55	60	65
프라이버시	0.13	0.19	0.23	0.31

정확도를 유지하면서 *Eps*값을 낮추고, *MinPt*값을 높게 설정할 경우 보다 높은 밀도를 가진 지역만이 군집으로 처리되기 때문에 보다 많은 가짜 데이터를 삽입해도 정확도를 유지할 수 있게 된다. 하지만 지나치게 *MinPt*값과 *Eps*값을 조절할 경우 정확도에 문제가 생기게 된다. 하지만 정확도를 유지하는 정도의 *MinPt*값과 *Eps*값의 조절을 통해 원하는 프라이버시를 일정 수준으로 보장할 수 있다.

따라서 본 기법은 사용자의 요구에 따라 프라이버시가 조절 가능한 효율적인 기법으로 요구에 따라 일정 수준의 프라이버시까지 보장할 수 있다.

VII. 결 론

본 논문에서는 다자간 환경에서 프라이버시를 보호하는 효율적인 DBSCAN 군집화 기법을 제안했다. 기존의 프라이버시 보존 군집화 기법에 한번 도 시도되지 않은 가짜 데이터 삽입 기법을 통해 밀도 차이에 기반으로 실행되는 DBSCAN 군집화 기법을 다자간 환경에서 프라이버시를 보호하는 효율적인 기법으로 확장했다. 본 기법은 기존의 다자간 환경에서 제안된 기법들이 주로 사용하는 안전한 다자간 계산과 암호학적 기법을 하지 않아 매우 효율적이며, 비 다자간 환경에서도 활용 가능한 실용적인 기법이다.

참고문헌

[1] J. Han and M. Kamber, Data Mining:

Concepts and Techniuqes, 2th Ed., Morgan Kaufmann Publishers, Jan. 2006.

[2] A.K. Jain and R.C. Dubes, Algorithms for Clustering Data, Prentice-Hall, Mar. 1998.

[3] R. Agrawal and R. Srikant, "Privacy-preserving data mining," In Proceedings of the 2000 ACM SIGMOD Conference on Management of Data, pp. 439 - 450, May 2000.

[4] Y. Lindell and B. Pinkas, "Privacy preserving data mining," In Advances in Cryptology - CRYPTO 2000, pp. 36-54, Aug. 2000.

[5] D. Agrawal and C.C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms," In Proceedings of the 20th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 247-255, Mar. 2001.

[6] A. Evfimevski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," In Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 211-222, June 2003.

[7] S. Oliveira and O. Zaiane, "Privacy Preserving Clustering By Data Transformation," In Proceedings of the 18th Brazilian Symposium on Databases, pp. 304-318, Oct. 2003.

[8] Z. Huang, W. Du, and B Chen, "Deriving Private Information from Randomized Data," In ACM SIGMOD, pp. 37-48 , June 2005.

[9] S.R.M. Oliveira and O.R. Za¨iane, "Achieving Privacy Preservation When Sharing Data For Clustering," In Proc. of the Workshop on Secure Data Management in a Connected World (SDM'4) in conjunction with VLDB'004, pp. 67-82, Aug. 2004.

[10] K. Liu and H. Kargupta, "Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining," IEEE TKDE, vol. 18, no. 1, pp. 92-106 , Nov. 2006.

[11] J. Vaidya and C. Clifton, "Privacy-Preserving K-Means Clustering Over Vertically Partitioned Data," In Proc. of the 9th ACM SIGKDD Intl. Conf. on Knowlegde Discovery and Data Mining, pp. 206-215, Aug. 2003.

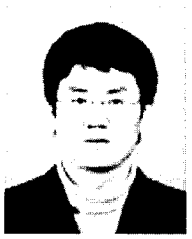
[12] A. Inan, S.V. Kaya, Y. Saygin, E. Savas, A.A. Hintoglu, and A. Levi, "Privacy preserving clustering on horizontally partitioned data," Data & Knowledge Engineering (DKE), vol. 63, no. 3, pp. 646-666, Oct. 2007.

[13] S. Jha, L. Kruger, and P. McDaniel, "Privacy preserving clustering," In Proceedings of 10th European Symposium on Research in Computer Security (ESORICS '5), Milan, pp. 397-417 , Sep. 2005.

[14] M. Ester, H.P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, pp. 226 - 231, Aug. 1996.

〈著者紹介〉

김 기 성 (Gi Sung Kim) 학생회원
2008년 2월: 서울시립대학교 수학과 졸업
2008년 3월~현재: 고려대학교 정보경영공학전문대학원 석사과정
〈관심분야〉 프라이버시향상기술(PET), 데이터베이스 보안, 암호 이론



정 익 래 (Ik Rae Jeong) 정회원
1998년 2월: 고려대학교 전산학과 학사 졸업
2000년 2월: 고려대학교 전산학과 석사 졸업
2004년 8월: 고려대학교 정보보호대학원 박사 졸업
2006년 6월~2008년 2월: 한국전자통신연구원 암호기술연구팀 선임연구원
2008년 3월~현재: 고려대학교 정보경영공학전문대학원 조교수
〈관심분야〉 프라이버시향상기술(PET), 데이터베이스 보안, 암호 이론