순서 유지 암호화 기반의 k-최근접 질의처리 알고리즘 설계

김용기*, 최기석*
*한국과학기술정보연구원 R&D시스템개발실
e-mail:ykkim@kisti.re.kr

Design of k-Nearest Neighbor Query Processing Algorithm Based on Order-Preserving Encryption

Yong-Ki Kim*, KiSeok Choi*
*Dept. of R&D System Development, Korea Institute of Science and Technology Information

요 약

최근 모바일 사용자의 안전한 위치기반 서비스의 사용을 위한 아웃소싱 데이터베이스에서 객체 및 사용자의 위치 정보를 보호하는 연구가 위치 데이터를 보호하기 위한 연구가 활발히 진행되고 있다. 그러나 기존 연구는 불필요한 객체 정보를 요구하기 때문에, 높은 질의 처리 시간을 지니는 단점을 지닌다. 이러한 문제점을 해결하기 위해, 본 논문에서는 기준 POI를 중심으로 객체의 방향성 정보와 변환된 거리를 이용하여, 사용자와 객체의 정보를 보호하는 k-최근접 질의처리 알고리즘을 제안한다.

1. 서론

최근 모바일 사용자의 안전한 위치기반 서비스의 사용을 위한 아웃소싱(Outsourcing) 데이터베이스에서 객체 및 사용자의 위치 정보를 보호하는 연구가 위치 데이터를 보호하기 위한 연구가 활발히 진행되고 있다[1-3]. 이를 위한 기존 아웃소싱 데이터베이스에서 사용자 위치 정보 보호를 위한 연구는 공간 좌표 변환에 관한 연구와 암호화 변환에 관한 연구로 나뉘어 있다. 첫째, 기존 공간 좌표 변환에 관한 연구는 위치 데이터의 변환을 통해 데이터를 보호한다. 둘째, 기존 암호화 변환에 관한 연구는 데이터를 암호화하여 데이터베이스에 저장하는 구조를 지닌다. 그러나, 기존 공간 좌표를 변환하는 연구 및 암호화 변환에 관한 연구의 대부분은 키가 노출되는 경우 객체에 대한 보호가 이루어지지 않는다. 이러한 문제점을 해결하기 위해, 객체 및 사용자의 위치 정보 대신 거리를 저장하여 데이터 보호를 수행하는 연구가 제안되었다[3]. 이 기법은 기준 POI(Point Of Interest)와 이웃하는 POI 간의 거리를 순서는 유지하나, 거리를 변환하여 객체 좌표 대신 변환된 거리만을 저장함으로써, 데이터의 위치 정보를 보호한다. 변환된 객체 정보는 서비스 제공자(Service Provider)에게 제공함으로써, 서비스 제공자로부터 안전하게 데이터를 지키는 장점을 지닌다. 그러나, 좌표정보 대신 기준 POI(anchor)로부터 거리정보만을 지니고 있으므로, 질의지점, 앵커 및 객체간의 방향성 정보를 알 수 없다. 따라서 질의 수행 시 불필요한 POI 요구가 발생하며, 이는 빠른 서비스 응답을 제공하지 못한다.

이러한 문제점을 해결하기 위해, 본 논문에서는 객체가 존재하는 영역을 앵커를 기준으로 N*M개의 영역으로 나누어 POI의 거리 정보와 방향성 정보를 유지하는 거리 순서 유지 (Distance order preserving) 암호화 기반의 k-최근접 질의처리 알고리즘을 제안한다. 제안하는 k-최근접 질의처리 알고리즘은 방향성을 고려하여 일정 영역으로 나누어 저장하기 때문에, 불필요한 객체 정보의 송수신을 줄이고, 빠른 질의처리 시간을 제공한다.

본 논문은 다음과 같이 구성된다. 2장에서는 사용자의 정보와 객체 정보를 보호하는 관련연구를 소개한다. 3장에서는 k-최근접 질의처리 알고리즘을 제안한다. 마지막으로, 5장에서는 결론과 향후 연구를 제시한다.

2. 관련연구

아웃소싱 데이터베이스에서 공격자로부터 객체의 위치 데이터를 보호하기 위한 기존 데이터의 암호화 변환에 관한 연구는 사용자의 위치 정보가 신뢰할 수 없는 서버에 저장될 경우, 악의적인 사용자에 의해 이러한 정보가 노출되는 것을 방지하기 위해 그리드 기반의 사용자 위치 암호화 알고리즘[2], R-tree를 이용하여 암호화하는 CRyptographic Transformation (CRT) 알고리즘[1], Metric 거리 순서를 유지하는 암호화를 수행하는 Metric Preserving Transformation (MPT) 알고리즘[3]이 존재한다. 첫째, A. Khoshgozaran과 C. Shahabi의 연구[2]는 객체 단위로 암호화를 수행하기 때문에, 영역 질의 결과 분석을 통해 데이터의 위치를 쉽게 유추할 수 있다. 둘째, CRT는 트리의 노드 레벨에 따라 순차적으로 접근하기 때문에, 튜플 연관성 파악이 용이하고 위치 데이터 노출에 매우 취약하다. 또한, A. Khoshgozaran과 C. Shahabi의 연구와 CRT는 암호화에 사용된 키가 노출되면 위치 데이터가 유추되는 단점을 지니고 있다. 셋째, M. L. Yiu et al. 은 객체의 좌표를 보호하기 위하여, 앵커(anchor)가 지

닌 일정 거리 이내의 객체에 대하여 거리 순서로 정렬하고 거리를 목표 분포로 변환하여 저장한다. 그러나, MPT 알고리즘은 앵커가 지니고 있는 거리가 크기 때문에, 후보 집합이 많은 문제점을 지니고 있다.

3. 거리순서 유지 암호화 기반의 k-최근접 질의처리 알고리즘

기존 MPT 알고리즘은 변환된 데이터의 방향성을 고려하지 못하기 때문에, 질의지점, 객체 및 앵커간의 거리를 이용하여 최적의 가지치기를 수행할 수 없다. 따라서, 질의 처리에 필요한 객체 수가 증가하여, 연산과 I/O 수가 증가한다. 이러한 문제점을 해결하기 위해, 데이터의 방향성을 고려한 영역 분할을 이용한 거리순서 유지 암호화 k-최근접 질의처리 알고리즘을 제안한다. 이를 위해, 첫째, 앵커가 커버하는 범위를 N개의 섹터로 나눈다. 사용자가 앵커의 위치와 자신의 위치를 이용하여 자신이 위치한 방향을 계산한다. 이는 앵커로부터 사용자의 반대편에 위치한 데이터를 고려하지 않음으로써, 질의처리에 수행에 불필요한 객체의 반환을 없앨 수 있다. 둘째, 나눠진 섹터 범위 내에서 거리를 기반으로 M개의 영역으로 나눈다. 이 때, M은 프라이버시의 레벨을 의미하며, M의 크기에 따라 반환되는 객체의 수를 줄일 수 있다. 정의 1은 객체 p의 표현을 나타낸다. OPE[4]를 이용하여 변환된 앵커와 객체간의 거리는 서버에 저장되고, 사용자가 요청할 때 해당 영역에 위치한 객체를 반환한다. 이 때, 객체 p, p'에 대해 OPEa(p)가 OPEa(p')보다 크면, 앵커 a로부터 p의 실제거리는 p'의 실제거리보다 크다.

정의 1. 객체 p(x,y)가 앵커 a(x', y')로부터 방향성 ε를 지니고, 거리 r로부터 변환된 OPE(r)을 지닌다고 할 때, p = (a, OPEa(r),ε)라 표현한다.

거리 기반 암호화 알고리즘은 다음과 같이 수행된다. 첫째, 객체 집합 P로부터 앵커 집합A를 선택하고, 각 앵커가 가질 버켓의 수 N, M 을 결정한다. 둘째, N과 M을 이용하여 길이 B와 앵글 C를 결정하고, 각 버켓의 영역에 포함된 객체를 해당 버켓Bi, j에 할당한다. 각 버켓 내의 모든 POI에 대해, POI p가 위치한 앵글 ε를 계산하고 서버에 OPE로 암호화된 거리 정보와 함께 보낸다.

한편, 기존 MPT 알고리즘의 경우, 최근접 anchor를 통해 생성된 탐색 영역만으로 POI를 반환하기 때문에 불필요한 후보 셋을 반환한다는 문제점을 지닌다. 제안하는 암호화된 공간 데이터를 위한 k-최근접점 질의처리 알고리즘은 사용자 위치와 앵커 간의 각도를 계산한 후, 해당 영역에서 샘플 데이터 k개를 송수신하여 질의지점으로부터 k번째 거리인 maxdist 를 설정한다. 설정된 maxdist 는 질의지점과 앵커 간의 각도를 고려하기 때문에, 최적의 한계치를 설정할 수 있으며, 정의 3을 이용하여 가지치기를 수행함으로써 질의 결과 후보 셋의 크기를 줄인다.

정의 2. 앵커 a가 거리 R을 N개의 원과 M개의 섹터로 나눌 때, n = sqrt(pow(x-x')+pow(y-y')) % (R/N), m = ε%(2 π/M) 번째 위치하는 공간에 위치한다.

정의 3. 샘플링 데이터 중에서 질의점 q로부터 떨어져 있는 객체를 q_k가 존재할 때, 다음과 같은 조건을 만족하는 p_i가 존재하는 경우, p_i는 p_k에 의해 가지치기된다.

$$Maxdist(p_k, q) \leq Mindist(p_i, q)$$

제안하는 k-NN 질의처리 알고리즘은 그림 1과 같다. 첫째, 질의 지점 q와 앵커와의 각도 ε를 계산한다. 전체 객체에서 지역을 대표하는 앵커 집합 A를 만든다. 둘째, 각도 ε를 가진 버켓에 존재하는 k개의 샘플링을 서버에게 요청한다. 셋째, 전송받은 k개의 POI에서 maxdist 설정하고, maxdist를 이용하여 후보 버켓 설정하고, 해당 버켓에 존재하는 질의 후보 집합을 요청한다. 마지막으로, 전송받은 질의 후보 집합에서 실제 거리를 계산하여, 최종 질의 결과를 반환한다.

1. 질의점과 앵커와의 각도 ε 계산
2. ε 버켓내에 존재하는 k개 샘플 데이터 요청
3. 샘플 데이터로부터 k번째 거리 maxdist 계산
4. maxdist로부터 후보 버켓 탐색 후 POI 셋 요청
5. POI 셋으로부터 각각의 POI에 대해 질의점과의 거리 연산
6. k개의 POI 반환

(그림 1) 거리 기반 k 최근접 질의 처리 알고리즘

4. 결론

본 논문에서는 아웃소싱 데이터베이스에서 공격자로부터 객체의 위치 데이터를 보호하기 위한 거리 순서 유지 암호화 기반의 k-최근접 질의처리 알고리즘을 제안한다. 제안하는 기법은 불필요한 객체 정보의 송수신을 줄이고, 빠른 질의처리 시간을 제공하기 위해, 객체가 존재하는 영역을 앵커를 기준으로 N*M개의 영역으로 나누어 POI 방향성 정보와 거리 순서를 유지한다. 향후 연구로는 제안하는 알고리즘의 구현 및 성능 평가를 통해, 알고리즘의 효율성을 입증하는 것이다.

참고문헌

[1] M. L. Yiu et al. "Enabling Search Services on Outsourced Private Spatial Data", VLDB Journal, 2010.

[2] A. Khoshgozaran and Cyrus Shahabi, "Private Buddy Search: Enabling Private Spatial Queries in Social Networks", Int'l Conference on Computational Science and Engineering, 2009.

[3] M. L. Yiu, I. Assent, C. S. Jensen, and P. Kalnis, "Outsourced Similarity Search on Metric Data Assets", IEEE TKDE journal, Vol. 24, No. 2, pp. 338-352, 2012.

[4] R. Agrawal et al. Order-Preserving Encryption for Numeric Data. In SIGMOD, pp. 563 - 574, 2004.