도로 네트워크 환경에서 암호화된 공간데이터를 위한 K-최근접점 질의 처리 알고리즘

A K-Nearest Neighbour Query Processing Algorithm for Encrypted Spatial Data in Road Network

장 미 영* 장 재 우**
Mi Young Jang Jae Woo Chang

요 약 최근 클라우드 컴퓨팅의 발전에 따라, 데이터베이스 아웃소싱(Outsourcing)에 대한 연구가 활발히 진행되고 있다. 또한 무선 통신 기술 및 모바일 기기의 발전으로 인해 위치 기반 서비스를 이용하는 사용자의 수가 증가하였다. 따라서 개인 또는 소규모의 사업자는 데이터 저장 및 관리 비용을 줄이기 위해 그들의 공간 데이터를 위치 기반 서비스 제공자에게 아웃소싱 한다. 그러나 사용자의 위치 정보는 시간대별 방문 장소 및 개인 정보를 지니고 있기 때문에, 이에 대한 허용되지 않은 접근 시 개인 정보 유출 문제가 발생한다. 따라서 위치 정보 아웃소싱을 위한 개인 정보 보호 연구가 필요하다. 이러한 문제를 해결하기 위해, 본 논문에서는 아웃소싱 환경에서 도로네트워크를 고려한 암호화된 공간 데이터베이스 기반 k-최근접점 질의 처리 알고리즘을 제안하였다. 제안하는 기법은 데이터베이스 아웃소싱을 위해 위치 데이터를 네트워크 거리 정보로 변환 및 암호화한 가공데이터를 생성하여 이를 서비스 제공자에게 전송한다. 또한, 전처리 과정을 통해 네트워크 노드와 POI 거리를 미리 저장하여 네트워크 탐색을 빠르게 수행하며, 질의 수행 시 최근접 대표 POI 및 암호화된 거리 정보를 이용하여 질의 결과 후보 집합을 탐색한다. 마지막으로, 질의 영역 재설정 과정을 통해 불필요한 후보 탐색을 줄임으로써 효율적으로 POI를 탐색한다. 마지막으로, 성능평가를 통해 제안하는 기법이 기존 방법에 비해 우수함을 보인다.

키워드 : 아웃 소싱 공간 데이터베이스, 공간데이터 암호화 기반 k-최근접점 탐색, 개인정보 보호 지원 질의처리 알고리즘

Abstract Due to the recent advancement of cloud computing, the research on database outsourcing has been actively done. Moreover, the number of users who utilize Location-based Services(LBS) has been increasing with the development in wireless communication technology and mobile devices. Therefore, LBS providers attempt to outsource their spatial database to service provider, in order to reduce costs for data storage and management. However, because unauthorized access to sensitive data is possible in spatial database outsourcing, it is necessary to study on the preservation of a user's privacy. Thus, we, in this paper, propose a spatial data encryption scheme to produce outsourced database from an original database. We also propose a k-Nearest Neighbor(k-NN) query processing algorithm that efficiently performs k-NN by using the outsourced database. Finally, we show from performance analysis that our algorithm outperforms the existing one.

Keywords : Outsourced Spatial Database, Encrypted Spatial Database Based k-NN Query Processing, Privacy-preserving Query Processing

† 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임. (과제번호 2010-0023800)

* 전북대학교 컴퓨터 공학과 박사과정 brilliant@jbnu.ac.kr
** 전북대학교 IT정보공학과 교수 jwchang@jbnu.ac.kr(교신저자)

1. 서 론

최근 클라우드 컴퓨팅에 대한 관심이 고조됨에 따라, 이를 활용한 데이터베이스 아웃소싱(Outsourcing)에 대한 연구가 활발히 진행되고 있다[3, 4, 13]. 데이터베이스 아웃소싱이란 데이터 소유자와 서비스 제공자를 분리하여, 데이터 소유자는 데이터베이스를 구축하고, 서비스 제공자는 데이터소유자로부터 전송받은 데이터베이스를 관리하는 시스템 구조를 말한다. 또한, 인증된 사용자는 서비스 제공자를 통해 데이터베이스에 접근하여 정보를 검색한다. 한편, 최근 GPS를 장착한 스마트폰, PDA, 휴대폰과 같은 모바일 기기들이 빠르게 보급됨에 따라, 이를 이용한 위치 기반 서비스가 확산되고 있다. 아울러 소셜 네트워크와 위치 정보를 결합한 서비스 모델의 등장으로 위치 데이터를 생성하는 사용자의 수가 급격히 증가하였다. 그 결과, 기존 위치기반 서비스의 데이터 소유자는 증가하는 사용자를 수용하고, 효율적인 질의 수행을 위해 공간 데이터를 아웃소싱 하고자 한다. 하지만, 사용자로부터 수집한 공간 데이터베이스를 그대로 아웃소싱 할 경우, 서비스 제공자는 이를 취득하여 제 3자에게 매도하거나 다른 용도로 데이터를 악용할 가능성이 있다. 따라서 데이터 소유자는 데이터베이스를 아웃소싱 할 때, 서비스 제공자를 포함하여 인증되지 않은 모든 사용자에게 원본 공간 데이터베이스가 유출되지 않기를 바란다. 또한, 질의 요청자는 질의를 수행하는 동안 자신의 위치가 노출되지 않으면서 정확한 질의 결과를 탐색하기를 원한다. 이를 위해, 공간 데이터 보호를 지원하는 암호화 기법 및 질의처리 알고리즘[1, 25 ,6 ,12 ,14 ,15]이 제안되었다. 그러나 기존 기법들은 유클리디언 공간 데이터를 위해 제안된 기법이며, 실제 사용자가 이동하는 도로 네트워크를 고려하지 않는다. 따라서 현재 아웃소싱 환경에서 도로네트워크를 고려한 공간 데이터 보호 기법에 대한 연구는 미약한 실정이다. 따라서, 본 논문에서는 기존 공간 데이터 암호화 기법을 아웃소싱 환경에서 도로 네트워크 기반 공간 데이터 보호를 위한 가공 데이터 생성 기법 및 암호화 기법으로 확장하여 수행한다. 아울러, 이를 이용하여 효율적으로 k-최근접점 질의를 수행하기 위한 k-최근접점 질의처리 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 공간데이터 보호를 위한 변환 기법에 대해 소개한다. 3장에서는 제안하는 도로 네트워크 기반 암호화 알고리즘을 설계하고 질의처리 알고리즘을 기술한다. 4장에서는 제안하는 기법과 기존연구와의 성능 평가를 수행한다. 마지막으로 5장에서는 결론 및 향후연구를 제시한다.

2. 관련 연구

본 장에서는 아웃소싱 데이터베이스에서 위치 데이터를 보호하기 위해 제안된 공간좌표 변환기법[5, 6, 15]과 거리 변환기법[7] 및 질의처리 알고리즘을 소개하고, 각 알고리즘의 장단점을 논의한다.

2.1 공간 좌표 변환기법

공간 좌표 변환기법은 아웃소싱된 공간 데이터베이스를 보호하기 위하여, 실제 공간 좌표를 변환한 후, 이를 서비스 제공자에게 전송한다. 서비스 제공자는 변환된 데이터베이스를 이용하여 사용자에게 서비스를 제공한다. 이를 위한 대표적인 연구로는 M. L. Yiu et al.[6]이 제안한 Hierarchical Space Division(HSD) 알고리즘과, Error-Based Transformation(ERB) 알고리즘, 두 기법을 결합한 HSD* 알고리즘이 있다. 또한 변환된 데이터베이스를 이용하여 수행하는 질의처리 알고리즘을 제안하였다. 먼저, HSD 알고리즘은 공간 데이터의 분포(distribution)를 변형시켜 데이터를 변환하는 기법이다. 이를 위해 데이터 소유자는 kd-tree로 분할된 원본 데이터베이스를 저장하며, 공격자에게 혼란을 주기 위해 원본 데이터의 분포를 타겟 분포를 이용하여 변환한다. <그림 1>은 HSD의 공간 좌표 변환 기법의 변환 예제를 보여준다.

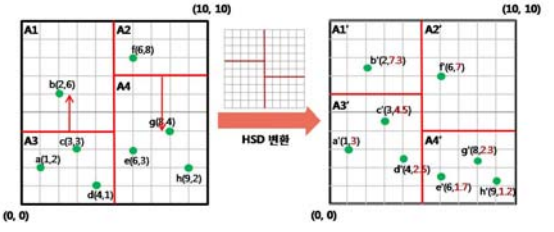


그림 1. HSD 변환 기법

둘째, ERB는 데이터에 에러를 삽입하여 공간 좌표를 변형하는 기법이다. 데이터 소유자는 에러율과

SHA-512[8] 보안 해시 함수(secure hash function)를 사용하여 공간 좌표 변환을 수행한다. 보안 해시 함수란 임의의 길이를 갖는 메시지를 입력받아, 고정 길이의 메시지를 생성하는 해시 함수를 의미한다. ERB는 사용자가 지정한 에러율과 보안 해시 함수를 이용하여 원본 데이터 좌표에서 일정 범위 내로 좌표를 변환한다. <그림 2>는 ERB 공간 좌표 변환 기법의 변환 예를 보인다. 에러율을 0.1 로 할 때, 좌표 a(1, 2)은 (0.9, 1.8), (1.9, 2.8)을 양 끝점으로 하고, 반지름이 $\sqrt{2}/2$인 원 영역 내 임의의 점으로 변환된다.

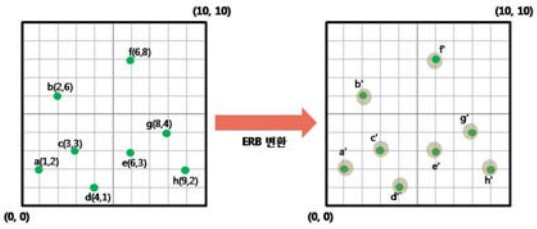


그림 2. ERB 변환 기법

셋째, HSD*는 HSD와 ERB를 결합한 기법으로, 데이터의 분포를 변환하고 또한 에러를 삽입한다. 데이터 소유자는 HSD와 동일하게 타겟 데이터 분포를 사용하여 원본 데이터베이스를 변환한 후, k-d 인덱스의 각 분할 영역 별로 보안 해시 함수를 사용하여 에러를 삽입함으로써 공간 좌표 변환을 수행한다.

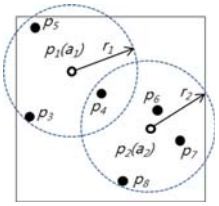
서비스 제공자에서 수행되는 변환된 데이터베이스 기반 질의 처리 알고리즘은 다음과 같다. 데이터 소유자는 공간 좌표 변환에 사용한 변환 키(key)를 인증된 사용자에게 전송하고, 인증된 사용자는 전송받은 변환 키를 통해 질의 영역을 변환한 후 서비스 제공자에게 전송한다. 마지막으로 서비스 제공자는 변환된 영역에 대해 질의를 처리한 후, 질의 결과를 인증된 사용자에게 전송하게 된다.

이와 같이, 공간 좌표 변환 기법은 변환된 가공 데이터를 서비스 제공자에게 전송하여 원본 데이터베이스의 노출확률을 감소시킨다. 하지만 기존의 공간 좌표 변환 기법의 경우, 공격자에게 원본 데이터베이스의 일부가 노출되면, 이를 변환된 데이터베이스와 대조하여 변환에 이용된 공식을 유도할 수 있다. 따라서 변환된 데이터에 이를 적용하면 원본 데이터베이스를 유추 가능하다. 또한, 변환된 데이터 간의 상대적인 거리가 유지되므로 원본 데이터베이스의 일부를 이용하여 데이터의 위치를 유추할 수 있다.

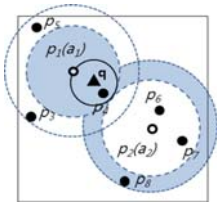
2.2 거리 변환기법

거리 변환기법은 실제 공간 데이터 간의 거리를 계산한 후, 계산된 거리 값을 변환한 가공데이터를 생성하여 서비스 제공자에게 전송한다. 서비스 제공자는 변환된 데이터간 거리를 기반으로 질의처리를 수행한다. 이를 위한 대표적인 연구로는 M. L. Yiu et al.[7]이 제안한 MPT (Metric Preserving Transformation)알고리즘이 있다. MPT 알고리즘의 수행과정은 다음과 같다. 먼저 데이터 소유자는 원본 데이터베이스에서 M-tree[13]를 이용해 공간 데이터를 그룹화 하기위한 대표 anchor 집합을 선정하고, 이로부터 인접한 데이터를 그룹화한 버켓을 구성한다. 구성된 버켓을 바탕으로 각 데이터와 anchor 간의 거리를 계산하고, 이를 OPES(Order Preserving Encryption Scheme) [10]을 적용하여 변환한다. OPES는 대표적인 순서유지 암호화 기법으로 데이터간의 순서(order)는 유지하면서 값의 분포를 변환하는 기법이다. 따라서 각 anchor에서 데이터까지의 순서가 유지된 변환 거리값 만을 서비스 제공자에게 전송하며 이를 바탕으로 최근접점 탐색 질의를 수행한다. <그림 3>은 MPT 기법의 변환 예제를 보여준다. 그림 (a)는 선택된 anchor와 생성 된 버켓의 모습을 나타내며 (c)는 각 버켓 내의 POI와 anchor간의 유클리디언 공간에서의 거리를 계산한 후 OPE를 적용한 결과이다.

MPT 거리 변환기법의 질의 처리 알고리즘은 다음과 같이 수행된다. 데이터 소유자는 거리 변환 에 사용한 변환 키와 anchor 집합의 정보를 인증된 사용자에게 전송한다. 인증된 사용자는 질의를 요청할 시, 자신으로부터 가장 가까운 anchor를 선택하고, 서비스 제공자에게 해당 anchor가 저장하고 있는 임의의 데이터 일부를 요청한다. 인증된 사용자는 전송 받은 임의의 데이터를 이용하여 최근접점 탐색 질의를 위한 영역을 계산하고, 암호화 키를 이용하여 OPES 적용 변환 후, 서비스 제공자에게 전송한다. 마지막으로 서비스 제공자는 변환된 영역을 바탕으로 최근접점 탐색 질의를 수행하고, 사용자에게 결과를 전송한다. <그림 3-(b)>는 질의 처리를 위한 최근탐색 영역 생성 예제를 보여준다.



(a) 대표 POI 선정 및 버켓 생성



(b) 질의 처리 과정

ID	Anchor ID	Distance
1	1	OPE(0.00) = 0.00
2	2	OPE(0.00) = 0.00
3	1	OPE(0.35) = 1.00
4	1	OPE(0.20) = 0.60
5	1	OPE(0.32) = 0.80
6	2	OPE(0.10) = 0.20
7	2	OPE(0.18) = 0.40
8	2	OPE(0.34) = 0.90

(c) 암호화된 데이터베이스

그림 3. MPT 기법의 예제

MPT 알고리즘은 OPES를 적용함으로써 위치 좌표가 아닌 변환된 거리만을 서비스 제공자에게 전송한다. 따라서, 공간 좌표 변환 기법에서 공격자가 변환 유도식 또는 일부 데이터 집합을 이용해 원본 좌표를 유추할 수 있는 공격 가능성을 감소시킨다. 아울러 저장된 거리 값을 획득하였다 할지라도 실제 데이터의 좌표를 알 수 없다는 장점을 지닌다. 하지만 MPT 알고리즘은 유클리디언 환경을 고려하여 제안된 기법으로 실제 사용자가 이동하는 도로 네트워크를 고려하지 않는다. 또한, 질의 수행 시 anchor로부터 POI까지 거리를 기반으로 수행하기 때문에 실제 질의지점과 멀리 위치한 POI가 탐색 영역 안에 포함되는 문제점을 지닌다. 따라서 후보 집합 반환 시 불필요한 POI를 포함하여 네트워크 효율을 감소시킨다. 또한, MPT 알고리즘은 질의 지점으로부터 가장 인접한 POI만을 탐색하는 최근접점 탐색 질의 처리만을 지원한다는 문제점을 지닌다.

3. 도로 네트워크 환경에서 암호화된 공간데이터를 위한 K-최근접점 질의 처리 알고리즘

본 장에서는 M. L. Yiu et al.[7]이 제안한 MPT 기법을 확장한 도로 네트워크 기반 암호화된 공간 데이터를 위한 K-최근접점 질의처리 알고리즘을 제안한다. 제안하는 기법은 전처리 단계를 통해 도로 네트워크 탐색 비용을 줄인다[16, 17]. 아울러, 암호화된 공간 데이터를 위한 최근접점 및 k-최근접점 질의 처리 알고리즘을 제안한다. 제안하는 질의 처리 알고리즘은 다음과 같은 장점을 지닌다. 첫째, 도로 네트워크 기반 질의 처리에서 발생하는 높은 네트워크 탐색 비용을 줄이기 위해 노드에 대표 POI까지의 거리를 전처리단계에서 저장함으로써, 최근접 대표 POI 탐색 시 노드 탐색만을 통해 빠르게 탐색 가능하다. 둘째, 샘플 POI를 통해 k-최근접접 탐색 영역을 설정하여 k-최근접접 질의 처리를 지원한다. 마지막으로 k-최근접점 탐색 영역을 재설정함으로써 질의 탐색 영역을 줄이고, 아울러 반환되는 후보 집합의 크기를 감소시킨다.

3.1 시스템 구조

본 논문에서 제안하는 기법의 시스템 구조는 <그림 4>와 같다. 전체 시스템은 데이터 소유자, 서비스 제공자 그리고 인증된 사용자로 구성된다. 아웃소싱 환경에서 위치 기반 서비스는 데이터 소유자와 서비스 제공자를 분리하여, 데이터 소유자는 데이터베이스를 구축하고 암호화하여 서비스 제공자에게 전송하고, 서비스 제공자는 인증된 사용자가 암호화된 공간 데이터베이스에 접근하여 정보를 검색할 수 있는 환경을 제공한다. 이는 개인 또는 회사가 IT 정보 서비스의 경험이 부족하다 할지라도 적은 비용으로 용이하게 공간 데이터베이스를 관리하고 위기기반 서비스를 제공할 수 있도록 해준다. 하지만 동시에, 데이터 소유자는 권한을 가진 사용자만이 암호화된 공간 데이터베이스에 접근하여 서비스를 사용하여, 자신의 데이터를 보호하기를 원한다. 따라서, 인증된 사용자에게만 암호화된 공간 데이터베이스에 대한 접근을 허용하고, 서비스 제공자에게 질의를 요청하여 서비스를 제공받도록 한다. 이를 위해, 데이터 소유자는 자신이 수집한 원본 공간 데이터베이스(P)를 보호하기위해 거리 값으로 가공된 데이터베이스(P′)를 생성하여 서비스 제공자에게 전송한다(1-a). 또한, 인증된 사용자에게 질의 수행을 위한 데이터 암호화 및 복호화 키를 전송한다(1-b). 향후 데이터 사용자는 필요에 따라 데이터베이스를 관리하기 위한 정보를 서비스 제공자

에게 전송한다(1-c).

한편, 아웃소싱 데이터베이스에서 암호화된 공간 데이터 기반 k-최근접점 질의 수행을 위해 인증된 사용자는 가공된 데이터에서 처리할 수 있는 형태로 질의를 변환해야 한다. 따라서, 데이터 소유자로부터 전송받은 키를 이용하여 질의를 변환하고, 이를 서비스 제공자에게 전송한다(2-a). 서비스 제공자는 처리한 질의 결과를 인증된 사용자에게 반환한다(2-b, 2-c). <표 1>은 본 논문에서 사용된 표기법 및 그 의미를 요약하여 정리한 것이다.

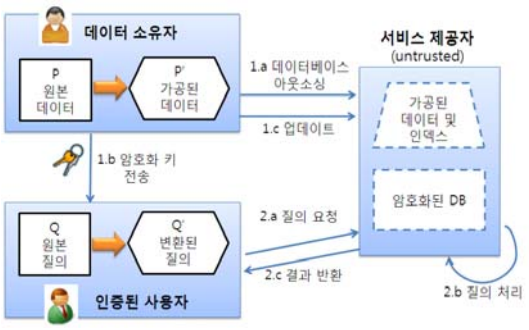


그림 4. 시스템 구조

3.2 K-최근접점 질의 처리를 위한 전처리 과정

3.2.1 버켓 생성 단계

데이터 소유자는 공간 데이터를 그룹화하기 위한 대표 POI를 선정한다. 이때 균일한 분포로 대표 POI를 선택하기 위해 기존 연구와 동일한 M-tree[13]를 이용하여 균일 분포를 지니는 대표 PO를 임의로 추출한다. 다음으로 선택된 대표 POI와 인접한 POI를 묶어 버켓을 생성하고, 도로 네트워크 확장을 통해 대표 POI와 POI간의 네트워크 거리를 계산하여 이를 저장한다.

3.2.2 OPES 적용 단계

버켓 생성 단계를 수행한 후 각 버켓의 대표 POI와 POI의 실제 도로 네트워크 거리 값에 OPES를 적용하여 암호화하는 단계이다. 가공 데이터 생성 및 암호화 단계는 네트워크 확장을 통해 탐색한 거리를 이용하여 MPT 기법[12]과 동일하게 수행된다. 이를 통해 생성된 가공 데이터는 서비스 제공자에게 전송되고, 암호화를 위해 사용된 암호화 키는 질의 수행을 위해 인증된 사용자에게 전송된다.

표 1. 논문에 사용된 표기법 및 의미

항목	의 미
P	원본 공간 데이터베이스
P′	가공된 공간 데이터베이스
POI	Point of Interest
POIs	A set of POI
(ai, ri)	anchor ai의 최대 거리 영역 ri
q	질의 지점 q=(x, y)
p.id	POI p의 id
pk	k-최근접 POI
dist(pi, pj)	두 개의 POI pi와 pj 사이의 거리
CK	암호화 key
OPE(v)	거리(value)에 OPE를 적용한 결과 값
E(x, y)	암호화된 공간 좌표
r	샘플링 데이터 수(전체 데이터의 %)
ɣ	k-최근접점 탐색의 최대 거리 값
k	k-최근접점 탐색 수
[min, max]	질의 범위의 최소, 최대 거리 값

3.2.3 버켓 내 대표 POI-노드 거리 계산 단계

M. L. Yiu et al.[7]이 제안한 MPT 기법은 유클리디언 공간을 기반으로 하기 때문에, 좌표 정보만을 이용하여 POI간 거리를 쉽게 계산 가능하지만, 도로 네트워크에서는 실제 네트워크거리를 탐색해야하므로 추가적인 탐색비용이 발생한다. 예를 들어 <그림 5>와 같이 사용자가 질의를 요청할 경우, 질의 지점으로부터 최근접 대표 POI를 찾기 위해 도로 네트워크를 확장하여 탐색한다. 최근접 대표 POI를 탐색하면 인증된 사용자는 서버에 샘플 데이터를 요청하고, 서버로부터 반환된 샘플데이터까지의 실제 거리를 계산하기 위해 다시 질의 지점에서부터 네트워크 탐색을 통해 각 POI까지의 실제 거리를 계산한다. 따라서 도로네트워크에서 최근접 대표 POI 및 샘플 POI를 탐색하기 위해 중복 확장을 수행하고, 마지막으로 탐색 영역 안에 포함되는 다른 대표 POI가 있는지 확인하기 위해 탐색 영역 내에 위치한 모든 도로 네트워크에 대해 추가 확장을 수행해야 한다. 이를 해결하기 위해, 제안하는 알고리즘에서는 데이터 소유자가 인증된 사용자에게 도로 네트워크의 각 노드에서 인접한 대표 POI까지의 거리정보를 추가 전송함으로써 네트워크의 확장 탐색 없이 노드탐색만으로 인접 대표 POI를 탐색 할 수 있도록 한다. 따라서 데이터 소유자는 인증된 사

용자에게 도로 네트워크를 구성하기 위한 노드 인덱스 V 및 도로 인덱스 E, 대표 POI 정보(위치 좌표 및 버켓 크기), 노드-대표 POI 간 거리정보 그리고 OPE 암호화 키를 전송한다. <그림 5>는 인증된 사용자가 전송받은 네트워크 정보를 나타낸다. 노드 n_5, n_6는 어느 대표 POI가 포함하는 영역에 위치하지 않기 때문에 인접한 대표 POI 정보를 지니지 않는다. 이와 같이, 전처리 단계에서는 원본 공간 데이터를 암호화를 수행하여 데이터 소유자의 공간 데이터 정보를 보호하는 동시에, 인증된 사용자가 암호화된 공간 데이터에서 효율적인 k-최근접점 탐색 질의 처리를 제공할 수 있도록 한다. k-최근접점 질의처리를 위한 전처리 과정은 알고리즘 1과 같다. 첫째, 도로네트워크 거리를 계산하기 위해 노드와 도로 인덱스를 메모리에 로딩하고(line 1), M-tree를 이용하여 균일한 분포로 대표 POI를 선택한다(line 2). 아울러 각 대표 POI의 버켓 내에 적정이상의 POI가 할당되도록 최소 포함 POI 수를 설정한다(line 3). 둘째, 버켓 생성을 위하여 모든 POI에서 대표 POI 까지 네트워크 거리를 계산하고, 가장 가까운 대표 POI의 버켓에 할당한다(line4-6). 셋째, 버켓 생성이 끝나면 대표 POI id를 생성하고, 각 대표 POI id에 할당된 POI리스트를 넣어주게 준다. 아울러, 각 대표 POI의 버켓 크기를 계산하고, 버켓 영역 안에 있는 모든 노드에 대표 POI 까지의 거리를 저장한다(line7-11). 넷째, 모든 버켓에 OPE를 적용하여 대표 POI와 POI의 거리를 이용한 가공데이터를 생성한다(line 12-13). 마지막으로 생성된 가공 데이터베이스를 서비스 제공자에게 전송하고, 인증된 사용자에게 도로네트워크 정보 및 대표 POI 정보, 노드-대표 POI 정보와 암호화 키를 전송한다(line14-15).

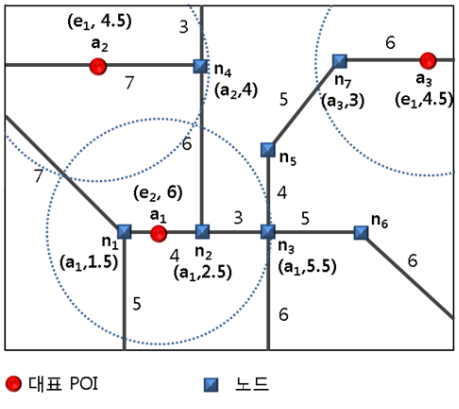


그림 5. 인증된 사용자에게 전송된 정보

Algorithm 1. Spatial Data Encryption for Data Owner

Input : Dataset P, Encryption Key CK, Integer A

Output : Encrypted P′, E(P),

```
1: loadMemory(Road_Network_index);
2: select a set of A anchor points from P;
3: B:= ⌈ |P|/A ⌉ ; //할당되는 POI 최소갯수 설정
4: for (each POI)
5:     calculate dist(p_i, ∀anchor);
6:     assign each POI to nearest anchor, until each
       anchor contains at least B number of POIs;
7: for 1 to A do
8:     let a_i be the i-th anchor point;
9:     let a_i.S be the set of POIs assigned the
       anchor a_i
10:    r_i:=max_{p∈ai.S} dist(a_i, p);//a_i의 버켓 사이즈 계산
11:    calculate dist(a_i, n.id) in r_i;
12:    for each POI p∈a_i.S do
13:        apply OPE to dist(a_i, p_i);
14:        send the encryped data<p.id, a.id, OPE(p_i,
           a_i), E(P)> to the server
15:    send Road_Network_index,
       encryption key CK, dist(a_i, n.id), a_i(e.id, x, y, r_i)
```

End Algorithm

3.3 암호화된 공간 데이터를 위한 K-최근접점 질의 처리 알고리즘

제안하는 기법은 변환 및 암호화된 데이터베이스에서 질의를 수행할 수 있도록 인증된 사용자가 데이터 소유자와 동일한 방식으로 자신의 위치 좌표 및 질의 영역을 변환 및 암호화하여 서비스 제공자에게 전송한다. 따라서, 본 절에서는 서비스 제공자가 소유한 변환된 공간 데이터를 이용한 k-최근접점 질의처리 알고리즘을 제안한다. 이때, k는 인증된 사용자가 탐색하고자 하는 POI의 수를 의미한다. 전처리 단계에서 데이터 소유자로부터 전송받은 정보를 바탕으로 인증된 사용자는 질의 요청을 위해 최근접 대표 POI를 탐색하여 자신의 위치 좌표를 거리 값으로 변환한다. 아울러 이를 암호화하여 변환된 질의 탐색 영역과 함께 서비스 제공자에게 전송한다. 기존 연구의 경우, 최근접 대표 POI를 통

해 생성된 질의 탐색 영역만을 이용하여 POI를 반환하기 때문에 불필요한 후보 집합을 반환한다는 문제점을 지닌다. 따라서 제안하는 암호화된 공간데이터를 위한 k-최근접점 질의처리 알고리즘은 최초 질의 탐색 영역에 포함된 대표 POI의 샘플 데이터를 송수신하여 k-최근접점 탐색 범위를 재조정함으로써 후보 집합의 크기를 줄인다. 또한 질의 탐색 영역에 포함된 모든 POI 탐색 시 전처리 과정을 통해 노드에 저장된 노드-대표 POI 간 거리정보를 이용하여 도로네트워크 탐색 수를 감소시킨다.

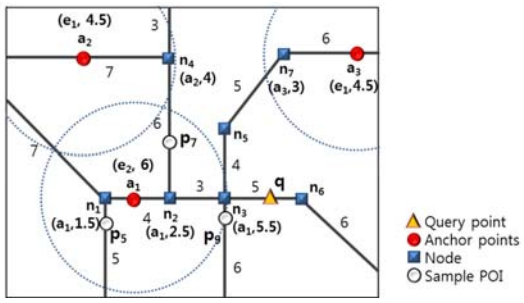


그림 6. 최근접 대표 POI 탐색

3.3.1 최근접 대표 POI 탐색 및 샘플 POI 탐색 단계

인증된 사용자는 데이터 소유자로부터 전송받은 도로 네트워크 정보 및 대표 POI 정보를 이용하여 자신의 위치에서 가장 가까운 대표 POI를 탐색한다. 이때 사용자는 대표 POI-노드 거리 정보를 이용하여 빠른 대표 POI 탐색을 수행하고, 서비스 제공자에게 최근접 대표 POI가 포함하는 r%의 샘플 POI를 요청한다. <그림 6>은 최근접 대표 POI 탐색 및 k-최근접점 탐색 질의 범위 설정 예제를 나타낸다. 인증된 사용자는 질의지점 q로부터 가장 가까운 대표 POI를 탐색하고자 할 때, 먼저 질의지점에서 인접한 노드 n_3, n_6를 탐색하고 노드에 대표 POI 정보가 있는지를 확인한다. 노드 n_3의 경우, (a_1, 2.5)의 정보를 통하여, n_3으로부터 가장 인접한 대표 POI는 a1이며, 거리는 2.5임을 알 수 있다. 반면 노드 n_6은 어떠한 대표 POI에도 속해 있지 않기 때문에 추가 정보를 저장하지 않는다.

3.3.2 k-최근접점 탐색 범위 설정 단계

인증된 사용자는 서비스 제공자에게 대표 POI a_1에 포함된 POI 중 r%의 샘플 POI를 요청하여 전송받는다. 이때 샘플 POI는 임의로 선택된다고 가정한다. 인증된 사용자는 질의 지점으로부터 전송받은 샘플 POI까지 거리를 계산하여 k번째 POI를 탐색한다. 아울러, 이를 이용하여 식(1)과 같이 k-최근접점 탐색 영역을 설정한다.

$$\gamma := \min dist(q, p_k) \quad (1)$$

예를 들어, 서비스 제공자에게 전송받은 샘플 POI의 정보가 <표 2>와 같을 때, 인증된 사용자는 자신의 위치 q로부터 도로 네트워크를 확장하여 샘플 POI에 대해 2-최근접점 질의를 수행한다. 이때, 질의 지점으로부터 샘플 POI {p_5, p_7, p_9}까지의 네트워크 거리가 각 {9, 8.5, 3.5}로 탐색되었을 때, 샘플 POI 중 질의 지점에서 2번째로 가까운 POI는 p_7이 된다. 따라서 식 (1)과 같이 질의 지점에서 p_7까지의 거리를 2-최근접점 탐색을 위한 최대 거리(γ)로 선택하고, 최대 거리 내에 다른 대표 POI의 버켓이 있는지 추가 확인한다.

표 2. 대표 POI 1(a1)의 샘플 POI

대표 POI id	POI id	OPE(dist)
1	5	0.5
1	7	7.5
1	9	1

이때, k-최근접점 탐색을 위한 질의 영역은 샘플 POI를 이용하여 계산된 k-최근접점 탐색 최대 거리 (γ)와 각 대표 POI의 범위 (a_i, r_i)를 이용하여 계산되며, 사용자의 위치와 대표 POI의 위치 관계에 따라 세 가지 경우로 계산 될 수 있다.

• 경우 1. 질의지점이 대표 POI 범위 밖에 위치한 경우

질의지점과 최근접 대표 POI 까지의 거리가 대표 POI가 가지는 버켓 크기보다 큰 경우, 즉 dist(q, a_i) > (a_i, r_i) 일 때, 최근접 대표 POI a_i로부터 k-최근접점 탐색을 위한 영역은 식(2)와 같이 계산된다.

$$[\min, \max] = [dist(a_i, p_k), (a_i, r_i)] \quad (2)$$

즉, a_i는 최대 버켓 크기 내에 있는 POI만을 반환하면 된다. <그림 7>은 대표 POI 범위 밖의 질의 지점에 대한 탐색 영역을 보여준다.

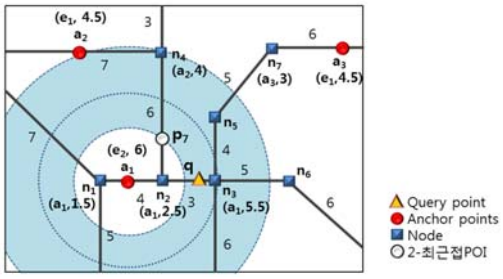


그림 7. 질의지점이 대표 POI 범위 밖에 위치한 경우

• 경우 2. 질의지점과 대표 POI 사이에 k-최근접점이 위치한 경우

질의지점과 대표 POI 사이에 k-최근접점이 위치할 때, 즉 dist(q, a_i) > dist(q, p_k) 의 거리 관계를 나타낼 때, 최근접 대표 POI a_i로부터 k-최근점점 탐색을 위한 질의 영역은 식(3)과 같이 계산된다.

$$[\min, \max] = [dist(a_i, p_k), dist(a_i, q) + dist(q, p_k)] \quad (3)$$

즉, 최근접 대표 POI a_i을 기준으로 생성되는 탐색영역은 a_i와 k-최근접점 p_k 간 거리부터 질의 지점 q와 a_i간 거리값에 q와 p_k 거리를 더한 영역으로 생성된다. <그림 8>은 질의지점과 대표 POI 사이에 k-최근접점이 위치한 경우를 보여준다.

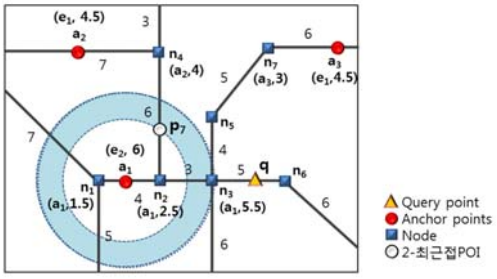


그림 8. 질의지점과 대표 POI 사이에 k-최근접점이 위치한 경우

• 경우 3. 질의지점과 대표 POI의 거리가 k-최근접점보다 가까운 경우

질의지점과 대표 POI와의 거리가 질의지점과 k-최근접점보다 가까운 경우, 즉 dist(q, a_i) < dist(q, p_k) 의 경우, 최근접 대표 POI a_i로부터 k-최근점점 탐색을 위한 질의 영역은 식(4)와 같이 계산된다.

$$[\min, \max] = [(a_i, 0), dist(q, p_k)] \quad (4)$$

질의지점 q와 최근접 대표 POI a_i가 k-최근접점 p_k보다 가까운 위치에 있다면, 실제 q와 a_i 거리 내에 있는 POI는 후보 집합이 될 수 있다는 것을 의미한다. 아울러, q에서부터 p_k거리 내에 있는 POI도 후보 집합이 될 수 있다. 즉, a_i 의 최소 거리 0으로부터 질의 지점 q와 p_k 간 거리 질의 탐색 영역으로 설정한다. <그림 9>는 질의지점과 대표 POI의 거리가 k-최근접점보다 가까운 경우를 나타낸다.

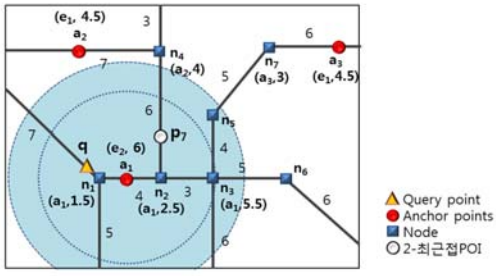


그림 9. 질의지점과 대표 POI의 거리가 k-최근접점보다 가까운 경우

3.3.3 탐색 범위내의 후보 대표 POI 탐색 및 질의 결과 영역 설정 단계

이전 수행단계를 통해 k-최근접점 탐색 영역이 설정되면 인증된 사용자는 생성된 최대 질의 범위에 포함되는 모든 대표 POI를 탐색한다. 이는 다른 대표 POI가 저장한 POI가 질의 범위 내에 위치할 가능성이 있기 때문이다. 따라서 노드-대표 POI간 거리정보를 이용하여 후보 대표 POI를 탐색한다. 만약 후보 대표 POI가 존재한다면, 샘플 POI를 요청하여 질의 지점부터 k개의 POI를 찾을 때까지 탐색하고 질의 지점까지의 거리(dist(q, c.pk))를 계산한다. 만약 기존에 설정된 탐색 영역 중 max의 값이 dist(q, c.pk)보다 크다면 max에 dist(q, c.pk)을 넣고, k-최근접점 질의 탐색 영역을 다시 생성한다. 마지막으로 재조정된 영역에 OPES를 적용하여 이를 서비스 제공자에게 전송하고, 서비스 제공자는 암호화된 탐색 영역을 바탕으로 암호화된 공간 데이터베이스에서 질의를 처리하여 후보 집합을 반환한다. 이때, 서비스 제공자는 질의지점의 위치 정보 및 POI의 실제 위치 정보를 알 수 없으므로, 거리를 기반으로 질의 결과 후보 집합을 생성하며, 해당하는 POI에 대해 암호화된 공간 정보를 함께 전송한다. 따라서 인증된 사용자는 데이터 소유자로부터

전송받은 키를 이용하여 질의 결과 후보 집합을 복호화하고, 실제 네트워크 거리를 계산하여 실제 K-최근접점을 선택한다. 기존연구의 경우 초기 설정된 질의 영역에 대해 모든 POI를 탐색하므로. 포함되는 대표 POI수가 많고 질의 결과 후보 POI 또한 증가한다. 그러나 제안하는 기법은 질의 영역 재설정 과정을 통해 질의 영역을 작은 영역으로 변경하면서 후보 결과 POI를 탐색하기 때문에, 결과적으로 질의 결과 POI 수를 감소시켜 네트워크 효율을 높이는 장점을 지닌다.

<그림 10>에서 대표 POI a_1에 대한 2-최근접점 탐색 영역이 [3.5, 9]로 생성 되었을 때, 인증된 사용자는 영역 안에 다른 대표 POI가 포함되어 있는지를 탐색한다. 이를 위해, q지점에서 도로를 확장하며 노드에 대표 POI 정보가 저장되어 있는지를 확인한다. 노드 n_4에 저장된 정보를 이용하여 노드 n_4는 대표 POI a_2의 버켓에 포함되어 있으며 거리 값은 4 임을 알 수 있다. 반면 n_5는 인접한 대표 POI정보를 저장하지 않으며, n_7의 경우 저장된 대표 POI 정보 a_3까지의 거리가 질의 탐색 영역 밖에 존재하고 있으므로 더 이상 탐색을 하지 않는다. n_4에 저장된 a_2의 거리를 통해 q에서부터 a_2까지의 거리는 12 라는 것을 계산할 수 있다. 인증된 사용자는 이를 통해, a_2의 범위 [3, 4.5] 영역이 생성된 질의 탐색 영역 안에 포함됨을 알 수 있다. 따라서 인증된 사용자는 a_2가 포함하는 영역에서 질의 영역과 교차하는 범위에 위치한 샘플 데이터를 서비스 제공자에게 요청한다. 샘플 데이터를 전송받으면 사용자는 이를 바탕으로 최근접 대표 POI와 동일한 방법으로 k-최근접점 질의를 수행하며, k번째 POI를 탐색하여 초기 설정한 질의 탐색 영역보다 작은 영역을 포함할 경우 이를 재조정한다.

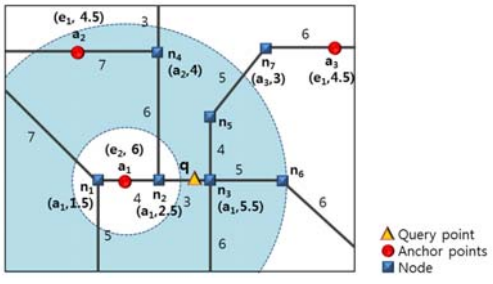


그림 10. 질의 거리 영역 내의 모든 대표 POI 탐색

<그림 11>은 a_2에 대해 생성된 k-최근접 질의 영역과 재설정된 a_1의 질의 탐색 영역을 나타낸다. 따라서 사용자는 재설정된 a_1질의 탐색 영역에 대한 POI 후보 집합을 서비스 제공자에게 요청하고, 후보 집합을 반환받는다.

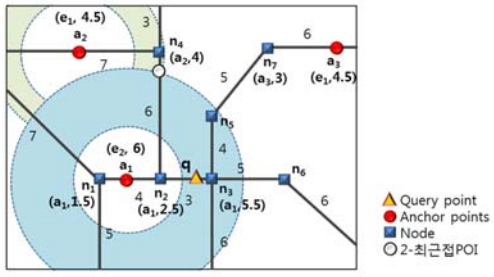


그림 11. k-최근접점 질의 탐색 영역 재설정

알고리즘 2는 암호화된 공간 데이터를 위한 k-최근접점 질의 처리 시, 인증된 사용자와 서비스 제공자 간에 수행되는 알고리즘을 나타낸다. 먼저, 인증된 사용자는 데이터 소유자로부터 전송받은 각 대표 POI의 정보 및 도로네트워크 정보를 로딩한다 (line 1). 로딩한 도로네트워크 정보를 바탕으로 자신의 위치에서 가장 가까운 대표 POI를 탐색한다 (line 2-3). 탐색된 최근접 대표 POI에 대해 샘플 데이터를 요청하고 샘플 데이터를 반환받는다 (line 4-5). 이를 기반으로 인증된 사용자는 질의 탐색 영역을 생성한다(line 6-7). 두 번째 단계에서는 질의 범위 내 다른 후보 대표 POI가 위치하는지 확인하기 위해 질의 범위 내 노드를 탐색한다(line 8). 만약 노드에 대표 POI정보가 있다면, 탐색 질의 영역과 가장 많이 겹쳐지는 대표 POI를 확인하고 서비스 제공자에게 해당 대표 POI에 대한 샘플 데이터를 요청한다(line 9-10). 전송받은 샘플 데이터를 바탕으로 q에서부터 k-최근접점 POI를 탐색한다(line 11). 탐색된 POI에서 q까지 거리를 maxDist로 설정하고, 최근접 대표 POI에서 설정된 질의 탐색 영역을 재설정한다(line 12-14). 마지막으로, 이를 OPES 적용시킨 후 서비스 제공자에게 전송하고, 서비스 제공자는 질의 영역 내의 후보 집합을 반환한다 (line 15-17).

Algorithm 2. Spatial k-NN Query Processing algorithm for Client

Input : Query point q, Encryption Key CK, Integer A, the range of anchors (a_i, r_i), random value r, the number of searching POIs k(k-NN), network information

Output : Query result set R

/* the first round */

1: loadMemory(Road_Network_index);

2: γ:=k-$min_{i\in[1,A]}$ dist(q,a_i);

3: let a_q be the nearest anchor to q;

/* the second round */

4: request the server for r number of sample POIs whose anchor ID equals to that of a_q;

5: let S be the set of random samples;

6: **for** each p∈S **do**

7: γ:=k-$min_{i\in[1,A]}$ dist(q,p);
[min,max]=(dist(q,a_i))-γ, (dist(q, a_i))+γ];

/* the third round */

8: expansion 노드s to find candidate anchor;

9: if(노드_flag) *//노드에 인접 anchor 정보 확인*

10: request the server for r number of sample POIs whose anchor ID equals to that of range[min,max];

11: find k-NN based on sample POIs from q;
let $c.p_k$ be the k-NN POI;

12: **if**(maxDist > dist(q, c.pk))

13: maxDist = dist(q, c.pk);

14: [min,max] = [dist(q,a_i), maxDist]; *//질의 결과 영역 설정*

15: request the server for all POIs whose OPE(dist(ai,p)) falls into the range [OPE([min,max])];

16: let R be the set of decrypted POI information from the received data

17: **return** the POI p∈R whose distance satisfies k-NN to q

4. 성능평가

4.1 실험 환경

본 장에서는 제안하는 공간 데이터 암호화 기반 k-최근접점 질의 처리 알고리즘의 우수성을 검증하기 위하여 성능평가를 수행한다. 성능평가 항목으로는 탐색 횟수, 질의 처리 시간, 후보 집합 크기를 비교한다. 여기서 탐색 횟수는 대표 POI를 찾기 위한 도로 탐색 및 샘플데이터를 탐색한 총 횟수를 의미한다. 한편 성능평가의 실험 환경은 <표 3>와 같다.

표 3. 실험 환경

항목	성능
CPU	Intel Pentium(R) Dual-Core E6600 3.06GHz
Memory	2GB
OS	Windows 7
Compiler	Microsoft Visual Studio 2008

성능평가는 223,200개의 도로와 175,344개의 노드로 구성되는 미국 샌프란시스코의(600㎢) 실제 도로 네트워크 데이터를 이용하여 수행하였다. 또한 성능평가에 사용된 POI 데이터는 Net- work-based Generator of Moving Objects[11]를 사용하여, 미국 샌프란시스코의 도로 네트워크를 기반으로 생성하였다. 전체 네트워크의 POI 밀도가 0.1인 경우 생성된 POI 수가 전체 네트워크 도로 수의 0.1% 임을 의미하며, 총 22,025개의 POI가 생성된다. <표 4>는 성능평가에 사용된 매개변수를 나타낸다. 아울러, 기존 기법인 MPT는 유클리디언 공간을 기반으로 제안된 기법이므로 성능 비교를 위해 도로네트워크로 확장하여 제안하는 기법과 성능비교를 수행한다.

표 4. 실험 환경 매개변수

parameter	range	default
k (요구 POI 수)	1, 3, 5, 10	5
대표 POI (선택한 대표 POI의 수)	100, 200, 300, 400, 500	300
density (전체 네트워크의 POI 밀집도)	0.01, 0.02, 0.05, 0.1	0.1

4.2 인덱스 생성시간에 대한 성능평가

<그림 12>는 POI의 밀도에 따른 전처리 시간을 나타낸다. 여기서 density가 0.1인 경우는 0.01인 경우에 비해 상대적으로 POI의 밀집도가 상대적으로 높음을 의미한다. 전처리 시간은 공간 데이터 인덱스를 생성하고 이를 아웃소싱 하기위해 가공데이터 생성 및 암호화하는 단계로 버켓 생성시간과 OPES 적용 시간을 포함한다. 제안하는 기법의 경우 대표 POI가 포함하는 영역 내에 위치한 노드를 탐색하는 시간이 추가된다. 버켓 생성 시간은 도로 네트워크 인덱스 로딩시간과 대표 POI 선택 및 POI를 할당

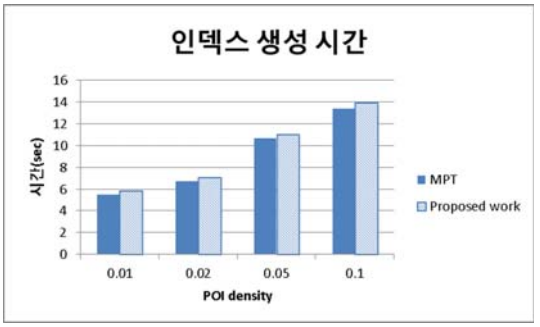


그림 12. 전체 네트워크의 POI 밀집도에 따른 인덱스 생성 시간(k=5, 대표 POI=300)

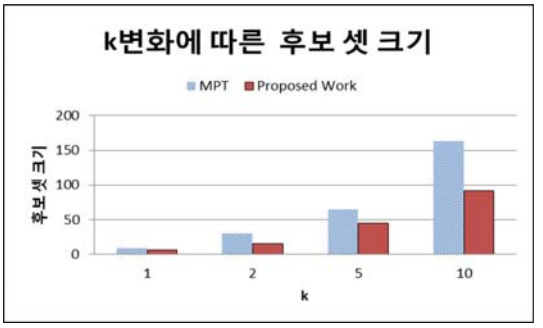


그림 13. k 변화에 따른 후보 집합 크기 (대표 POI=300, density=0.1)

하는 작업에 소요되는 시간을 의미한다. OPES 적용 시간은 버켓 생성 후 대표 POI와 POI간 실제 도로 네트워크의 거리를 암호화하는데 소요되는 시간을 의미한다. 대표 POI 정보 생성 시간은 대표 POI의 버켓 내에 있는 노드와 대표 POI간의 거리를 계산 후 이를 각 노드에 저장하는 시간을 의미하며, 질의 지점으로부터 대표 POI 탐색 수를 줄이기 위해 제안하는 연구에서 추가로 수행된다. 버켓 생성 시간의 경우, POI 수가 가장 많은 0.1이 다른 POI 수에 비해 많은 시간이 필요함을 알 수 있다. 이는 POI 수의 양이 많을수록 대표 POI와 POI 간 거리 계산 양이 많기 때문이다. 또한, OPES 적용시간도 POI의 수가 많을수록 증가함을 알 수 있다. 대표 POI 정보 생성 시간의 경우, POI 수가 증가함에 따라 미세한 차이를 보이는데, 이는 POI 수가 적더라도 대표 POI 가 가지는 버켓 사이즈가 크기 때문에 저장하는 노드 수가 비슷하기 때문이다.

4.3 k 변화에 따른 성능평가

<그림 13>은 탐색 POI 수 k에 따른 후보 집합 크기를 나타낸다. 모든 기법이 k가 증가함에 따라 후보 집합 크기가 증가함을 보인다. 이는 사용자가 요구하는 k가 클수록, 인증된 사용자에 의해 생성되는 질의 영역이 커지기 때문이다. MPT는 인접한 대표 POI를 기반으로 생성된 질의 영역이 클 경우, 영역 안의 모든 대표 POI를 재탐색하고 탐색된 대표 POI에 포함되는 모든 후보 집합을 반환한다. 반면, 제안하는 기법은 추가 대표 POI를 탐색 후 탐색 질의 영역을 점진적으로 재설정함으로써 후보 집합의 불필요한 탐색을 줄인다. 따라서 제안하는 기법이 MPT에 비해 모든 k에서 평균 약 41% 감수된 후보 집합의 수를 반환한다.

<그림 14>는 k에 따른 네트워크 탐색 수를 나타낸다. 모든 기법이 k가 증가함에 따라 네트워크 탐색 수가 증가함을 보인다. 이는 사용자가 요구하는 k가 클수록, 질의 지점으로부터 먼 지점까지 네트워크를 확장하여 후보 POI 및 대표 POI를 탐색하기 때문이다. MPT의 경우, 질의 지점에서 대표 POI 탐색을 위한 확장과 샘플 데이터를 탐색하기 위한 확장, 질의 지점에 포함되는 추가 대표 POI를 탐색하기 위한 확장이 필요하다. 반면, 제안하는 기법은 노드에 인접 대표 POI까지의 거리를 저장하고 있기 때문에, 질의 지점에서 인접한 노드만을 탐색하여 빠르게 대표 POI를 탐색할 수 있다. 아울러, 생성된 질의 영역 안에 존재하는 다른 대표 POI도 추가 확장 없이 탐색이 가능하다. 따라서 제안하는 기법은 모든 경우에서 기존연구에 비해 탐색 수가 평균 약 44% 감소함을 보인다.

<그림 15>는 k 변화에 따른 질의 처리 시간을 나타낸다. 모든 기법이 메모리기반의 탐색을 수행하였으며, k가 증가할수록 질의 처리 시간이 증가함을 알 수 있다. 이는 k가 증가함에 따라 필요한 네트워크 탐색 시간이 증가하기 때문이다. MPT의 경우, k변화에 따른 질의 처리 시간은 각각 0.124초, 0.36초, 0.842초, 1.826초로 증가한다. 제안하는 기법의 경우, 0.102초, 0.32초, 0.636초, 1.442초로 기존 연구보다 약 18~21%의 빠른 질의 처리 시간을 보인다. 이는 제안하는 기법은 전처리 과정을 통해 노드에 저장된 대표 POI까지의 거리를 이용하여 빠르게 대표 POI 탐색을 수행하기 때문이다. 아울러, 추가 대표 POI 탐색 시에도 질의 영역을 모두 탐색하지 않고 노드 정보를 통해 인접 대표 POI를 빠르게 탐색

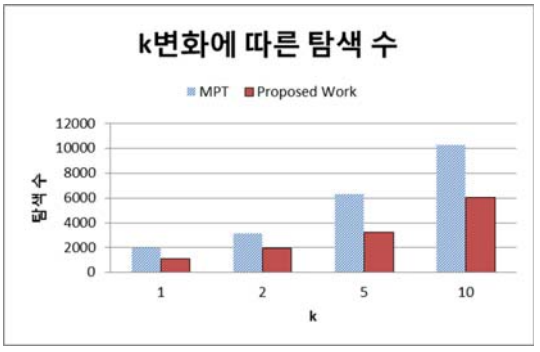


그림 14. k 변화에 따른 질의 처리 시간
(대표 POI=300, density=0.1)

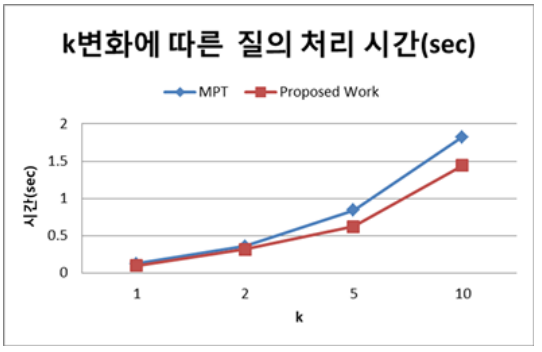


그림 15. k 변화에 따른 질의 처리 시간
(대표 POI=300, density=0.1)

하기 때문이다.

4.4 대표 POI 수 변화에 따른 성능평가

<그림 16>은 대표 POI 수 변화에 따른 후보 집합 크기를 나타낸다. 모든 기법이 대표 POI 수가 증가함에 따라 반환되는 후보 집합의 크기가 감소함을 알 수 있다. 이는 대표 POI 수가 많아지면 각 대표 POI가 포함하는 POI수가 감소하고, 버켓의 크기가 작아져 비교적 작은 질의 탐색 영역을 생성하기 때문이다. 제안하는 기법의 경우, 최근접 대표 POI를 통해 생성된 질의 영역에 겹쳐지는 다른 대표 POI를 통해 탐색 영역을 재조정하기 때문에 MPT에 비해 더 적은 양의 후보 집합을 반환한다. 한편, 두 기법 모두 대표 POI가 400개와 대표 POI가 500개 일 때, 후보 집합 크기의 감소차가 적어지는 것을 볼 수 있다. 이는 질의지점이 대표 POI와 가깝다 할지라도 샘플 데이터를 탐색하여 생성된 질의 탐색 영역은 일정 크기 이하로 작아지지 않기 때문이다. 이를 통해, 대표 POI 수가 POI의 수의 일정 비율 이상 생성되면 성능에 큰 영향을 끼치지 않음을 나타낸다.

<그림 17>은 대표 POI 수 변화에 따른 탐색 수를 나타낸다. 모든 기법이 대표 POI 수가 증가함에 따라 탐색 수가 감소하는 것을 알 수 있다. 이는 대표 POI의 수가 증가하면 질의 지점에서부터 인접한 대표 POI가 근거리에 확률이 높아져 탐색 수가 감소하기 때문이다. 아울러, 질의 지점과 대표 POI의 거리가 작아지면 탐색 하는 질의 영역의 크기도 감소하게 된다. 또한 제안하는 기법은 노드와 대표 POI간 거리정보를 이용하여 MPT보다 평균 약 51% 탐색 수 감소를 보인다.

<그림 18>은 대표 POI 수 변화에 따른 질의 처리 시간을 나타낸다. 모든 경우에 있어서 제안하는 기법이 MPT에 비해 좋은 성능을 나타냄을 알 수 있다. 제안하는 기법이 질의 처리 시간 측면에서 MPT보다 평균 약 1.5배의 우수한 성능을 보인다. POI의 수가 증가하면, 대표 POI 수 또한 상대적으로 증가하여 모든 기법에서 추가적으로 탐색되는

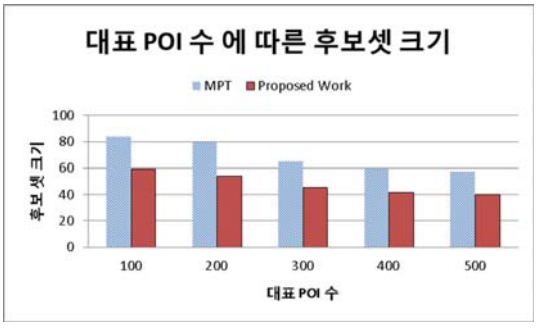


그림 16. 대표 POI 수 변화에 따른
후보 집합 크기 (k=5, density=0.1)

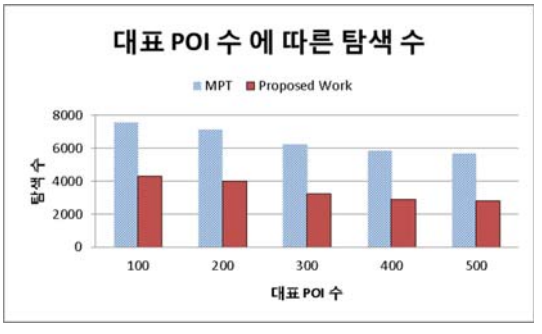


그림 17. 대표 POI 수 변화에 따른
네트워크 탐색 수 (k=5, density=0.1)

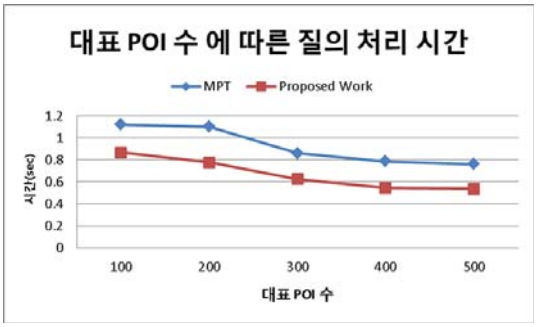


그림 18. 대표 POI 수 변화에 따른 질의 처리 시간(k=5, density=0.1)

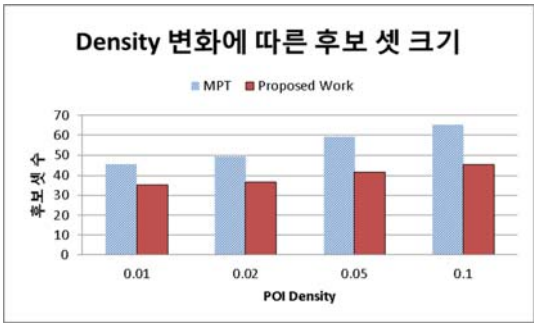


그림 19. 전체 네트워크의 POI 밀집도에 따른 후보 집합 크기 (k=5, 대표 POI=300)

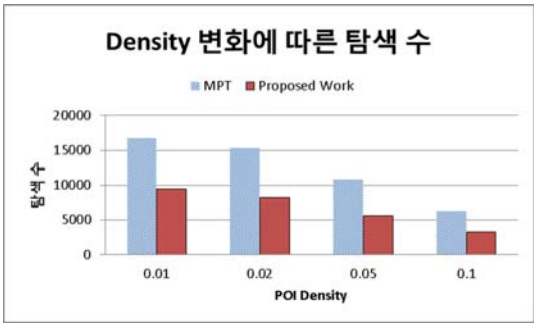


그림 20. 전체 네트워크의 POI 밀집도에 따른 네트워크 탐색 수 (k=5, 대표 POI=300)

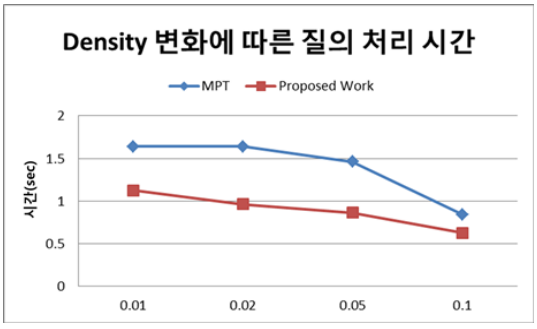


그림 21. 전체 네트워크의 POI 밀집도에 따른 질의 처리 시간 (k=10, 대표 POI=300)

POI의 수가 증가한다. 특히 MPT의 경우, 생성된 질의 탐색 영역에 포함된 대표 POI 탐색을 위해 추가적으로 네트워크 확장을 수행해야 한다. 반면, 제안하는 기법은 질의 탐색 영역이 크게 생성되더라도 노드에 저장된 대표 POI간 거리를 통해 생성된 질의 탐색 영역 안에 포함된 대표 POI를 빠르게 탐색 할 수 있다.

4.5 전체 네트워크의 POI 밀집도 변화에 따른 성능 평가

<그림 19>는 총 POI 수 변화에 따른 후보 집합 크기를 나타낸다. 모든 기법이 POI 밀집도가 감소함에 따라 후보 집합의 크기 또한 감소함을 보인다. 이는 density에 따라 생성된 질의 탐색 영역의 크기가 유사할지라도, 그 안에 위치하고 있는 POI의 수가 적기 때문이다.

<그림 20>은 총 POI 수 변화에 따른 탐색 수를 나타낸다. 모든 기법이 POI 밀집도가 작아짐에 따라 POI 탐색 수가 증가함을 보인다. 이는 POI 밀집도가 작을수록 주변에 위치한 대표 POI 및 POI 수가 적기 때문에, 사용자가 요구하는 k개의 POI를 찾기 위해 보다 많은 확장이 필요하기 때문이다. MPT의 경우 POI의 밀집도가 낮을수록 대표 POI 탐색 및 샘플 데이터의 탐색 수는 큰 폭으로 증가함을 알 수 있다. 반면 제안하는 기법은 POI의 수와 대표 POI의 수가 적다할지라도 대표 POI 정보를 가지고 있는 노드까지만 탐색을 수행하기 때문에 불필요한 탐색을 줄인다.. 따라서 모든 경우에서 제안하는 기법의 네트워크 탐색이 MPT에 비해 약 50% 정도 감소함을 알 수 있다.

<그림 21>은 전체 POI 수 변화에 따른 질의 처리 시간을 나타낸다. 모든 기법이 전체 POI 수가 증가함에 따라 질의 처리 시간이 감소함을 보인다. 이는 총 POI 수가 많을수록 질의 영역 내 포함되는 POI가 증가하여 k를 만족하기 위한 확장을 수행하지 않기 때문이다. MPT의 경우, POI의 수가 적으면 대표 POI 탐색 및 샘플 데이터 탐색에 많은 시간을 소요한다. 한편, 제안하는 기법은 전처리를 통

해 대표 POI를 찾기 위한 탐색 수가 감소하는 만큼 질의 처리에 필요한 시간도 감소한다. 이는 네트워크를 확장하며 대표 POI를 탐색할 필요 없이, 노드에 저장된 대표 POI간 거리를 이용하여 빠르게 질의 처리를 수행하기 때문이다. 이를 통해, POI 탐색 횟수가 질의 처리 시간에 영향을 미친다는 것을 알 수 있으며, 제안하는 기법이 모든 경우에서 개선된 성능을 보임을 알 수 있다.

5. 결 론

본 논문에서는 아웃소싱 환경에서 도로네트워크를 고려한 암호화된 공간 데이터베이스 기반 k-최근접점 질의 처리 알고리즘을 제안하였다. 이를 위해, 데이터 소유자는 버켓 내의 대표 POI 및 POI간 네트워크 거리 정보를 이용하여 가공 데이터를 생성하고, 이를 암호화 하여 서비스 제공자에게 전송한다. 따라서 신뢰할 수 없는 서비스 제공자는 도로 네트워크 정보를 알고 있다 하더라도 암호화된 데이터의 거리 값만을 가지고 있으므로 실제 위치 좌표를 유추할 수 없다. 한편, 서비스 제공자가 가공 데이터를 이용하여 효율적으로 질의를 수행하기 위한 k-최근접점 질의처리 알고리즘을 제안한다. 제안하는 기법은 질의 처리 시 도로네트워크 확장을 빠르게 수행하기 위해, 전처리 과정을 통해 대표 POI에서 각 POI까지의 도로네트워크 거리를 노드에 미리 저장한다. 따라서, 질의 요청자는 자신의 위치, 즉, 질의 지점으로부터 가장 인접한 대표 POI를 빠르게 탐색한다. 또한, 거리 기반으로 수행되는 질의에 따른 불필요한 후보 탐색을 감소시키기 위해, k- 최근접점 질의 영역을 재설정함으로써 후보 집합의 크기를 감소시킨다. 마지막으로, 성능평가를 통해 제안하는 기법이 기존 연구에 비해 후보 집합의 크기 및 탐색 수, 질의 처리 시간 측면에 있어 우수한 성능을 보임을 입증하였다.

향후 연구로는 거리 기반 가공데이터를 이용한 질의 수행에서 도로 네트워크의 방향성을 고려하여 후보 집합의 크기를 줄이고, 불필요한 확장을 감소하는 기법을 연구하는 것이다.

참 고 문 헌

[1] A. Khoshgozaran, C. Shahabi, 2007 “Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy,” Proceedings of the 10th international conference on Advances in spatial and temporal databases, pp.239-257.

[2] A. Khoshgozaran and C. Shahabi, 2009, “Private Buddy Search: Enabling Private Spatial Queries in Social Networks,” Proceedings of the International Conference on CSE, Volume 04, pp.166-173.

[3] D. Sacharidis, K. Mouratidis and D. Papadias, 2010, “k-Anonymity in the Presence of External Databases”, Journal IEEE TKDE, Volume 22 Issue 3, pp.392-403.

[4] L. Qiu, Y. J. Li and X. Wu, 2008, “Protecting business intelligence and customer privacy while outsourcing data mining tasks”, Journal Knowledge and Information Systems, Volume 17 Issue 1, pp.99-120

[5] M. L. Yiu, G. Ghinita, C. S. Jensen and P. Kalnis, 2009, “Outsourcing of Private Spatial Data for Search Services”, Proceedings of the IEEE ICDE, pp.1140-1143.

[6] M. L. Yiu, G. Ghinita, C. S. Jensen and P. Kalnis, 2010, “Enabling Search Services on Outsourced Private Spatial Data,” The International Journal on VLDB, Volume 19 Issue 3, pp.363-384.

[7] M. L. Yiu, I.Assent, C. S. Jensen and P. Kalnis, 2012, “Outsourced Similarity Search on Metric Data Assets,” Journal IEEE TKDE, Volume 24 Issue 2, pp.338-352.

[8] National Institute of Standards and Technology. Secure Hashing. http://csrc.nist.gov/groups/ST/toolkit/secure_hashing.html

[9] P. Ciaccia, M. Patella, and P. Zezula. “M-tree: An Efficient Access Method for Similarity Search in Metric Spaces,” In VLDB, pages 426 - 435, 1997

[10] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu., 2004, “Order-Preserving Encryption for Numeric Data,” Proceedings of the ACM SIGMOD, pp.563-574.

[11] T. Brinkhoff, 2002, “A Framework for Generating Network-Based Moving Objects,” Journal Geoinformatica Volume 6 Issue 2, pp.153-180.

[12] T. Wang, L. Liu, 2009, “Privacy-Aware Mobile Services over Road Networks,” Journal Proceedings of the VLDB Endowment, Volume 2 Issue 1, pp.1042-1053.

[13] Wong, W. K., Cheung, D. W., Kao, B., Mamoulis, 2009, “Secure k-NN computation on encrypted databases,” Proceedings of the 35th SIGMOD, pp.139-152

[14] X. Jiang, J. Gao, T. Wang, and D. Yang, 2010, “Multiple sensitive association protection in the outsourced database,” Proceedings of the 15th international conference on DSFAA, Volume Part II, pp.123-137.

[15] Y. Yang, S. Papadopoulos, D. Papadias and G. Kollios, 2008, “Spatial Outsourcing for Location-based Services”, Proceedings of the IEEE 24th ICDE, pp.1082-1091.

[16] 김용기, 김아름, 장재우, 2008, “공간 네트워크 데이터베이스에서 공간 제약을 고려한 경로 내 최근접 질의처리 알고리즘” 한국공간정보시스템학회 논문지, 제10권, 제3호, pp. 19-30.

[17] 김용기, 장재우, 2007, “공간 네트워크 데이터베이스에서 POI 기반 실체화 기법을 이용한 Closest Paris 및 e-distance 조인 질의처리 알고리즘” 한국공간정보시스템학회 논문지, 제9권, 제3호, pp. 67-80.

논문접수 : 2012.02.15
수 정 일 : 2012.06.12
심사완료 : 2012.06.18

장 미 영

2009년 전북대학교 컴퓨터공학과 졸업 (학사)
2011년 전북대학교 컴퓨터공학과 졸업 (석사)
2011년~현재 전북대학교 컴퓨터공학과 박사과정
관심분야는 공간 데이터베이스, 위치 정보 보호, 궤적 데이터 마이닝

장 재 우

1984년 서울대학교 전자계산기 공학과 (공학사)
1986년 한국과학기술원 전산학과(공학석사)
1991년 한국과학기술원 전산학과(공학박사)
1996년~1997년 Univ. of Minnesota, Visiting Scholar
2002년~2004년 Penn State Univ., Visiting Scholar
1991년~현재 전북대학교 IT 정보공학과 교수
관심분야는 공간데이터베이스, 클라우드 컴퓨팅, 데이터베이스 정보보호