위치 기반 서비스에서 K-anonymity를 보장하는 가중치 근접성 그래프 기반 최근접 질의처리 알고리즘

A Nearest Neighbor Query Processing Algorithm Supporting K-anonymity Based on Weighted Adjacency Graph in LBS

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

요 약 무선 통신 기술 및 GPS(Global Positioning System)등의 발달로 인하여 위치 기반 서비스(Location-Based Services: LBS)가 크게 발전하는 추세이다. 그러나 위치 기반 서비스를 이용하기 위해 질의 요청자는 자신의 정확한 위치 정보를 위치 기반 서비스 제공자에게 전송해야 한다. 따라서 위치 기반 서비스를 제공하면서 질의 요청자의 위치 정보를 보호하는 것이 중요한 과제이다. 이 문제를 해결하기 위해, 기존 기법은 실제 사용자의 위치를 숨기며 네트워크 사용을 줄일 수 있는 2PASS 기법을 사용하였다. 그러나 이 기법은 실제 사용자 분포를 고려하지 않기 때문에 실제 사용자 위치 보호를 완전히 보장하지 않는다. 따라서 본 논문에서는 K-anonymity를 보장하는 가중치 근접성 그래프 기반 최근접 질의처리 알고리즘을 제안한다. 제안하는 알고리즘은 질의 영역 내 K-anonymity를 보장함으로써 사용자의 위치 정보를 보호할 뿐만 아니라 불필요한 질의 결과 탐색을 줄여 네트워크 효율을 증가시킨다. 마지막으로, 성능평가를 통해 제안하는 기법이 기존 연구에 비해 질의 처리 시간 및 네트워크 효율 측면에서 우수함을 보인다.

키워드 : 위치 기반 서비스, 질의 처리 알고리즘, 사용자 정보보호, 네트워크 효율

Abstract Location-based services (LBS) are increasingly popular due to the improvement of geo-positioning capabilities and wireless communication technology. However, in order to enjoy LBS services, a user requesting a query must send his/her exact location to the LBS provider. Therefore, it is a key challenge to preserve user's privacy while providing LBS. To solve this problem, the existing method employs a 2PASS cloaking framework that not only hides the actual user location but also reduces bandwidth consumption. However, 2PASS does not fully guarantee the actual user privacy because it does not take the real user distribution into account. Hence, in this paper, we propose a nearest neighbor query processing algorithm that supports K-anonymity property based on the weighted adjacency graph(WAG). Our algorithm not only preserves the location of a user by guaranteeing k-anonymity in a query region, but also improves a bandwidth usage by reducing unnecessary search for a query result. We demonstrate from experimental results that our algorithm outperforms the existing one in terms of query processing time and bandwidth usage.

Keyword : Location Based Service(LBS), Query Processing Algorithm, User Location Preservation, Network Bandwidth Usage

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

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

1. 서 론

최근 위치 측정 장치의 다양화 및 무선 이동 통신의 발달로 인하여 사용자들의 위치를 바탕으로 서비스를 제공하는 위치 기반 서비스(LBS: Location Based Services)가 꾸준히 발전하고 있다. 위치 기반 서비스는 크게 스마트폰, PDA(Personal Digital Assistants) 등을 소지한 사용자와 위치 기반 서버로 구성된다. 먼저, 질의를 요청하는 사용자(질의 요청자)는 가장 근접한 POI(Point of Interest)탐색 또는 질의 영역 내 POI 탐색 등을 요청하기 위해 무선 통신 단말을 통해 자신의 위치와 원하는 질의를 서비스 제공자에게 전송한다. 위치 기반 서버는 사용자의 위치 정보를 기반으로 데이터베이스를 탐색하여 결과 POI를 탐색하는 질의 처리 알고리즘 수행하여 질의 요청자에게 결과를 반환한다. 그러나 현재 위치 기반 서비스를 이용하기 위해 사용자는 자신의 위치 정보를 수시로 위치 기반 서버에 전송해야 하며, 이러한 정보를 접근 권한이 없는 공격자(Adversary) 혹은 신뢰할 수 없는 서비스 제공자가 획득하는 경우 사용자의 연속적인 위치 추적 및 생활 패턴 노출 등의 개인 정보 유출 문제로 확대 될 수 있다. 이러한 정보가 유출되는 경우 공격자는 위치 기반 서비스 이용자들의 방문 빈도가 높은 장소 또는 주요 방문 시간대 등의 정보를 파악하여, 집 또는 회사 등의 주요 행선지를 분석하고 생활스타일, 질병 정보, 종교 등의 개인 정보를 획득할 수 있다. 따라서 최근 위치 기반 서비스 이용자들의 개인 정보 보호 요구가 크게 증가하고 있는 추세이다[11, 13].

현재까지 위치 기반 서비스에서 서비스 이용자의 위치 정보 및 개인 정보 보호를 위한 다수의 기법들이 연구되었다. 이러한 기법들은 질의 요청자의 정확한 위치 정보를 전송하지 않고도 신뢰할 수 있는 질의 결과를 얻기 위한 알고리즘이다. 대표적인 사용자 위치 정보 보호 기법 연구로는 클로킹(Cloaking) 영역 기반 질의 처리 연구가 있다[1-5, 8-10, 12]. 클로킹 영역 기반 질의 처리 알고리즘은, 먼저 질의를 수행하기 위해 사용자의 위치 보호를 위한 클로킹 영역을 생성한다. 이때, 클로킹 영역은 사용자의 위치 좌표를 특정 사용자 요구 사항인 Privacy profile (i.e., K-anonymity, L-diversity 등)을 만족하도록 확장한 질의 영역으로써, 위치 기반 서버에서 질의 수행 시 사용자의 위치 좌표 대신 클로킹 영역에 대한 질의를 수행하게 된다. 따라서 클로킹 영역 생성을 통해 질의 요청자는 자신의 정확한 위치 좌표를 전송하지 않고도 원하는 질의를 요청함으로써 자신의 개인 정보를 보호한다. 그림 1은 K-anonymity[7]를 적용한 Cloaking 영역 생성 예제이다. K=5일 때, 질의 요청자는 자신의 위치 q를 감추기 위해 K-1, 즉 4명의 다른 서비스 이용자를 포함하는 Cloaking 영역을 생성하였다. 따라서 질의 영역 내에 위치한 서비스 이용자의 정보(예: id 등)가 유출되더라도, 질의 요청자의 노출 확률을 최대 1/K로 감소시킨다(예제에서는 1/5로 감소되었다). 한편, 위치 기반 서버는 질의 영역 내에 사용자의 정확한 위치를 알 수 없으므로 정확도 높은 질의 결과를 반환하기 위해 전체 클로킹 영역에 대한 결과 후보 집합을 탐색하고, 사용자에게 전송한다. 따라서 클로킹 영역 기반 질의처리 알고리즘은 사용자가 원하는 정보 보호 수준이 높을수록 더 넓은 클로킹 영역을 생성하게 되고, 아울러 질의 결과 후보 집합의 수도 증가 하여 질의 처리 및 결과 전송 오버헤드가 증가하게 되는 단점을 지닌다.

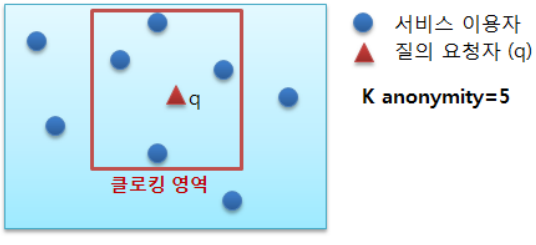


그림 1. 클로킹 영역 생성 기법

이러한 문제점을 해결하기 위해, H. Hu[6]는 보로노이 다이어그램(Voronoi diagram)을 이용한 POI 인덱스인 가중치 근접성 그래프를 정의하고, 이를 기반으로 질의를 수행하는 2PASS (2-Phase Asynchronous Secure Search) 최근접 질의 처리 알고리즘을 제안하였다. 2PASS 알고리즘은 사용자가 질의 요청 시 자신의 위치 정보 또는 클로킹 영역을 생성하는 대신, 자신이 설정한 정보 보호 수준인 가중치 임계값 τ (최소 결과 영역 크기)를 만족하도록 가중치 근접성 그래프를 탐색하여 최근접 질의 결과 POI 및 추가 POI를 직접 선택하여 질의를 수행한다. 따라서 자신의 위치를 노출하지 않고도 인덱스 탐색을 통한 질의를 수행함으로써 정보

보호를 수행하며, 결과 POI를 직접 선택함으로써 기존 연구에 비해 송/수신되는 POI수를 크게 감소시켜 네트워크 효율을 증가시킨 기법이다. 그러나 2PASS 알고리즘은 실제 사용자의 위치를 고려하지 않고 질의를 수행하기 때문에 최종 질의 결과 POI 영역이 최소 영역 크기 임계값인 α를 만족하더라도 실제 영역 내 사용자의 밀도가 매우 낮은 경우, 질의 요청자의 위치가 노출될 확률이 높다.

따라서 본 논문에서는 해당 연구를 확장하여 질의 결과 POI 영역 선택 시 실제 사용자의 수를 고려한 K-anonymous 가중치 근접성 그래프 기반의 최근접점 질의 처리 알고리즘을 제안한다. 제안하는 기법의 특징은 크게 두 가지로 정의할 수 있다. 첫째, 질의 요청자와 위치 기반 서버 사이에 제 3의 서버인 Anonymizer를 위치시켜 질의 결과 선택 시 사용자 정보를 함께 고려하도록 한다. 둘째, 기존 클로킹 기법의 장점인 K-anonymity를 보장함으로써 실제로 사용자 정보 보호를 수행하는 질의처리 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 클로킹 영역 기반 질의처리 알고리즘 및 2PASS 기법에 대해 설명한다. 아울러 제안하는 시스템의 구조 및 질의처리 알고리즘은 3장에서 자세히 다루도록 한다. 4장에서는 제안하는 기법과 기존연구와의 성능 평가 결과를 기술한다. 마지막으로 5장에서는 결론 및 향후연구에 대해 논의 한다.

2. 관련연구

최근 LBS에서 사용자 위치 정보 보호를 위해 다양한 연구가 수행되었다[1-5, 8-10, 12]. 대표적인 연구로는 사용자 위치 정보를 기반으로 클로킹 영역을 생성하고 질의 처리를 수행하는 클로킹 영역 기반 질의처리 알고리즘이 존재한다. 사용자는 질의 전송 시 자신의 위치정보가 다른 사용자와 구분되지 않을 만큼 최소한의 정보를 노출하면서, 효율적으로 질의를 수행할 수 있도록 한다.

대표적인 클로킹 영역 기반 질의처리 알고리즘은 M. Mokbel 그룹에서 제안한 New Casper[9] 가 있다. Mokbel 그룹은 그리드 기반 피라미드 구조(grid-based pyramid structure)를 사용하여 클로킹 영역을 생성하고, 이를 기반으로 수행되는 k-최근접점 탐색 질의처리 알고리즘을 제안하였다. 먼저,

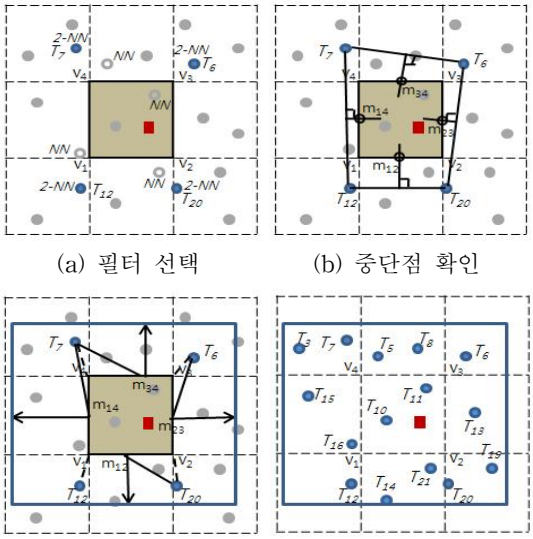


그림 2. New Casper 질의 처리 과정[9]

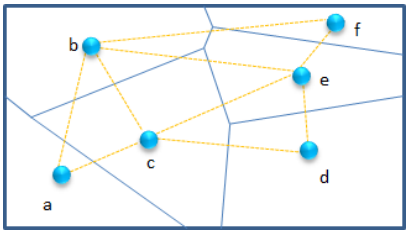
클로킹 영역 생성을 위해 전체 데이터 영역을 각 레벨마다 다른 크기를 가진 그리드 셀로 구성하여 피라미드 구조로 표현하고, 서버에서는 사용자가 원하는 privacy profile 을 만족하는 셀을 찾을 때까지 하위 레벨부터 상위 레벨까지 피라미드 구조를 탐색한다. 또한, 사용자가 클로킹 영역과 함께 질의를 요청하면, 위치 기반 서버는 이를 기반으로 최근접점 탐색 질의를 수행한다. 수행과정은 다음과 같다. 먼저 클로킹 영역의 각 꼭지점(vertex)을 기준으로 최근접점을 찾아 확장 기준점(target point) T로 정하고, 각 꼭지점의 확장 기준점 T를 연결하여 수직 이등분점을 클로킹 영역에 사상한다(그림 2a-b). 이때 사상한 수직 이등분점과 T 까지의 거리, 그리고 클로킹 영역의 에지를 이루는 두 개의 꼭지점에서 각각 T까지 거리 중에 가장 큰 값을 선택하여 클로킹 영역 생성 방향으로 확장한다(그림 2c-d). 이는 선택된 거리가 질의 영역의 최근접점이 될 수 있는 POI까지의 거리 중 가장 먼 거리이기 때문이다. 따라서, 해당 거리까지 영역을 확장하여 POI를 탐색할 경우, 질의 영역의 질의 결과로 선택가능한 모든 POI를 탐색 할 수 있다. 이러한 방법으로 확장된 영역 안의 POI들을 검색하여 최근접점 질의 결과 후보 집합을 사용자에게 반환한다.

그러나 클로킹 영역 기반 질의 처리 기법은 질의 처리 시 실제 사용자의 위치가 아닌 클로킹 영역에

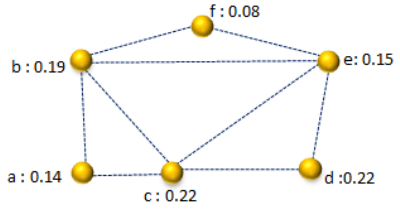
대해 탐색한 POI 후보 집합을 전송 받기 때문에 영역의 크기에 따라 결과 집합 후보의 크기가 증가하여 질의 처리 및 전송 오버헤드가 커진다. 특히, 사용자의 위치 정보 보호 요구 수준이 높을수록, 사용자의 밀도가 높을수록 질의 영역의 크기가 증가하여 질의 결과 후보 집합으로 탐색되는 POI의 양은 증가한다.

이러한 문제점을 해결하기 위해, H. Hu 그룹에서는 보로노이 다이어그램 (Voronoi diagram) 을 이용한 POI 인덱스인 가중치 근접성 그래프를 정의하고, 이를 기반으로 질의를 수행하는 2PASS (2-Phase Asynchronous Secure Search) 최근접 질의 처리 알고리즘을 제안하였다[13]. 먼저, POI 인덱스를 생성하기 위해 보로노이 다이어그램을 이용하여 영역을 분할하고, 각 보로노이 셀의 인접 셀 간의 연결성을 기반으로 각 셀의 크기에 따라 가중치를 부여하는 WAG(Weighted Adjacency Graph)를 생성한다. 그림 3a와 같이, 전체 데이터 영역은 보로노이 다이어그램을 이용하여 각 한 개의 POI를 포함하는 보로노이 셀로 분할되었다. 또한 그림 3b에서는 각 POI의 보로노이 셀 인접성을 연결한 WAG의 생성 결과를 나타낸다. 여기서 각 POI가 가지는 가중치는 해당 POI를 포함하는 보로노이 셀의 크기를 나타낸다.

최근접점 탐색 질의 처리는 크게 두 단계로 수행된다. 먼저, 사용자는 위치 기반 서버에 WAG 정보



(a) 보로노이 다이어그램



(b) 가중치 근접성 그래프(WAG)

그림 3. 2PASS 인덱스 구조[6]

를 요청한다. 그리고 자신이 속한 보로노이 셀 정보를 바탕으로 자신의 위치로부터 가장 가까운 POI를 탐색하고, 해당 POI의 보로노이 셀 가중치를 확인한다. 이때, 가중치 값이 자신이 정한 위치 정보 보호 수준인 τ를 만족하지 못하면, WAG를 탐색하여 인접한 POI 중 가중치가 높은 영역을 지니는 POI를 선택하여 확장을 수행하고, τ를 만족하면 종료한다. τ는 사용자 파라미터로써, 질의 결과 POI를 포함하는 최소 크기 영역이다. 둘째, 사용자는 최근접 POI 및 확장을 통해 탐색한 POI의 실제 위치 정보 및 부가 정보를 위치 기반 서버에 요청함으로써 질의 결과를 수신한다.

2PASS 질의 처리 기법은 자신의 위치나 클로킹 영역 정보를 전송하지 않음으로써 자신의 위치를 노출하지 않고도 인덱스 탐색을 통한 질의를 수행하여 사용자 위치 정보 보호를 수행하는 기법이다. 즉, 질의 영역을 전송하여 영역으로부터 최근접 POI를 탐색하는 기존 질의처리 알고리즘과 달리, 질의 요청자가 인접한 질의 결과 POI를 선택하여 요청함으로써 기존 기법에 비해 송/수신되는 POI 수를 감소시켜 네트워크 효율을 극대화 하였다. 그러나 이 기법은 실제 사용자의 위치를 고려하지 않기 때문에 최종 질의 결과 POI 영역이 최소 영역 크기 임계값인 τ를 만족하더라도 실제 영역 내 사용자의 밀도가 매우 낮은 경우, 질의 요청자의 위치가 노출될 확률이 높다. 예를 들어, 동일한 임계값을 적용하여 질의를 수행하는 경우, 도심지에서는 좁은 면적에서는 정보 보호를 보장하는 수의 사용자가 포함될 수 있지만, 사용자의 밀도가 상대적으로 낮은 농촌 등의 영역에서는 결과 영역에 질의 요청자만을 포함하거나 매우 적은 수의 사용자를 포함할 수 있다. 따라서 질의 영역에 포함된 사용자 중 실제 질의를 요청한 사용자가 노출 될 수 있는 문제점이 존재한다. 따라서 질의 결과 영역의 크기만이 아닌 실제 사용자 분포를 고려하여 질의를 수행하는 연구가 필요하다.

3. 가중치 근접성 그래프 기반 최근접 질의처리 알고리즘

3장에서는 기존 H. Hu 그룹의 연구인 2 PASS 최근접 질의 처리 알고리즘[6]을 확장하여 질의 결과 POI 영역 선택 시 실제 사용자의 수를 고려한

K-anonymous 가중치 근접성 그래프 기반의 최근접점 질의 처리 알고리즘을 제안한다.

제안하는 기법의 특징은 다음과 같다. 첫째, 질의 요청자와 위치 기반 서버 사이에 사용자 위치 정보를 관리하는 Anonymizer를 위치시켜 질의 결과 선택 시 사용자 정보를 함께 고려하도록 한다. Anonymizer는 일반적으로 모바일 서비스를 제공하는 사업자나 모바일 서비스 제공사의 협력 업체로 가정하며, 신뢰할 수 있는 서버로써 사용자의 모든 위치 정보를 관리한다. 이를 위해, 주기적으로(또는 필요에 따라) 사용자 위치를 업데이트하여 관리하고, 질의 요청 시 질의 시점의 질의 요청자 및 사용자의 위치를 이용하여 질의를 수행한다. 둘째, 질의 수행 시 결과 POI를 선택할 때 해당 POI 영역의 보로노이 셀 크기만이 아닌 실제 사용자의 수인 K-anonymity를 고려하여 가중치를 계산한다. 이를 위해, 기존의 가중치 근접성 그래프인 WAG에 K-anonymity 정보를 고려하는 K-WAG를 제안한다. 따라서 사용자 정보 보호 임계값을 만족하는 질의 결과 POI 집합을 선택할 때, 실제 사용자의 분포를 고려함으로써 정보 보호 수준을 높인다.

3.1 시스템 구조

제안하는 기법의 시스템 구조는 그림 4와 같다. 본 논문에서는 질의 요청자와 서버 사이에 Anonymizer를 위치하여 서비스를 이용하는 모든 사용자의 위치 데이터를 관리한다. 따라서 Anonymizer를 이용한 질의 처리 시나리오는 다음과 같다. 그림 4와 같이 사용자가 자신의 위치 정보와 함께 질의를 전송하면, Anonymizer는 사용자가 위치한 영역의 WAG를 위치 기반 서버에 요청한다. 다음으로 Anonymizer는 WAG를 통해 각 영역의 크기를 기반으로 서버에서 계산한 가중치를 확인하고, 각 보로노이 셀 내의 사용자 수를 이용하여 각 노드의 최종 가중치를 계산한다. 따라서 사용자가 위치한 최근접 POI셀부터 가중치를 더해가면서 사용자가 요청한 정보 보호 가중치 τ를 만족하는 모든 POI를 탐색하고, 이에 대한 정보를 위치 기반 서버에 요청한다. 그림 4의 질의 처리 시나리오에서 사용자가 “현재 위치에서 가장 가까운 레스토랑을 검색하여 메뉴정보를 탐색하라”는 질의를 요청하였을 때, Anonymizer는 앞에서 서술한 일련의 과정을 통하여 실제 사용자와 가장 가까이 위치한 맥도날드를 포함한 롯데리아, 피자헛을 최종 결과 POI로 탐색하고, 위치 기반 서버에 해당 레스토랑의 메뉴 정보를 요청함으로써 질의를 수행한다.

3.2 K-WAG 생성 및 제안하는 질의 처리 알고리즘

먼저 K-WAG를 생성하기 위해, 위치 기반 서버는 전체 영역을 POI 정보와 보로노이 다이어그램을 이용하여 보로노이 셀 영역으로 분할한다. 각 보로노이 셀 영역은 하나의 POI 정보를 저장하며, 셀 내의 임의의 모든 점에 대해 최근접 POI는 해당 보로노이 셀에 저장된 POI 이다. 따라서 사용자가 위치한 보로노이 셀을 탐색하는 것만으로도 최근접 POI를 쉽게 탐색할 수 있는 장점을 지닌다. 그러나 최근접 POI만을 탐색하여 질의 결과를 요청하는 경우 사용자의 위치가 해당 POI 영역 내에 포함된다는 것을 유추할 수 있으므로 사용자의 위치 정보가 노출될 수 있는 위험이 존재한다. 따라서 제안하는 기법에서는 사용자가 정의한 정보보호 임계값을 만족하도록 POI를 추가 탐색하여 질의 결과를 요청한

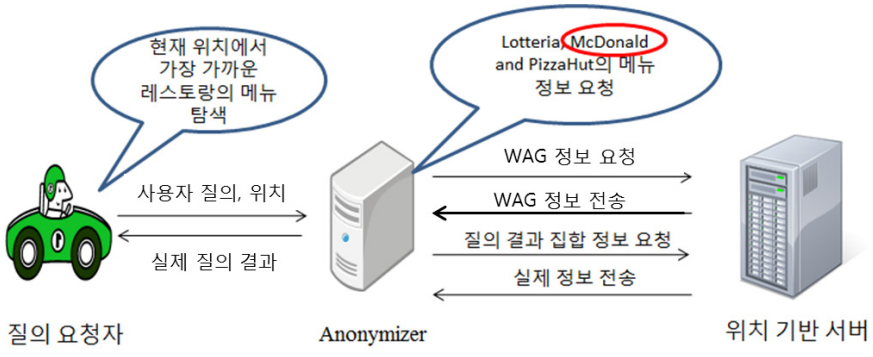


그림 4. 전체 시스템 구조 및 질의 처리 시나리오

다. 또한, 정보 보호 임계값 계산 시 보로노이 셀 크기 뿐만 아니라 각 보로노이 셀 영역에 위치한 사용자 수를 고려하여 가중치를 계산함으로써 실제 사용자의 분포를 고려한 질의 처리 알고리즘이다. K-WAG 생성 알고리즘은 그림 5와 같다.

K-WAG 생성 알고리즘
Input V_i : 보로노이 셀 i 의 크기 정보 U_i : 보로노이 셀 i 에 위치한 사용자 수
Output V_w(각 노드의 가중치)기반 K-WAG
1: compute area size for each voronoi cell (V_i) 2: calculate number of user on that cell (U_i) 3: give vertex weight (V_w) 4: set WAG (sum of the vertex weight) 5: end

그림 5. K-WAG 생성 알고리즘

K-WAG 생성 알고리즘은 Anonymizer가 위치 기반 서버로부터 사용자가 포함된 영역의 WAG 정보를 수신한 후 수행된다. 먼저, Anonymizer는 사용자가 속한 영역의 인접 보로노이 셀 크기 및 가중치(V_a)를 확인한다(line 1). 또한 각 보로노이 셀에 위치한 서비스 이용자의 수를 계산하고 식 (1)을 이용하여 최종 가중치를 계산 한다(line2-3). 식에서 α와 β는 보로노이 셀 크기와 사용자 수의 가중치 비율을 나타낸다. 사용자 수 가중치는 전체 사용자 수 ($Total_{Un}$)에서 질의 결과 영역에 포함된 사용자의 수(u_n)와의 비율을 이용하여 계산한다. 본 논문에서는 보로노이 셀 크기 가중치를 0.4, 영역 내 사용자 수 가중치를 0.6으로 가정하고 질의를 수행하는 알고리즘을 설명한다. 4장 성능평가에서 임계값 α와 β의 비율을 변경하여 실험을 수행함으로써, 두 임계값의 비율을 조정하는데 따른 질의 결과를 비교한다. 마지막으로 K-WAG 에 가중치 값을 저장함으로써 알고리즘을 종료한다(line 4-5).

$$v_w = (v_a \times \alpha) + \left(\frac{u_n}{Total u_n} \times \beta \right) \quad (1)$$

본 논문에서 제안하는 K-WAG를 이용한 최근접점 탐색 질의 처리 알고리즘은 그림 6과 같다. 설명의 편의를 위해 K-WAG 기반 질의 처리 알고리즘의 수행예제는 그림 7과 그림 8의 예제를 통해 설명한다.

먼저 사용자가 요청한 K-anonymity 값이 2이고, 정보 보호 임계값 τ가 0.5일 때, Anonymizer는 사용자가 위치한 영역의 WAG 정보를 저장하고 있는지 확인하고, 정보가 없을 경우 이를 서비스 제공자에게 요청한다(line1-2). 이 때, 서비스 제공자가 저장한 WAG 정보는 POI 위치를 기반으로 생성하여 각 보로노이 셀의 크기만을 이용한 가중치 값을 저장하고 있기 때문에, 실제 사용자 정보 보호 요구사항인 K-anonymity를 반영하지 못한다. 따라서

K-WAG 기반 최근접 탐색 알고리즘
Input q : 질의 요청자의 위치 k : 사용자가 요청한 K-anonymity 정보(최소 사용자 수) τ : 사용자 정보 보호 임계값
Output 질의 결과 POI 집합
1: if *iWAG* -tree is not available then 2: Obtain *iWAG*-tree from server 3: S = the snippet contains q 4: U_n = number of user on S 5: while area (S) < τ and U_n < k do 6: move one level up the tree and add snippets to S 7: request S' from the server 8: join the received S' into K-WAG 9: find objects from K-WAG 10: calculate approximate MVWCC (K-WAG, objects, τ, k) 11: request complete records of the objects 12: return objects and its attributes to the user

그림 6. K-WAG 기반 최근접점 탐색 알고리즘

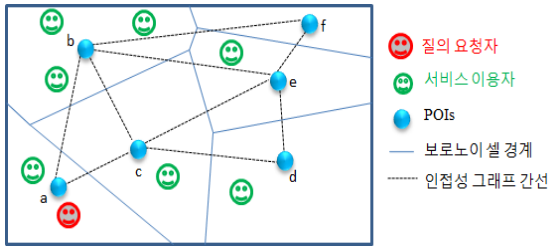


그림 7. 보로노이 셀 정보 및 사용자 분포

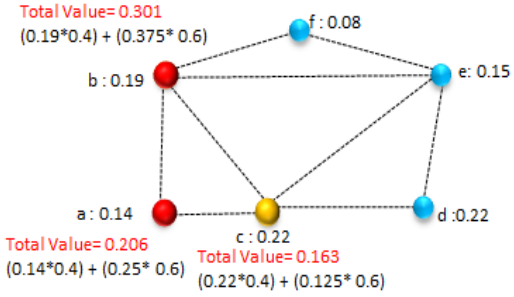


그림 8. K-WAG 가중치 계산

각 보로노이 영역에 위치한 사용자 수를 확인하고 K-anonymity를 고려한 질의 수행을 위해 앞에서 설명한 그림 5의 알고리즘을 통해 K-WAG를 생성 및 최종 가중치를 계산한다(line2-5). Anonymizer는 실제 질의 요청자의 위치를 알고 있기 때문에, 사용자가 속한보로노이 셀 영역을 탐색하고, 최근접 POI를 저장한다. 아울러, 해당 영역에서부터 사용자가 요청한 정보 보호 임계값 τ를 만족하기 위해 질의 결과 POI 집합에 포함하기 위한 추가 POI를 탐색하기 위한 확장을 수행한다. 그림 7과 같이 사용자가 위치한 보로노이 셀로부터 최근접 POI가 a임을 확인할 수 있다. POI a를 저장하는 보로노이 셀 영역의 크기 및 사용자 수를 고려한 K-WAG 가중치는0.206으로써 사용자 요구사항인 정보 보호 임계값 τ=0.5를 만족하지 못한다. 따라서 POI a로부터 인접한 b, c를 탐색하게 되고,WAG로부터 계산된 최종 K-WAG 가중치를 비교한다. POI b의 경우 해당 보로노이 셀의 영역만으로 계산된 가중치는 0.19였으나 영역 내 포함된 사용자의 비율을 고려하였을 때 0.301로 가중치가 재조정된다. 또한 POI c의 경우 위치 기반 서버로부터 전송받은 가중치는 0.22로 POI b보다 높은 값을 지니고 있었지만, 영역 내 포함된 사용자의 수가 적어 최종 가중치는 0.163으로 저장되었다. 따라서 Anonymizer에서는 b를 추가 POI로 저장하고, 가중치를 확인한다. 이 때, 두 POI가 지니는 가중치의 합은 0.507(0.206+0.301)로 임계값을 초과하였으므로 K-anonymity를 만족하는 지를 확인한다. POI a의 보로노이 셀 영역에 포함된 서비스 이용자의 수는 2, POI b의 보로노이 셀 영역에 포함된 서비스 이용자 수는 3이므로 K-anonymity=2를 만족한다. 따라서 Anonymizer는 질의 처리 알고리즘을 종료하고 POI a와 b를 최종 질의 결과 집합으로 반환한다.

4. 성능평가

본 장에서는 제안하는 K-anonymity를 고려한 가중치 근접성 그래프 기반의 최근접점 질의 처리 알고리즘의 우수성을 검증하기 위하여 기존 연구인 2PASS[6] 질의 처리 기법과 성능 평가를 수행한다. 성능평가를 위한 실험은 CPU Intel Xeon 2.00 GHz와 램 2GB, Windows XP 환경에서 Visual Studio 2010을 이용하여 수행되었다. 성능평가에 사용된 POI 데이터는 미국 북동부 지역의 실제 주소 데이터 집합[14]으로, 중복된 POI를 제외한 119898개의 POI를 사용하였다. 성능평가를 위한 매개변수는 표 1로 정리한다. 먼저 사용자의 수는 약 24만명으로 전체 데이터 영역에 랜덤하게 생성한 위치 값을 지닌다. 또한 사용자가 설정한 정보 보호 임계값은 질의 결과 영역의 크기와 사용자의 수를 비율적으로 합한 값으로 최소 0.000001에서 최대 0.001의 값을 지닌다. 기존 2PASS 알고리즘의 경우 정보 보호 임계값을 질의 결과 POI 영역의 크기만으로 나타낸다. 또한 질의 결과 영역 내 포함되는 사용자의 수인 K-anonymity는 2에서 10까지 5개의 데이터를 적용하였으며, 기존 연구에서는 이를 고려하지 않으므로 제안하는 기법에 대해서만 변수를 적용하여 성능평가를 수행하였다. 마지막으로 랜덤하게 분포된 전체 사용자 중 하나의 보로노이 셀 내에 포함된 평균 사용자의 수는 2이다.

표 1. 성능평가 변수

변수	범위	기본값
총 사용자 수	239,668	-
사용자 정보 보호 임계값(τ)	0.000001, 0.00001, 0.0001, 0.001	0.0001
K-Anonymity	2, 4, 6, 8, 10	6
보로노이 셀 평균 사용자수	2	-

성능평가를 위해 각 알고리즘의 최근접점 탐색 질의 처리 시간과 질의 결과 후보 셋의 수를 측정하여 비교 평가하였다. 질의 처리 시간의 경우, 알고리즘 수행시간을 측정하였으며, 실험의 결과 값은 10,000개의 질의를 수행한 평균값을 나타낸다.

그림 9 및 그림 10은 사용자가 요구하는 정보 보호 수준인 임계값 τ와 K-anonymity의 변화에 따른 질의 처리 수행 시간을 나타낸다. 제안하는 기법과 기존 기법 모두 τ와 K-anonymity가 증가할수록 질의 처리 시간 또한 증가한다. 이는 두 변수가 커질수록 사용자가 요구하는 정보보호 수준이 높아져 더 많은 후보 POI 및 서비스 이용자를 탐색하기 위해 가중치 기반 근접 그래프를 계속해서 탐색하고, 이에 따른 가중치 연산이 증가하기 때문이다. 특히, 기존 연구의 경우, 질의 결과를 포함하는 영역의 크기만을 정보 보호 수준으로 고려하기 때문에 임계값 설정 시 사용자 수까지 고려하는 제안하는 기법

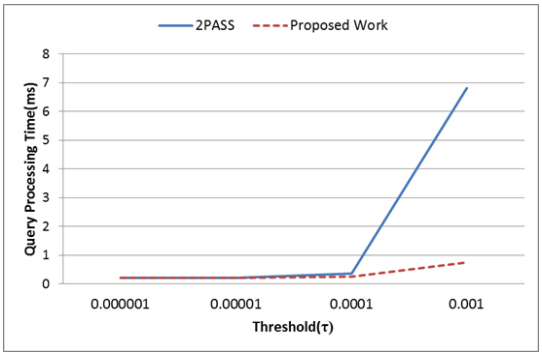


그림 9. τ변화에 따른 질의 처리 시간 (k-anonymity=6)

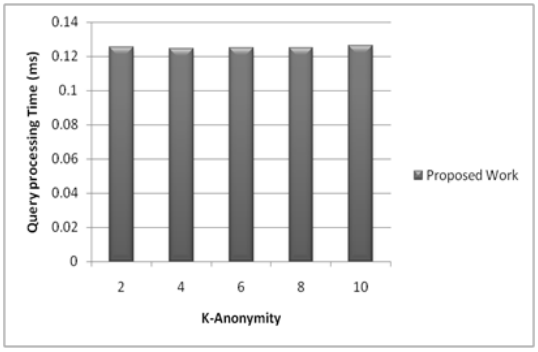


그림 10. K-anonymity변화에 따른 질의 처리 시간(τ=0.0001)

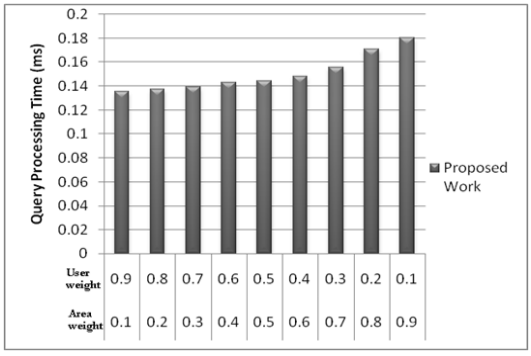


그림 11. τ가중치 변화에 따른 질의 처리 시간 (τ=0.0001)

보다 임계값을 만족하기 위해 더 많은 확장을 수행하므로 더 큰 질의 처리시간을 지닌다. 성능평가 결과, 실제 τ값이 증가 할수록 제안하는 기법의 질의 처리 시간은 약 0.12초, 0.125초, 0.15초, 0.7초로 증가 하였으며 기존 기법의 경우 각각 0.12초, 0.13초, 0.24초, 6초로 증가하였다. 또한 그림 10과 같이, 제안하는 기법의 τ가 0.0001일 때, K-anonymity 변화에 따른 질의 처리 시간은 K-anonymity 증가에 따라 약 0.01초 보다 작은 증가율을 나타냄으로써, 높은 K-anonymity 수준을 적용하더라도 질의 결과 영역 크기와 동시에 고려함으로써 질의 처리 시간에는 크게 영향을 끼치지 않음을 알 수 있다.

그림 11은 제안하는 기법에서 정보 보호 임계값 τ를 결정하는 변수인 사용자 수와 질의 결과 영역 크기의 비율을 다르게 수행했을 때 질의 처리 시간을 나타낸다. 임계값의 전체 가중치를 1이라 했을 때, 질의 결과 영역 내 사용자 수에 대한 가중치가 높을 수 록 질의 처리 시간이 감소하는 것을 보였다. 이는 정보 보호 임계값을 만족하기 위해 질의 결과 영역의 크기만이 아닌 실제 사용자의 분포를 고려함으로써 불필요한 확장을 감소시켜 나타나는 이점이다.

그림 12 및 그림 13은 사용자가 요구하는 정보 보호 수준인 임계값 τ와 K-anonymity의 변화에 따른 질의 결과 후보 집합의 크기를 나타낸다. 네트워크 효율을 증가시키기 위해서는 반환하는 POI 수를 감소시키는 것이 필요하다. 성능평가 결과, 제안하는 기법과 기존 기법 모두 τ와 K-anonymity가 증가할수록 사용자가 요구하는 정보 보호 수준이 높아져 더 많은 후보 POI를 탐색하게 된다. 그러나 제안하는 기법의 경우 질의 결과 POI 탐색 시 질의 결과 영역의 크기 뿐 아니라 사용자의 수를 고려하므로 기존 기법에 비해 작은 영역을 질의 결과로 선택하게 된다. 따라서, 결과 영역에 포함되는 POI 수가 기존 기법에 비해 크게 감소함을 알 수 있다. 성능평가 결과, 실제 τ값이 증가 할수록 제안하는 기법의 질의 결과 후보 집합의 크기는 약 80배 증가 하였으며 기존 기법의 경우 평균 후보 집합의 수가 약 200배 증가하였다. 이는 제안하는 기법이 동일한 임계값에 대해 탐색하는 POI 수를 크게 감소시킴으로써 사용자와 위치 기반 서버 간에 통신 오버헤드를 최대 2.5배 정도 감소시킴을 알 수 있다. 또한 제안하는 기법에서 K-anonymity 변화에 따른 질의 결과 후보 집합의 수는 K-anonymity 값을 15에서 60까지 증가시켰을 때 결과 값이 약 2배 증가함으로써 더 많은 사용자를 포함하여 사용자 정보 보호를 수행하는 것을 알 수 있다.

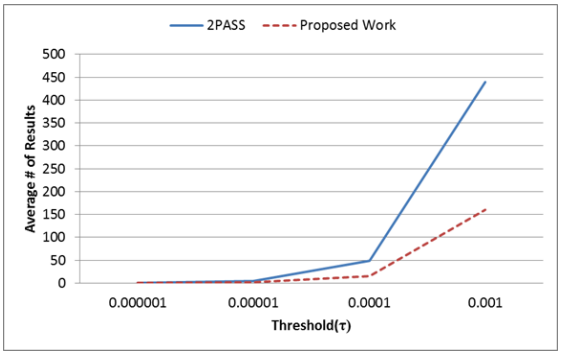


그림 12. τ변화에 따른 질의 결과 후보 집합의 수(k-anonymity=6)

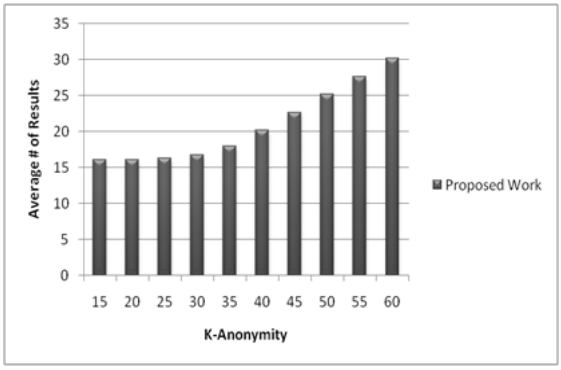


그림 13. K-anonymity 변화에 따른 질의 결과 후보 집합의 수(τ=0.0001)

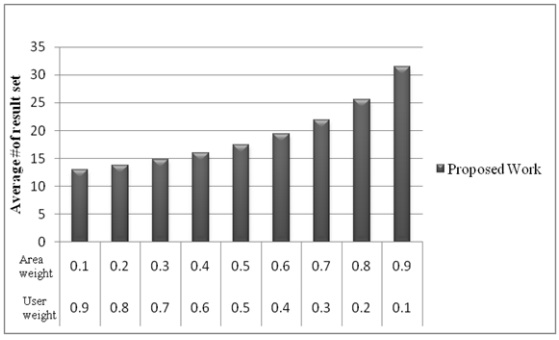


그림 14. τ가중치 변화에 따른 질의 결과 후보 집합의 수 (τ=0.0001)

마지막으로, 그림 14는 제안하는 기법에서 정보 보호 임계값 τ 를 결정하는 변수인 사용자 수와 질의 결과 영역 크기의 비율을 다르게 수행했을 때의 질의 결과 후보 집합의 크기를 나타낸다. 질의 처리 시간 결과와 동일하게, 질의 결과 영역 내 사용자 수에 대한 가중치가 높을수록 질의 결과 후보 집합의 수가 작아져 질의 결과 효율이 증가하는 것을 보였다. 이 또한 정보 보호 임계값을 만족하기 위해 질의 수행 시 실제 사용자의 분포를 고려함으로써 불필요한 확장을 감소시켜 나타나는 장점이다.

5. 결 론

최근 다양한 위치 기반 서비스가 꾸준히 발전되고 있다. 그러나 해당 시스템은 위치 기반 서비스 제공을 위해 사용자의 정확한 위치 정보를 필요로 하기 때문에 개인 정보 유출의 우려 또한 증가하고 있다. 따라서 본 논문에서는 가중치 기반 근접성 그래프를 이용한 K-anonymity를 보장하는 최근접 질의처리 알고리즘을 제안하였다. 제안하는 기법은 Anonymizer 서버를 이용하여 사용자 위치 정보를 관리한다. 아울러, 질의 수행을 위한 사용자 정보 보호 가중치 설정 시 질의 결과 영역의 크기와 질의 결과 영역 내 포함된 실제 사용자 수를 모두 고려함으로써 질의 결과 POI 탐색 수를 감소시켜 네트워크 효율을 증가시킨다. 성능평가를 통해 제안하는 기법이 기존 연구에 비해 질의 처리 시간 및 네트워크 효율 측면에서 우수함을 나타내었다.

향후 연구는 제안하는 기법을 k-최근접점 탐색 및 영역 질의를 수행하는 질의처리 알고리즘으로 확장하여 연구하는 것이다. 또한, Anonymizer의 사용에 따른 사용자 위치 정보 관리 비용에 대한 분석을 수행하는 것이다.

참 고 문 헌

[1] B. Bamba and L. Liu, 2008, “PRIVACYGRID: Supporting Anonymous Location Queries in Mobile Environments,” Proceeding of the 17th international conference on World Wide Web, pp. 237-246.

[2] C. Y. Chow, M. F. Mokbel, and X. Liu. A, 2006, “Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location-based Services,” In Proc. of the ACM International Symposium on Advances in Geographic Information Systems, pp. 171-178.

[3] G. Ghinita, P. Kalnis and S. Skiadopoulos, 2007,

"MobiHide: A Mobilea Peer-to-Peer System for Anonymous Location-Based Queries," In Proc. of International Symposium on Spatial and Temporal Databases, vol.4605/2007, pp. 221-238.

[4] G. Ghinita, P. Kalnis and S. Skiadopoulos, 2007, "PRIVE: Anonymous Location- Based Queries in Distributed Mobile Systems," In Proc. of World Wide Web, pp. 371-380.

[5] G. Ghinita, P. Kalnis, A. Khoshgozaran, C.Shahabi and K. L. Tan, 2008, "Private Queries in Location Based Services: Anonymizers are not Necessary," In Proc. of ACM SIGMOD international conference on Management of data, pp. 121-132.

[6] H. Hu and J. Xu, 2010, "2PASS: Bandwidth-Optimized Location Cloaking for Anonymous Location-Based Services." IEEE Transactions and Parallel on Distributed Systems, pp. 1458-1472.

[7] L. Sweeney, 2002, "k-anonymity: A model for protecting privacy." International Journal of Uncertainty, Fuzziness and Knowledge Based Systems 10(5), pp. 557 - 570.

[8] M. Gruteser and D. Grunwald, 2003, "Anonymous usage of Location-Based Services Through Spatial and Temporal Cloaking." In Proc. of the 1st international conference on Mobile systems, application and services, pp. 31-42.

[9] M. Mokbel, C. Chow, and W. Aref, 2006, "The New Casper: Query Processing for Location Services without Compromising Privacy," In Proc. of the International Conference on Very Large Data Bases, pp. 763-774.

[10] P. Kalnis, G. Ghinita, K. Mouratidis and D. Papadias, 2007, "Preventing Location-Based Identity Inference in Anonymous Spatial Queries," In Proc. of Transactions on Knowledge and Data Engineering, vol.19 no.12, pp. 1719-1733.

[11] Privacy concerns a major roadblock for location-based services say survey, 2007, http// www.Govtech.com/gt/article -s/104064, 2007.

[12] W. Ku, Y. Chen and R. Zimmermann, 2009, "Privacy Protected Spatial Query Processing for Advanced LBSs," Wireless Personal Communications, vol.51, no.1, pp. 53-65.

[13] W. R. Muntz, T. Barclay, J. Dozier, C. Faloutsos, A. Maceachren, J. Martin, C. Pancake and M. Satyanarayanan, 2003, "IT Roadmap to a Geospatial Future." The National Academics Press.

[14] Y. Theodoridis, 2003, "The R-tree portal, http:// www.rtreeportal.org," [Online]

[15] 김형일, 이아름, 장재우, 2010, "P2P 환경의 위치 기반 서비스에서 연속적인 질의 처리를 지원하는 그리드 기반 Cloaking 영역 설정 기법", 한국 GIS 학회지, 제 18권, 제 3호, pp 53-62.

[16] 장미영, 장재우, 2012, "도로 네트워크 환경에서 암호화된 공간데이터를 위한 k-최근접점 질의 처리 알고리즘", 한국공간정보학회지, 제 20권, 제3호, pp 67-81.

논문접수 : 2012.01.04
수 정 일 : 1차 2012.08.03 / 2차 2012.08.16
심사완료 : 2012.08.21

장 미 영
2009년 전북대학교 컴퓨터공학과 졸업 (학사)
2011년 전북대학교 컴퓨터공학과 졸업 (석사)
2011년~현재 전북대학교 컴퓨터공학과 박사과정
관심분야는 공간 데이터베이스, 정보 보호, 궤적 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,