프로필을 고려한 위치 기반 서비스 모델에서 사용자 식별 위협을 막는 익명화 기법

정승주, 박 석
서강대학교 컴퓨터공학과
pieces09@sogang.ac.kr, spark@sogang.ac.kr

An Anonymization Scheme Protecting User Identification Threat in Profile-based LBS Model

Seungjoo Chung, Seog Park
Department of Computer Science and Engineering, Sogang University

요 약

최근 무선 인터넷에서 사용자의 위치정보가 다양한 응용의 정보 요소로 활용되기 시작하였고, 이러한 응용의 하나로 위치기반 서비스(Location-Based Service: LBS)가 주목을 받고 있다. 그러나 위치기반 서비스에서는 서비스를 요청하는 사용자가 자신의 정확한 위치 정보를 데이터베이스 서버로 보내기 때문에 사용자의 개인 정보가 노출될 수 있는 취약성을 지니고 있다. 이에 모바일 사용자가 안전하고 편리하게 위치기반 서비스를 사용하기 위한 개인 정보 보호 방법이 요구되었다. 사용자의 위치 정보를 보호하기 위해 전통적인 데이터베이스에서의 개인정보 보호를 위해 사용되었던 K-anonymity의 개념이 적용되었고, 그에 따른 익명화를 수행할 수 있는 모델이 제시되었다. 하지만 기존 연구되었던 모델들은 오직 사용자의 정확한 위치 정보만을 민감한 속성으로 고려하여 익명화를 수행하였기 때문에, 이후 제시된 사용자의 프로필 정보를 고려한 모델에 대해서는 기존의 익명화만으로는 완전한 프라이버시를 보장할 수 없게 되어 추가적인 처리 과정을 필요로 하게 되었다.

본 연구는 프로필 정보를 고려한 위치기반 서비스 모델에서 Private-to-Public 질의가 주어지는 경우에 발생하는 추가적인 개인 식별의 위협에 관한 문제를 정의하고 이에 대한 해결책을 제시하며, 또한 제안 기법이 사용자 정보 보호를 보장하며 기존 방안보다 효율적임을 보인다.

1. 서 론

최근 무선 인터넷에서 사용자의 위치 정보가 다양한 응용의 정보 요소로 활용되기 시작하였고, 이 중 사용자의 위치 정보를 바탕으로 하여 서비스를 제공하는 위치기반 서비스(Location-based Service)가 주목을 받고 있다. 모바일 사용자가 교통 정보, 사람 찾기, 인접한 Point of Interest 찾기, 현재 위치의 날씨 정보 들의 서비스를 요청하면, LBS에서는 GPS와 같은 위치 식별 기기를 통해 사용자가 현재 위치한 장소가 어디인지를 파악하여 그에 기반한 효과적인 서비스를 제공하게 된다. 하지만 이런 서비스를 제공받기 위해서는 위치기반 서비스 제공자(Location-based Service Provider:LBSP)에게 자신의 실제 위치를 알려야만 하기 때문에 이로 인해 발생할 수 있는 개인 프라이버시 침해의 문제가 대두되었다. 공격자는 서비스 이용자가 어떤 곳에 있고 어느 장소를 언제 주로 방문하는지에 관한 정보를 수집하여 생활 패턴, 질병 정보 등의 개인 정보를 얻어낼 수 있다. 국외에서는 이런 정보를 활용한 스토킹 범죄 사례도 발생한 바 있다. 따라서 효과적으로 서비스를 제공하는 동시에 개인의 정확한 위치 정보를 악의적인 사용자로부터 보호하기 위한 기법이 연구되어 왔다.

기존 기법들 중 가장 활발히 연구된 것은 전통적인 데이터베이스에서의 정보 보호를 위해 사용되었던 k-anonymity[1]의 개념을 위치기반 서비스로 확장한 Location k-anonymity[2]이다. 이것은 질의를 요청할 때 사용자의 위치를 정확한 좌표가 아닌, 사용자와 그 외 k-1명 이상의 사용자가 포함된 영역을 만들어 전송하여 그 정보가 드러나더라도 질의 요청자가 식별될 확률을 1/k이하로 줄이겠다는 개념이다. 이 Location k-anonymity를 만족시키는 여러 가지

위치기반 서비스 모델이 연구되었고[3,4,5] 사용자의 프라이버시를 보호하기 위한 연구가 많이 이루어졌다. 하지만 더 나은 품질의 위치기반 서비스를 제공하기 위해 개인 사용자의 프로필 정보를 공개하는 시스템 모델이 제안되면서, 공개된 프로필 정보로 인해 발생할 수 있는 프라이버시 침해 위협이 새롭게 등장하였다. 따라서 기존의 위치기반 서비스에서 사용자의 위치 정보를 은폐하는 것에 더하여 프로필 정보에 관해서도 고려하는 기법에 대한 연구가 필요해졌다.

본 연구에서는 프로필 정보가 공개된 위치기반 서비스 모델에서 발생할 수 있는 프라이버시 침해 위협 중, Private-to-Public 질의가 사용자로부터 주어질 때 일어날 수 있는 질의 요청자 식별 문제에 대해 정의하고, 이를 막을 수 있는 기법에 대해 연구한다. 제안하는 기법은 프로필 정보의 효율적인 사용을 위해 정보 손실률을 줄이며 효율적으로 질의를 처리할 것이다.

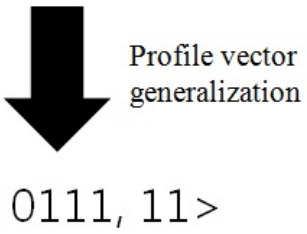
2. 연구 동기

위치기반 서비스에서 사용자 질의의 종류는 오브젝트와 서브젝트가 Private이나 Public인가에 따라 크게 세 가지로 구분될 수 있고, 질의의 종류에 따른 사용자 정보 보호 기법이 요구된다[3]. 따라서 새로운 모델이 제안되었을 경우, 질의 종류에 따른 사용자 정보 침해 위협을 고려해야 한다.

위치기반 서비스에서 더 나은 서비스의 제공을 위해 사용자의 프로필 정보를 공개함에 따라, 그로 인한 질의 요청자 식별 문제와 추론 공격에 대한 문제가 제기되었고, 그에 따른 새로운 익명화 모델에 대한 연구가 이루어졌다.

H. Shin et al.[4]은 Profile Anonymization Model을 제안하여, Location Server가 사용자의 프로필을 가지고 있고 Anonymous Group에 포함된 사용자들의 프로필 정보를 일반화 하는 모듈을 포함시켜 문제를 해결하려 했으나, 전체 프로필 정보를 일반화하는 방법을 사용했기 때문에 정보의 손실이 크게 나타날 수 있다는 단점이 있다. [그림 1]은 사용자의 프로필을 프로필 벡터로 표현한 것과, 그 프로필 벡터를 일반화 한 예를 보여주고 있다. 이런 식의 일반화 기법을 사용한 경우, 사용자 정보의 손실률이 늘어나고, 또한 서비스 제공자가 능동적인 서비스를 제공하는 데 어려움이 있을 수 있다는 단점이 있다.

Name	*Gender*	*Age*	*Salary*	*Profile Vector*
Doe	Male	35	$45,000	⟨10, 0010, 10⟩
Jane	Female	25	$85,000	⟨01, 0100, 01⟩
Robert	Male	42	$63,000	⟨10, 0001, 01⟩



<11, 0111, 11>

그림 1 Profile Anonymization Model에서 프로필 정보 일반화의 예

본 연구에서는 접근 방법을 달리하여, 이런 정보의 손실을 줄이면서 Private-to-Public 질의에서 나타나는 질의 요청자 식별 문제에 주목하여 이것을 해결하고자 했다.

사용자	성별	나이	직업	운전면허유무
A	남	43	자영업	유
B	남	23	대학생	무
C	남	28	회사원	무
D	남	16	중학생	무
E	여	36	전업주부	무

표 1 ASR에 속한 사용자들의 정보의 예

[표 1]은 프로필 정보가 공개된 ASR에 속한 사람들에 대한 정보이다. Location k-anonymity의 적용으로 인해 ASR에 속한 5명중 누가 질의자인지에 대한 확률이 1/5이하여야 하나, 질의에 포함된 정보가 사용자의 특정 속성과 연관된 경우 그 확률이 보다 커지게 된다. 예를 들어, 가장 가까운 주유소를 찾는 질의의 경우에는, 운전면허를 보유한 사용자가 A이기 때문에 A가 질의자일 확률이 높고, 여성 전용 의료기관에 대해 질의한 경우는 E가 질의자일 확률이 높게 나타나게 된다.

따라서 본 연구에서는 이런 특정 질의에 대한 사용자 식별 위협을 최소화하면서, 기존 기법에 비해 추가적으로 드는 비용을 최소화하며 사용자의 프로필 정보는 그대로 보존하는 기법에 대해 제안한다.

3. 제안 기법

그림 2 시스템 구조

[그림 1]은 본 연구에서 제안하는 시스템의 구조를 보여주고 있다. 사용자는 모바일 기기를 사용하여 자신의 정확한 위치와 질의 내용을 Location Anonymizer로 보낸다. Location Anonymizer에는 기존 연구들에서 제시되었던 익명화 기법을 사용하는 익명화 모듈과, 본 연구에서 새롭게 제시하는 질의 일반화 모듈이 포함된다. [그림 2]는 Location Anonymizer의 구조이다.

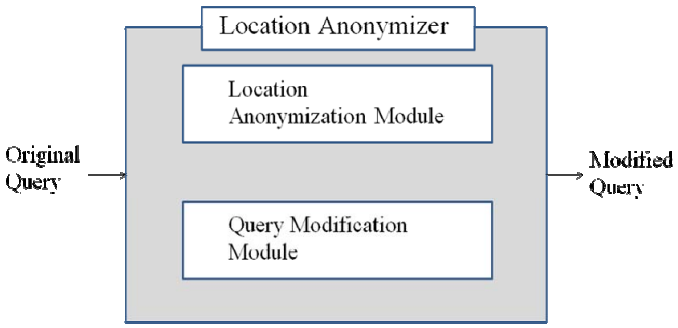


그림 3 Location Anonymizer의 구조

Location Anonymizer에서의 대략적인 정보 처리 과정은 다음과 같다. 먼저 위치 정보 익명화 모듈(Location Anonymization Module)에서 사용자가 포함된 Location K-anonymity를 만족하는 ASR을 만들고, 질의 일반화 모듈에서 질의 내용에 포함된 Object가 질의 요청자가 식별 되는 것을 유발하는 지를 확인한다. 이 때 개인 사용자의 식별 위협이 존재한다면, 질의에 포함된 Object를 일반화 하여 개인이 식별될 확률을 낮추게 된다.

Private-to-Public 질의에 의해 질의 요청자가 식별되지 않게 하기 위해서는, 질의의 Object와 그에 연관된 질의자의 프로필 속성에 대해, 해당 속성 값이 ASR에 포함된 사용자들 사이에서 유일해지지 않아야 한다. 따라서 이런 프라이버시 침해 문제를 해결할 수 있는 방법으로 두 가지를 제시할 수 있다. 첫째로, 위치 정보 익명화를 통해 ASR을 생성할 때부터 어떤 질의에 대해서도 해당 속성 값이 유일해지지 않도록 익명화를 수행하는 방법이 있을 수 있고, 둘째로는 질의의 Object를 보다 상위 카테고리의 Object로 수정하여, Object에 연관된 프로필 속성 값을 가진 사용자가 늘어나도록 하는 것이다.

이런 방법들을 통해 개인 식별의 위협이 없어졌다면, LA는 생성한 K-ASR과 수정된 질의 내용을 Location-based Service Provider(LBSP)에게 보낸다. LBSP는 보내진 정보를 이용하여 질의에 대한 답변의 후보군을 찾아내어 다시 LA에게 보내고, LA는 일반화 되기 전의 사용자의 질의를 바탕으로 후보군 중 최적의 답을 찾아내어 사용자에게 전송하는 것으로 프로세스가 종료된다.

다음 절에서는 위에서 소개한 두 가지 기법을 보다 자세히 설명하고, 문제를 해결하기 위해 어떻게 적용할 것인지를 다루도록 한다.

3.1 ASR 생성시의 조건 추가

프로필이 공개된 위치기반 서비스 모델에서 Private-to-Public 질의를 통한 질의 요청자의 식별 문제를 해결하기 위한 첫 번째 기법은, ASR을 생성할 때 질의 요청자의 모든 프로필 속성에 대해 유일한 값이 존재하지 않도록 하는 조건을 추가하여 K-ASR을 만드는 것이다.

[표 2]는 [표 1]의 프라이버시 침해 예제에서 A사용자가 질의를 한 경우, 초기 생성된 ASR 내에 다른 사용자를 추가로 포함한 ASR을 새로 생성하여 질의 요청자가 드러날 확률을 줄인 예이다.

사용자	성별	나이	직업	운전면허유무
A	남	43	자영업	유
B	남	23	대학생	무
C	남	28	회사원	무
D	남	16	중학생	무
E	여	32	전업주부	무
F	여	43	자영업	유
G	남	30	무직	유

표 2 수정한 ASR에 속한 사용자들의 예

[표 1]의 예제에서 운전 면허를 가진 사용자가 오직 A이기 때문에, 그에 관련된 질의가 주어졌을 경우에는 질의 요청자가 A일 확률이 높게 나타나게 된다. 하지만 사용자 F와 G를 추가로 포함하게 되면, 해당 속성을 가지고 있는 사용자가 늘어남으로 인해, 특정 질의의 경우에 A가 질의 요청자일 확률이 낮아진다.

3.2 질의 내용 변경을 통한 사용자 정보의 보호

본 논문에서 제안하는 두 번째 기법은, 질의 내용에 포함된 목표 오브젝트가 사용자 프로필에 있는 속성과 밀접한 연관이 있는 경우에 관련 오브젝트를 더 상위의 오브젝트로 수정하여, 질의 요청자가 가진 해당 속성이 유일하게 나타나지 않도록 한다. 이 때, 오브젝트의 수정은 [그림 4]와 같은 오브젝트 일반화 트리 구조를 이용한다.

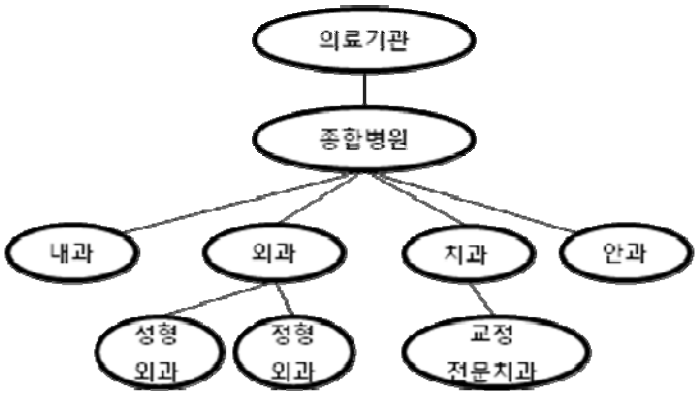


그림 4 오브젝트 일반화 트리의 일부

오브젝트 일반화 트리의 각 노드에는 통계 자료에 기반하여 지역 내 각 개체의 수를 예측한 값이 저장되어 있어, 이것을 통해 오브젝트가 일반화 되었을 때의 추가로 필터링 해야 하는 정보의 양을 예측하게 된다.

3.3 결합 기법

앞에서 설명한 두 기법에는 각각의 단점이 있다. 먼저 ASR 생성시의 조건을 추가하는 방법에서는 지역 내의 사용자의 분포가 편중되어있을 경우 ASR의 크기가 지나치게 커지는, 기존 ASR생성 기법과 같은 문제가 존재한다. 또한 질의 내의 오브젝트를 일반화하는 경우는 지나친 일반화로 인해, 서비스 제공자로부터 전송되는 결과값의 양이 너무 많아질 수 있다는 단점이 존재한다. 따라서 이 두 기법을 적절히 결합하여 서로의 단점을 보완하는 기법에 관하여 연구하였다.

[그림 5]는 결합 알고리즘으로, 먼저 질의를 일반화 한 뒤, ASR을 확장하여 생성하도록 한다. 오브젝트의 지나친 일반화를 방지하기 위해 Threshold값을 두고, 그 값을 Threshold 체크 함수를 통해 비교하여, 지나친 오브젝트의 일반화가 필요한 경우에는 ASR을 확장 생성하게 된다. Threshold 체크 시에는 오브젝트 일반화 트리에 부여했던 값을 활용하게 되고, ASR 확장 시에 ASR의 크기가 지나치게 커진 경우는 질의가 실패한 것으로 간주하게 된다.

Algorithm

```
1.  Area <- CreateTemporalCloakedRegion(user, k, l);
2.  if CheckUserAnonymity(area, user, query) then
3.      Label area as cloaked;
4.      Exit;
5.  end if
6.  while !CheckUserAnonymity(area, user, query) do
7.      ifCheckQueryGeneralizationThreshold(query,
    h_tree) then
8.          query <- GeneralizeQuery(query);
9.      end if
10.     if CheckUserAnonymity(area, user, query) then
11.         Label area as cloaked;
12.         break;
13.     end if
14.     Expand(area);
15.     if IsQueryFailed(area, Max_ASR) then
16.         Label area as failed;
17.         break;
18.     end if
19. end while
```

그림 5 결합 알고리즘

4. 실험

직접적으로 성능을 비교하기 적합한 대조군이 없기 때문에, 프로필이 공개된 모델에서 Private-to-Public질의가 주어졌을 경우 사용자 노출의 위협성을 고려하지 않은 기존 기법에 비해, 얼마만큼의 추가적인 비용이 드는지에 관하여 실험하였다.

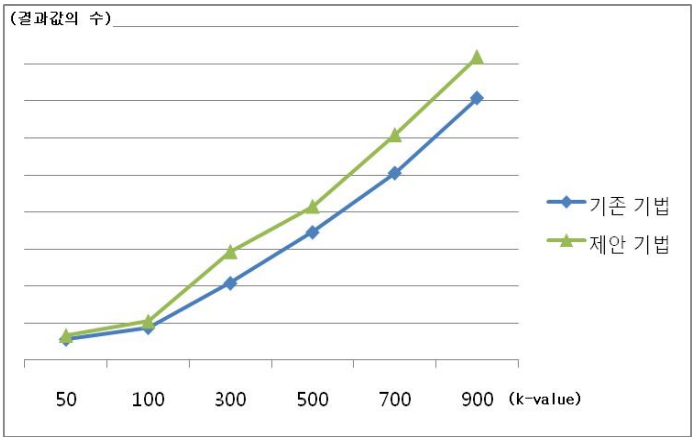


그림 6 질의에 대한 결과값의 수 비교

[그림 6]은 질의와 K값이 주어졌을 때, 얼마만큼의 결과값이 추가로 검색되어, 필터링 할 데이터의 양이 얼마나 늘어나는지에 대한 실험이다. 추가로 늘어나는 데이터 양을 백분율로 계산했을 경우, 약

10% 정도의 데이터를 Location Anonymizer 측에서 필터링 해야 하는 것을 확인할 수 있었다.
[그림 7]은 질의와 K값이 주어졌을 때, 사용자의 질의가 식별 위협의 가능성이 있는지를 판단하고 그에 따른 질의 일반화와 ASR 수정을 하는 시간에 대한 비교이다. 백분율로 15%정도의 추가 시간이 소요되었다.

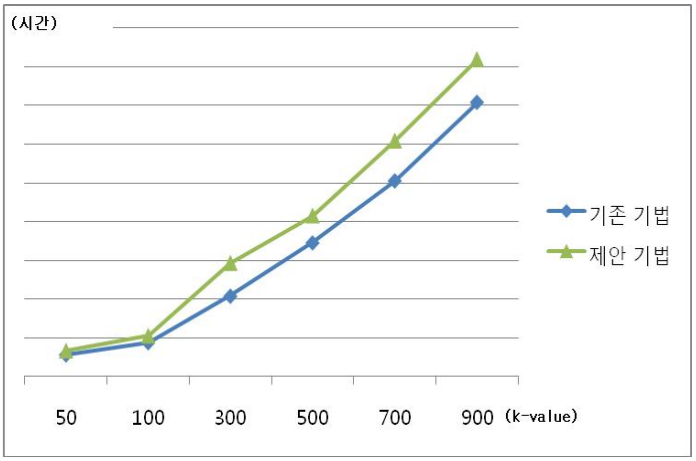


그림 7 총 질의 처리 시간 비교

두 실험의 결과에 따르면, 특정 질의에 대한 사용자의 프라이버시를 고려하지 않은 일반적인 위치 정보 익명화 기법에 비해 10%~15%정도의 추가적인 비용으로 사용자의 식별 위협을 막을 수 있는 것으로 측정되었다. 이 추가 비용은 사용자의 프라이버시를 보호하기 위한 추가 비용으로써 허용할만한 수준이라고 생각된다.

5. 결론 및 추후 연구

본 연구에서는 프로필을 고려한 위치기반 서비스 모델에서 특정 질의가 주어졌을 시 발생할 수 있는 사용자 식별 위협에 대한 문제를 제시하고, 그에 대한 해결책을 제안하였고, 실험을 통해 약간의 추가 비용을 부담하면 문제를 해결할 수 있음을 보였다.
추후 연구 계획으로는, 알고리즘을 개선하고 알고리즘에 적합한 위치 정보 익명화 기법을 사용하여 보다 빠르고 효율적인 처리를 가능하게 하려 한다. 또한 지금의 연구가 제한된 예제에 국한되어있기 때문에, 보다 큰 범위의 일반적인 데이터 예제와 질의 종류, 보다 정확한 통계 정보에 기반한 실험을 하려 한다.

참고문헌

[1] L. Sweeney, "K-anonymity: a model for protecting privacy", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10(5), pp. 557-570, 2002.

[2] M. Gruteser and D. Grunwald, “Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking,” Proceedings of the 1st international conference on Mobile systems, applications and services, pp. 31-42, 2003.

[3] M. Mokbel, C. Chow, and W. Aref, “The new Casper: query processing for location services without compromising privacy,” VLDB ’06, pp. 763-774, 2006.

[4] Heechang Shin, Vijayalakshmi Atluri, Jaideep Vaidya, "Profile Anonymization Model for Privacy in a Personalized Location Based Service Environment", 9th International Conference on Mobile Data Management,

[5] B. Gedik and L. Liu, “Location Privacy in Mobile Systems: A Personalized Anonymization Model,” Distributed Computing Systems, ICSCS ’05., pp. 620-629, 2005.

[6] M. Duckham and L. Kulik, “Location privacy and location-aware computing,” Book chapter in Dynamic & Mobile GIS: Investigating Change in Space and Time, pp. 35-51, 2006.