

도로 네트워크에서 연속적인 질의 처리를 지원하는 사용자 위치 정보 보호 기법†

나소라^o 유혜겸 장재우*

전북대학교 컴퓨터공학과, 영상정보신기술연구소*

{srna, hkyu, jwchang}@dblabb.chonbuk.ac.kr

Location Anonymization Algorithm Supporting Continuous Query Processing over Road Network

Sora Na^o Hye-Kyoum Yu Jae-Woo Chang*

Department of Computer Engineering, Chonbuk National University of Korea,

Center for Advanced Image and Information Technology*

1. 도로 네트워크에서 연속적인 질의 처리를 지원하는 사용자 위치 정보 보호 기법

최근 무선 통신 및 GPS 기술의 발달과 단말기의 보급으로 위치기반 서비스(Location-Based Services)의 이용이 확산되었다. 위치기반 서비스의 사용자는 GPS 기기로 측정된 자신의 위치정보를 서버에 보내어 가까운 주유소 검색, 교통 상황 검색 등의 각종 서비스를 이용할 수 있다. 그러나 위치기반 서비스(Location-Based Service)에서는 사용자의 정확한 위치 정보에 기반한 질의를 전송하기 때문에, 사용자의 개인 정보가 상대방에게 노출될 수 있다. 따라서 사용자가 안전하게 위치기반 서비스를 이용하기 위해서는 사용자 위치 정보 보호 방법이 요구된다.

도로 네트워크에서 사용자 위치 정보를 보호하기 위한 대표적인 연구로 T. Wang과 L. Liu가 제안한 XStar[1]가 존재한다. XStar는 도로 네트워크에서 노드(node)와 인접한 에지(edge)의 집합인 스타를 이용하여 사용자의 위치를 은닉한다. 아울러 cloaking 영역에 포함되는 사용자의 수(k) 뿐 아니라, 에지의 수(s)를 고려함으로써 보다 견고한 보호 수준을 지원한다. 그러나 XStar는 연속적인 질의 처리 시, 각 시간마다 생성되는 cloaking 영역에 포함된 사용자가 달라지기 때문에, 시간이 지남에 따라 사용자가 노출될 확률이 증가하는 문제점을 지닌다. 연속 질의 처리 시 사용자의 노출 문제를 해결하기 위한 연구로는 T. Xu등에 의해 제안된 Advanced KAA[2]가 존재한다. Advanced KAA는 cloaking 영역의 엔트로피를 통한 사용자 개인 정보 보호 수준을 정의하고, 이를 통해 연속 질의 처리 시 사용자 노출 확률을 감소시켰다. 그러나 이는 유클리디언 공간에서 정의되었기 때문에 도로 네트워크에 적합하지 않다는 단점이 존재한다.

본 논문에서는 XStar를 개선하여, 연속적인 질의 처리 수행 시에도 효율적으로 사용자의 위치 정보를 보호하는 기법을 제안한다. 이를 위해 엔트로피를 도로 네트워크에 적용하고, 엔트로피를 이용한 cloaking 영역 생성 알고리즘을 설계한다. 제안하는 정보 보호 기법은 스타 선택과 스타 병합의 두 단계로 구성된다.

단계 1. 스타 선택 단계

사용자의 위치를 사용자가 존재하는 스타로 대체함으로써 정확한 위치를 은닉한다. 이 때 효율적인 질의 처리를 지원하기 위해 스타의 에지 수 및 스타 엔트로피를 통해, 스타 병합 및 질의 처리에 필요한 비용인 스타 선택 비용을 정의한다. 스타 엔트로피는 이전 cloaking 영역에 존재하는 사용자(이후 참여자라 명명)가 특정 스타로 이동할 가능성을 의미한다. 스타 엔트로피 $H(S)$ 는 다음과 같이 정의된다.

정의 1. 스타 S에 포함된 에지가 $E_S = \{e_1, e_2, \dots, e_n\}$ 이며, 참여자가 $U_A = \{u_{A1}, u_{A2}, \dots, u_{Am}\}$, 참여자가 질의 요청 자일 확률이 $P_A = \{p_{A1}, p_{A2}, \dots, p_{Am}\}$ 이라 할 때, 이동 확률 행렬 M의 한 항 M_{ki} 및 참여자가 현재 스타 S로 이동할 확률 $P_{Es} = \{p_{E1}, p_{E2}, \dots, p_{En}\}$, 스타 엔트로피 $H(S)$ 는 각각 다음과 같다. 여기서 $oft(e_i, e_j)$ 는 전처리 과정을 통해 계산된 에지 e_i 에서 에지 e_j 로의 이동 횟수를 의미한다.

$$M_{ki} = oft(edge(u_{Ak}), e_i) / \sum_{i=1}^n oft(edge(u_{Ak}), e_i), \quad P_{Es} = P_A \cdot M, \quad H(S) = - \sum_{i=1}^n p_{Ei} \log p_{Ei}$$

스타 엔트로피가 높을수록 참여자가 해당 스타로 이동할 확률이 증가하므로 병합 비용이 감소하며, 스타의 에지가 많을수록 질의 처리 비용이 증가한다. 따라서 스타 선택 비용은 엔트로피와는 반비례 관계를, 에지의 수와는 비례관계를 지닌다. 따라서 스타 선택 비용은 다음과 같이 정의된다.

정의 2. 스타 S에 포함된 에지의 수를 NE(S), 스타 S의 엔트로피를 H(S)라 하면, 스타 S의 선택비용 $Cost(S)$ 는 다음과 같다. 여기서, WB와 WE는 각각 에지 수와 엔트로피의 가중치를 의미한다.

$$Cost(S) = NE(S) \times WB + WE / H(S)$$

† 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임 (No. 2010-0000202)

이를 통한 스타 선택 알고리즘은 다음과 같다. 첫째, 사용자의 좌표를 통해 사용자가 존재하는 에지가 포함된 두 스타를 검색한다. 둘째, 두 스타 중의 하나가 이미 선택된 경우, 해당 스타를 선택한다. 셋째, 둘 모두 선택되지 않은 경우, 두 스타의 선택 비용을 계산한다. 넷째, 스타 선택 비용을 적용하여 임의로 하나의 스타를 선택한다.

단계 2. 스타 병합 단계

이 단계에서는 다수의 스타를 병합하여 k-anonymity 및 s-diversity를 만족하는 슈퍼스타를 생성한다. 병합된 슈퍼스타의 적합성을 판단하기 위해 슈퍼스타의 엔트로피를 정의한다. 슈퍼스타의 엔트로피는 참여자가 질의 요청자 이면서, 현재의 슈퍼스타에 포함되었을 가능성으로 정의된다.

정의 3. 참여자가 질의 요청자일 확률이 $P_A = \{p_{A1}, p_{A2}, \dots, p_{Am}\}$ 이고, 이동 확률 행렬이 M일 때, 현재 슈퍼스타 SS에 포함된 사용자가 질의 요청자일 확률 $P = \{p_1, p_2, \dots, p_n\}$ 는 $P = P_A \cdot M$ 이며, 따라서 슈퍼스타의 엔트로피 $H(SS)$ 는 다음과 같다.

$$H(SS) = - \sum_{i=1}^n p_i \log p_i$$

슈퍼스타의 엔트로피가 높다는 것은 다수의 사용자가 질의 요청자일 가능성이 높은 것이므로, 사용자의 노출 확률이 감소함을 의미한다. 따라서 슈퍼스타의 엔트로피가 사용자 요구사항에 따른 엔트로피인 $\log k$ 이상의 값을 지니면, 해당 슈퍼스타는 cloaking 영역으로 적합하다. 제안하는 스타 병합 알고리즘은 다음과 같다. 첫째, 포함한 모든 질의에 대해 현재 슈퍼스타의 엔트로피를 계산하고, 현재의 슈퍼스타가 각 질의의 요구사항을 충족하는지 검사한다. 둘째, 요구사항을 충족하는 질의에 대해, 현재 슈퍼스타를 결과로 설정하고, 질의를 삭제한다. 셋째, 남은 질의를 처리하기 위해, 현재 슈퍼스타의 인접 활성스타를 검색하여 슈퍼스타를 확장한다. 넷째, 새로이 구성된 슈퍼스타의 hop수가 제한 hop수를 넘지 않으면 위의 과정을 반복한다.

2. 성능평가

본 논문에서는 제안하는 사용자 위치 정보 보호 기법을 Microsoft Visual Studio .net 2003으로 구현하여 Intel Core2 duo CPU E4500 2.2GHz와 RAM 2GB, Window XP 환경에서 성능평가를 수행하였다. 성능 평가에 사용된 도로 네트워크 데이터는 Sanfrancisco 도로 지도이며, BrinkHoff 알고리즘을 이용하여 최대 TimeStamp가 20인 210,000개의 사용자 궤적을 생성하였다. 각 TimeStamp마다 요청된 질의 수는 약 10,000~50,000개이다. 성능 비교 대상은 XStar이며, 성능평가 항목으로 k의 평균을 각각 3, 6, 9, 12로 변경하면서 cloaking 영역 설정 시간 및 성공 횟수를 측정하였다. k값의 분산은 1.5, s값의 평균은 6, 분산은 1.5, 제한 hop 수의 평균은 4, 분산은 1로 설정하였다.

그림 1은 k의 평균값 증가에 따른 TimeStamp 당 평균 cloaking 영역 설정 시간 및 평균 성공 횟수를 나타낸다. 첫째, cloaking 영역 생성 시간 측면에서는, 제안하는 기법이 스타 엔트로피의 적용을 통한 병합 시간의 감소로 인해 XStar보다 약 7% 정도의 성능 향상을 보인다. 둘째, 성공 횟수 측면에서는, 제안하는 기법이 사용자 개인 정보 보호 수준을 적용하여 보다 엄격한 cloaking 영역 생성 기준을 지님에도 불구하고 약 3% 정도의 성능 향상을 보인다. 이는 병합 비용을 계산하여 보다 적은 병합 비용을 지니는 스타를 선택하기 때문이다. 한편, 두 항목 모두 k값의 증가에 상관없이 거의 유사한 성능을 보인다. 이는 하나의 스타가 병합됨으로써 다수의 사용자가 cloaking 영역에 추가되어, 적은 계산량으로도 k-anonymity를 만족할 수 있기 때문이다. 성능 비교를 통해 제안하는 기법이 XStar 보다 엄격한 cloaking 영역 기준을 적용하였음에도, 모든 성능측면에서 XStar와 유사하거나 향상된 성능을 보임을 알 수 있다.



그림 1. cloaking 영역 설정 시간

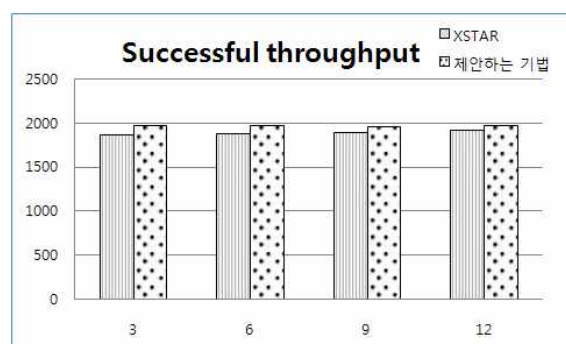


그림 2. cloaking 성공 횟수

참고문헌

[1] Ting Wang, Ling Liu, "Privacy-Aware Mobile Services over Road Networks", VLDB, 2009
 [2] Toby Xu and Ying Cai, "Location Anonymity in Continuous Location-based Services", ACMGIS, 2007, pp. 221-238.