

AMV: 클로킹 영역을 최소화하는 k-익명화 기법[☆]

AMV: A k-anonymization technique minimizing the cloaking region

송 두 희¹ 허 민 재¹ 심 중 원¹ 황 소 리¹ 송 문 배² 박 광 진^{1*}
Doohee Song Minjae Heo Jongwon Sim Sori Hwang Moonbae Song Kwangjin Park

요 약

본 논문에서, 우리는 이동 벡터(motion vector)를 이용하여 이동 중인 클라이언트의 k-익명화를 지원하는 AMV 기법을 제안한다. AMV는 이동 벡터 정보를 이용하여 사용자(클라이언트)들의 최소 클로킹 영역을 만들 수 있다. 클로킹 영역을 줄이는 주된 이유는 서버가 공간 질의를 요청한 모든 사용자에게 다수의 객체 정보(질의 결과)를 전송해야 하기 때문이다. 실험 결과를 통하여 기존 기법보다 AMV 기법의 성능이 우수함을 증명하였다.

☞ 주제어 : 클로킹, k-익명화, 위치기반 서비스, 공간 정보, 프라이버시

ABSTRACT

In this paper, we propose AMV scheme which supports k-anonymization by using vectors for mobile clients. AMV can produce the minimal cloaking area using motion vector information of users (clients). The main reason for minimizing cloaking area is a server has to send the object information to all users who request the spatial queries. The experimental results show that the proposed AMV has superior performance over existing methods.

☞ keyword : cloaking, k-anonymization, location based service, spatial data, privacy

1. 서 론

최근 GPS가 탑재된 모바일 기기들의 발전으로 위치기반 서비스의 이용이 확산되고 있다. 위치기반 서비스는 이동 통신망이나 위성항법장치 등을 통해 얻은 위치 정보를 바탕으로 이용자에게 여러 가지 공간 질의 서비스를 제공하는 시스템이다[1,2]. 위치기반 서비스를 활용한 예로는 네비게이션, 지도 서비스, 차량 추적, 주변 건물 정보 조회, POI(point of interest) 등이 있다[3,4].

위치기반 서비스는 사용자가 서버에게 자신의 실제 위치를 알려야 하기 때문에 사용자의 위치 정보 등이 노출될 수 있는 단점이 존재한다. 따라서 모바일 사용자의 개인 정보를 보호하기 위하여 클로킹(cloaking) 영역이나

k-익명화 기법 등이 제안되었다[5-9].

본 논문에서는 [4]에서 제안한 시스템 모델(질의 요청자 - 익명화(anonymizer) - 서버)를 적용하여 서비스를 이용하는 모든 사용자의 위치 데이터를 관리한다.

질의 요청자는 클로킹을 이용함으로써 서버에게 정확한 위치 정보를 제공하지 않으면서 신뢰할 수 있는 질의 결과를 얻을 수 있다. k-익명화는 질의를 요청한 사용자와 인접한 k-1개 이상의 사용자들을 서로 묶어서 클로킹 영역을 생성하는 기법이다. 이 기법의 장점은 제 3자에게 사용자의 정보가 노출되더라도 자신의 정보가 노출될 확률은 $1/k$ 로 줄일 수 있다는 점이다. 그러나 개인의 정보가 노출될 확률이 줄어드는 반면에 클로킹 영역이 설정된 상태에서 사용자가 공간 질의를 요청할 경우 서버는 질의자에게 클로킹 영역에 대한 다수의 객체 정보(질의 결과)를 전송할 가능성이 있는 객체 집합을 제공해야 하는 단점을 가진다. 반면 k-익명화의 k 값을 줄여서 클로킹 영역이 줄인다면 제 3자에게 사용자의 위치가 노출될 확률이 증가하게 된다. 우리는 이동 벡터를 이용하여 k-익명화를 만족시키면서 클로킹 영역을 줄일 수 있는 AMV (Anonymization of Minimum boundary rectangle using Vector)를 제안한다.

¹ Department of Information & Communication Engineering, Wonkwang University, Jeollabuk-do Iksan-si, 570-749, Korea.

² Mobile Communication Business, Samsung Electronics, Maetan-dong, Yeongtong-Gu, Suwon, Gyunggi-Do, Korea 443-742.

* Corresponding author (kjpark@wku.ac.kr)

[Received 22 July 2014, Reviewed 19 August 2014, Accepted 25 September 2014]

[☆] 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2013R1A1A1004593).

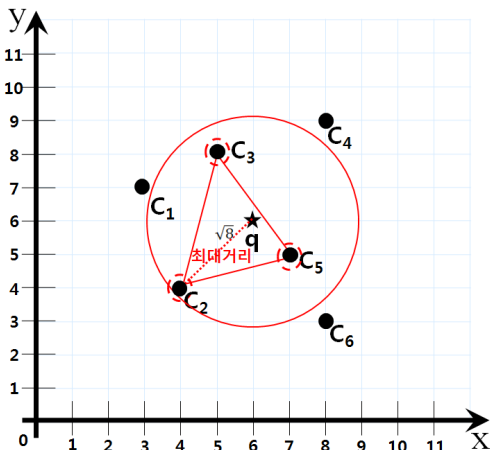
AMV의 주요 기여는 다음과 같다.

- k -익명화를 만족할 때 기존 기법의 클로킹 영역보다 AMV의 클로킹 영역이 작다. 따라서 서버가 AMV를 이용하여 객체 정보(질의 결과 될 가능성이 있는 객체 집합)를 질의자에게 제공할 경우 기존 기법보다 적은 객체 수를 수신하기 때문에 질의처리 시간을 줄일 수 있다.
- 익명자가 정확한 k -익명화 클로킹을 만족하기 위해서는 지속적인 갱신비용이 발생한다. 그러나 AMV를 이용하면 익명자는 t 에서 $t+n$ 까지 사용자의 위치를 확인하지 않아도 되기 때문에 익명자의 갱신 비용을 줄일 수 있다. 새

본 논문은 다음과 같다. 2장에서 기존의 클로킹(cloaking) 영역이나 k -익명화 기법에 대한 관련 연구를 설명하고, 3장에서는 AMV의 클로킹 영역을 줄이는 방법에 대하여 서술한다. 4장의 실험결과에서는 AMV와 기존 기법을 비교한다. 마지막으로 5장에서 결론을 맺는다.

2. 관련연구

그림 1은 4-익명화 기법에 대한 예제를 보여주고 있다. k 가 4를 만족시킬 수 있는 인접한 사용자의 최대거리는 $\sqrt{8}$ 임을 확인할 수 있다.



(그림 1) 클로킹 영역의 생성 예
(Figure 1) Creation example of cloaking range(4-anonymization)

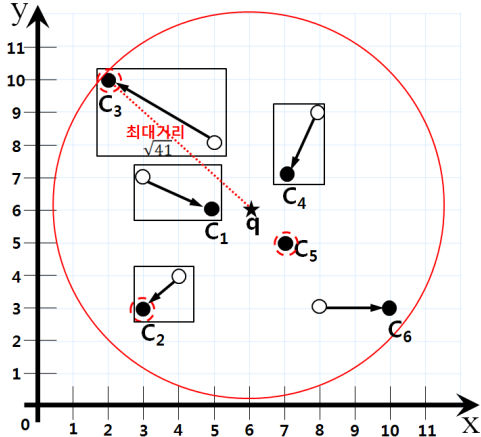
k -익명화 및 L-diversity를 고려한 그리드 기반 클로킹 기법에서는 L-diversity를 수행하여 L개의 건물을 탐색한 후 k -익명화를 수행하여 최소한의 넓이를 가지는 클로킹 영역을 생성한다[5]. L개의 건물을 탐색 후 임시 클로킹 영역을 설정하고 임시 클로킹 영역 안에 사용자의 수를 파악한 후에 질의 요청자가 원하는 사용자 수 보다 적으면 우선순위가 높은 순서대로 범위를 확장을 한다. M. Mokel가 제안한 New casper는 지역 익명화를 만족시키는 4가지 요소(정확성, 질, 효율성, 유연성)를 서술하였다[4]. [6]은 개인정보보호를 위해 Dummy를 이용한 클로킹 기법을 설계해 위치기반서비스에서 개인정보를 보호하였다. 기존의 Privacy Grid의 색인 구조에 건물의 그룹화 항목을 추가시켜 건물을 중복으로 카운트하던 문제를 해결하였고, 각 건물 모서리에 인접한 셀을 검색한 후 k 값을 증가시킴으로써 최소 클로킹 영역 설정으로 인해 개인정보보호 유출을 해결하였다. 그밖에 k -익명화를 이용한 정보 보호 연구들이 진행되었다[10]. 그러나 위에 서술된 연구들은 모두 이동성을 고려하지 않았다.

3. Anonymization of Minimum boundary rectangle using Vector(AMV)

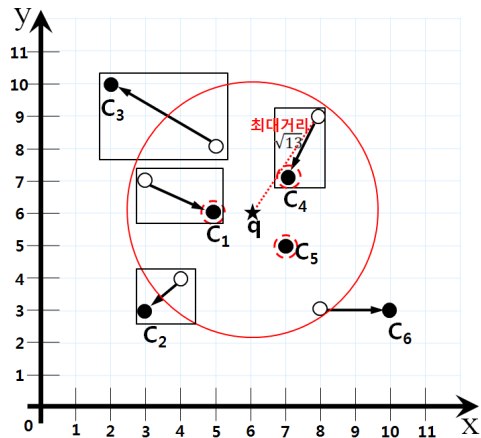
본 절에서 우리는 그림 1을 기반으로 그림 2와 그림 3에 대한 k -익명화 방법을 서술한다. 그림 1은 t 초일 때 q 를 포함하는 4-익명화 방식을 보여주고 있으며, 그림 2와 그림 3은 $t+n$ 초일 때 4-익명화를 구성한 그림이다. 그림 2는 기존 기법의 예제로서 질의를 요청한 사용자와 인접한 사용자가 k 를 만족하면, k 로 묶인 사용자들은 지속적으로 k -익명화를 유지한다고 가정한다. 그 이유는 인접한 사용자들이 변경될 경우 질의를 요청한 사용자의 정보가 노출될 확률이 증가하기 때문이다.

그림 2는 그림 1을 기반으로 구성된 클로킹 영역으로서 그림 1은 t 초 일 때 k -익명화를 만족시키는 최소 클로킹 영역이며(4-익명화: C_2, C_3, C_5, q), 그림 2는 $t+n$ 초일 때 C_2, C_3, C_5, q 를 유지한 상태에서 4-익명화를 만족시키는 클로킹 영역을 보여주고 있다. 즉, 붉은색 원으로 표시된 영역은 $t+n$ 초일 때 k -익명화를 만족하는 클로킹 영역이다. 그림 1은 t 초일 때 클로킹 영역을 나타내고 있으며, 클로킹의 면적은 25.12이다. 그림 2는 $t+n$ 초 일 때 클로킹의 면적은 128.74임을 확인할 수 있다. 이처럼, 기존 기법은 k 개의 사용자들을 유지해야 하기 때문에 클로킹 영역이

증가할 수 있다. 따라서 우리는 기존 문제점을 개선하기 위하여 그림 3에 대한 예제를 보여준다.



(그림 2) 4-익명화를 만족시키는 클로킹 영역 (기존 기법)
(Figure 2) Cloaking range satisfy 4-anonymization (original method)



(그림 3) 4-익명화를 만족시키는 클로킹 영역 (제안 기법)
(Figure 3) Cloaking range satisfy 4-anonymization (proposed method)

그림 3은 이동 벡터를 고려하여 $t+n$ 초일 때 4-익명화를 만족하는 클로킹 영역을 설정한 범위를 보여준다. AMV의 클로킹 면적은 40.82임을 확인할 수 있다. 이처럼 AMV기법은 k -익명화의 클로킹 영역을 기존 기법보다 효과적으로 줄일 수 있다.

기존의 연구들에 비하여 사용자의 속도와 방향의 유지

비용이 추가 발생하지만 사용자 입장에서 서버에게 위치와 익명화 k 의 정보를 제공할 때 속도와 방향 정보를 제공하면 되기 때문에 추가 발생하는 유지 비용은 미미하다.

4. 실험결과

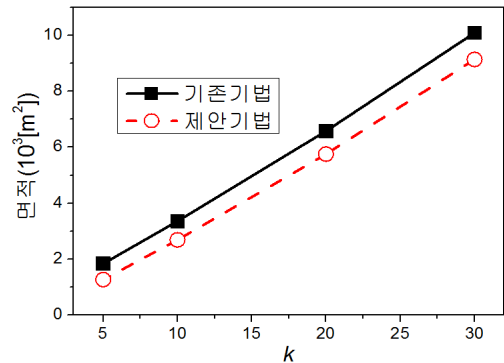
이 장에서는 기존 기법과 AMV 기법을 성능을 측정하고 이를 비교하였다. 모든 실험은 CPU 2.9Ghz의 프로세서와 4Gbyte의 메모리를 가진 컴퓨터에서 C++을 이용하여 실험을 수행하였다. 질의자는 고정된 상태에서 인접한 사용자들은 이동한다고 가정하였으며, 그리드 한칸의 거리는 1m로 설정하였다.

실험 환경에 대한 설정 값은 표 1과 같다.

(표 1) 실험 환경

(Table 1) Experimental environments

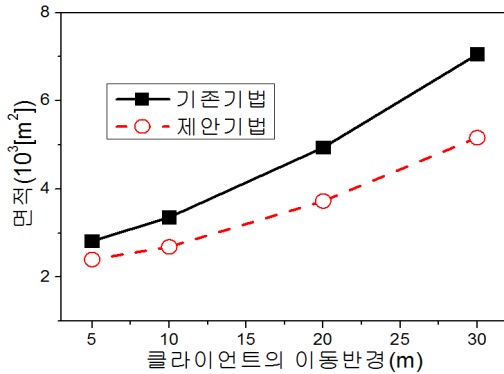
파라미터	설정 값
그리드 수	200*200
클라이언트 수	100
클라이언트의 이동반경	5, 10, 20, 30
k	5, 10, 20, 30
벡터의 시간(t)	1, 2, 3, 4
평균 실험 횟수	10,000



(그림 4) 클로킹 영역 비교
(Figure 4) Comparison of cloaking range(k)

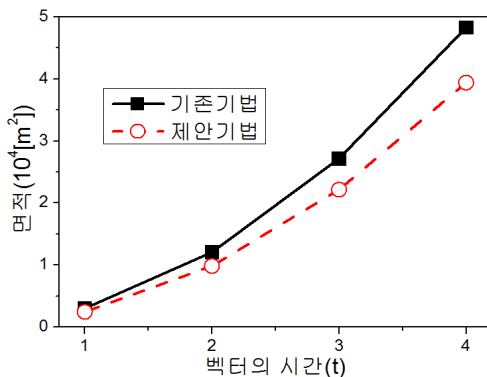
그림 4는 질의자가 요청한 k -익명화의 k 값에 따라 클로킹 영역이 변화하는 것을 보여준다. 실험환경은 클라이언트의 이동반경을 10m으로 고정하고 벡터의 시간을 1초 후라고 가정한다. 기존 기법과 AMV 기법을 비교한 결과 AMV 기법이 기존 기법보다 클로킹의 평균 면적이 13.7%

줄어든 것을 확인할 수 있었다. AMV가 기존기법보다 우수한 이유는 이동 벡터를 통해 얻어진 클라이언트의 이동 정보를 미리 계산할 수 있기 때문에 클로킹 면적을 줄일 수 있었다.



(그림 5) 클라이언트의 이동반경에 따른 클로킹 영역 비교
(Figure 5) Comparison of cloaking range (movement radius)

그림 5는 클라이언트가 이동할 수 있는 이동반경을 변수로 하여 기존기법과 AMV기법의 클로킹 영역을 비교한 것을 보여준다. 실험환경은 k -익명화의 k 값을 10으로 고정하고, 벡터의 시간을 1초 후라고 가정한다. 기존 기법과 AMV 기법을 비교한 결과 AMV 기법이 기존 기법보다 클로킹의 평균 면적이 23.1% 줄어든 것을 확인할 수 있었다. 클라이언트의 이동반경이 증가할수록 기존기법의 그래프 기울기에 비해 AMV 기법의 기울기가 낮은 것을 확인할 수 있었다.



(그림 6) 이동 벡터의 시간에 따른 클로킹 영역 비교
(Figure 6) Comparison of cloaking range(t)

그림 6은 이동 벡터의 속도와 방향 정보를 기준으로 t 초 이후에 목적지를 예측한 결과를 보여준다. 즉 t 초 이후에 클라이언트가 이동할 수 있는 클로킹 영역을 측정하는 결과이다. 실험환경은 k -익명화의 k 값을 10으로 고정하고, 클라이언트의 이동반경을 10m로 가정하였다. 기존 기법과 AMV 기법을 비교한 결과 AMV 기법이 기존 기법보다 클로킹의 평균 면적이 18.4% 줄어든 것을 확인할 수 있었다.

4. 결 론

본 논문에서는 질의자와 이동 중인 클라이언트의 k -익명화를 만족시키는 최소영역 클로킹에 대한 기법을 제안하였다. AMV 기법은 이동 벡터 정보를 통하여 클라이언트의 이동 거리를 추측할 수 있었다. 이를 통하여 기존 기법보다 클로킹 영역을 줄이면서 k -익명화를 만족시킬 수 있었다.

향후 연구는 사용자도 이동 객체로 설정한 후, 다양한 실험을 실시하여 AMV 기법의 우수함을 검증할 것이다.

참 고 문 헌 (Reference)

- [1] K. Park, "Efficient Data Access for Location-dependent Spatial Queries", Journal of Computer Science and Technology, Springer, vol. 29, no. 3, pp. 449-469, 2014.
- [2] D. Song and K. Park, "An Efficient Adaptive Bitmap-based Selective Tuning Scheme for Spatial Queries in Broadcast Environments", Journal of KSII Transactions on Internet and Information Systems, vol. 5, no. 10, pp. 1862-1878, 2011.
- [3] C.-G. Park, J.-J. Kim, K.-J. Han, "Efficient POI Search for Location-based Services in Road Networks", Journal of Computing Science and Engineering, vol. 40, no. 2, pp.141-151, 2013.
- [4] H. Sim, "LBS Description Element and Research Trend", Journal of Information Processing Systems, vol. 20. no. 6, pp. 3-12, 2013.
- [5] M. F. Mokbel, C.-Y. Chow and W. G. Aref, "The New Casper: Query Processing for Location Services without Compromising Privacy", Proceedings of the 32nd International Conference on Very Large Data Bases, pp. 763-774, 2006.

- [6] J. Kim, A. -R. Lee, Y. -K. Kim, J. -H. Um, J. -W. Chang, "Cloaking Method supporting K-anonymity and L-diversity for Privacy Protection in Location-Based Services", Journal of Korea Spatial Information Society, vol. 10, no. 4, pp.1-10, 2008.
- [7] J. -Y. Kim, E. -H. Jeong, B. -K. Lee, "A Design of Cloaking Region using Dummy for Privacy Information Protection on Location-Based Services", Journal of Communications and Networks, vol. 36, no. 8, pp. 929-938, 2011.
- [8] S. Park, J. Bai, S. Park, "An Anonymization Technique of Continuous Query and Query Log for Privacy in Location-Based Services", Journal of Computing Science and Engineering, vol. 38, no. 2, pp. 65-71, 2011.
- [9] J. Lee, "Hierarchical Clustering-Based Cloaking Algorithm for Location-Based Services", Journal of Korea Information and Communications Society, vol. 8, no. 8, pp. 1155-1160, 2013.
- [10] S. Park, S. J. Jeong, "An Anonymization Scheme Protecting Querist Identification in Location-based Service with User Profile", Journal of Computing Science and Engineering, vol. 38, no. 4, pp.201-207, 2011.

● 저 자 소개 ●



송 두 희 (Doohee Song)

2010년 원광대학교 정보통신공학과 학사
2012년 원광대학교 대학원 정보통신공학과 석사(공학석사)
2012년~현재 원광대학교 대학원 정보통신공학과 (박사수료)
관심분야 : 데이터베이스, 위치기반서비스, 공간질의처리
E-mail : songdoohee@navaer.com



허 민 재 (Minjae Heo)

2009년~현재 원광대학교 정보통신공학과 학사과정
관심분야 : 데이터베이스, 정보보호
E-mail : never5902@naver.com



심 종 원 (Jongwon Sim)

2009년~현재 원광대학교 정보통신공학과 학사과정
관심분야 : 데이터베이스, 위치기반서비스
E-mail : sjwmju@naver.com



황 소 리 (Sori Hwang)

2010년~현재 원광대학교 정보통신공학과 학사과정

관심분야 : 데이터베이스, 정보보호

E-mail : sori7378@navaer.com



송 문 배 (Moonbae Song)

2005년 고려대학교 대학원 컴퓨터학과 박사(이학박사)

2005년~2006년 고려대학교 연구조교수

2006년~2008년 Univ. of Tsukuba, JSPS Postdoctoral Fellow

2008년~2009년 성균관대학교 연구교수

2009년~ 현재 삼성전자 책임연구원

관심분야: 모바일 컴퓨팅, LBS, 서비스 디자인, 모바일 UI/UX, 서비스 플랫폼

E-mail : mbsong@gmail.com



박 광 진 (Kwangjin Park)

2000년 고려대학교 컴퓨터학과 학사

2002년 고려대학교 대학원 컴퓨터학과 석사(이학석사)

2006년 고려대학교 대학원 컴퓨터학과 박사(이학박사)

2006 - 2007 프랑스 국립컴퓨터과학연구소(INRIA) 박사 후 연구원

현재 원광대학교 정보통신공학과 교수

관심분야: 데이터베이스, 분산컴퓨팅

E-mail : kjpark@wku.ac.kr