

이동 컴퓨팅 환경에서 캐쉬 교체 알고리즘의 성능평가

정회원 김 호 숙*, 용 환 승**

Performance Evaluation of Cache Replacement Algorithm in Mobile Computing Environment

Ho-sook Kim*, Hwan-seung Yong** *Regular Members*

요 약

최근 저가의 무선 통신 기술의 발달과 고성능 이동 컴퓨팅 장비의 보급에 따라 이동 컴퓨팅 시장이 점차로 확대되면서 이동 컴퓨팅 환경의 제약성에 효율적으로 대처하기 위한 여러 방법이 제안되었고, 캐쉬 기법에 대한 많은 연구가 이루어지고 있다. 그 중에서 이동 호스트의 위치 변화에 따라 이동 호스트의 캐쉬 내에 저장된 데이터의 가치와 의미가 변경되는 것에 기반을 두고, 시간에 따른 사용자의 이동성과 데이터의 공간 속성을 효과적으로 지원하는 캐쉬 교체 방법으로 CR_LOC와 CR_REG가 제안되었다. 본 논문은 이동 컴퓨팅 환경 하에서 데이터의 공간 속성을 고려한 캐쉬 교체를 수행할 때 캐쉬 적중률에 영향을 미치는 다양한 요인들을 분석하고 이에 대한 성능평가를 수행한다. 이를 바탕으로 서로 다른 조건의 이동 호스트 각각에 대한 최적의 캐쉬 교체를 수행하기 위한 캐쉬 교체 선택 알고리즘을 제시한다.

ABSTRACT

The mobile computing service market is increasing rapidly due to the development of low-cost wireless network technology and the high-performance mobile computing devices. So it has been proposed the research to cope with the restrictions of characteristic of the mobile computing environments, especially much study is being done on the caching methods. Among them new cache replacement methods - CR_LOC and CR_REG efficiently support user mobility and spatial attributes of data. Those are based on the properties that the value and the semantics of the data, which are stored in the cache of a mobile host, change according to the movement of the mobile host. In this paper, we analyze a variety of factors which affect the cache hit-ratio of a mobile host that performs cache replacement considering the spatial properties of data in mobile environments. Moreover, we evaluate the performance of the cache replacement algorithm based on them. Finally, we propose a new cache replacement selection algorithm for the best cache replacement to each mobile host with a different condition.

1. 서론

무선 네트워크 기술의 발전과 이동 정보 기기의 발전은 컴퓨터 환경에 있어서 이동 컴퓨팅이라는 새로운 패러다임을 낳았다. 이동 컴퓨팅 환경(Mobile Computing Environment)은 사용자가 랩탑이나 노트북 등의 이동 단말기를 사용하여 무선 통신을 통해서 사용자의 물리적인 위치와 무관하게 서버 컴

퓨터의 자원과 함께 작업하는 환경을 의미한다^[1,2]. 최근 이동 장비의 소형화, 고성능화와 함께 무선 통신의 기반 구축에 힘입어, 작업 위치의 제약 없이 항상 최신의 정보를 얻고, 이를 통해 좀 더 높은 차원의 정보 서비스를 창출하고자하는 사용자들의 요구가 높아지면서 이동 컴퓨팅의 시장은 점차로 확대되고 있다^[3]. 이동 컴퓨팅 환경은 이동 지구국(Mobile Support Station)과 이동 호스트(Mobile

* 동의공업대학 컴퓨터 정보계열 전자계산과(hskim@dit.ac.kr)

** 이화여자대학교 컴퓨터학과 데이터베이스 연구실(hsyong@ewha.ac.kr)
논문번호 K01100-0308, 접수일자 : 2001년 3월 8일

Host 또는 Mobile Unit)로 구성된다. 이동 지구국은 고정 네트워크에 연결된 호스트들 중에서 이동 호스트에게 무선 통신 인터페이스를 지원하는 것으로서 데이터베이스 서버의 역할도 담당한다^[2]. 이동 컴퓨팅 환경에서는 무선 통신의 좁은 대역폭에 의해 전송 데이터 양의 제약을 받고, 전송 매체의 외부 간섭에 의한 잦은 단절이 발생할 수 있다. 이동 호스트는 시간에 따라 셀 간을 이동하면서 서로 다른 지구국과의 연결에 의해 트랜잭션을 지속하며, 전력 소모의 제한을 갖는다^[2]. 이러한 특징을 고려하여 지구국에서 데이터를 전송할 때 자주 요구되는 데이터를 방송(broadcast)을 통해 전달함으로써 이동 호스트가 데이터를 요구하는데 필요한 에너지 소모를 줄이고 사용자 수의 증가에 효율적으로 대처하는 방법이 제안되었다^[4]. 이와 함께 지구국에서 방송되는 데이터 중에서 향후 사용 가능성이 높은 데이터를 이동 호스트의 캐쉬에 저장하였다가 이후 사용자의 질의에 응답할 때 이용하는 기법이 도입되었다^[5]. 캐쉬 데이터의 사용은 질의 응답 지연 시간을 줄이고 지구국과 이동 호스트와의 데이터 전송량을 줄이며, 정상 상태 뿐 아니라 통신이 단절된 상태에서도 사용자의 질의에 대해 응답할 수 있는 장점을 갖는다^[5]. 이동 컴퓨팅 환경에서 동적으로 변화하는 지역의 지리 정보를 서비스하는 응용의 예처럼 데이터의 위치와 영역 특성을 고려하는 프로그램의 경우, 이동 호스트의 위치 변화에 따라서 이동 호스트의 캐쉬 내에 저장된 데이터의 가치와 의미가 변경되므로 데이터의 위치 특성과 영역 특성을 고려한 캐쉬교체 알고리즘이 요구된다^[6]. 즉 이동 호스트의 위치와 연관된 질의를 수행하는 경우, 데이터의 위치 특성과 영역 특성을 고려한 캐쉬 교체 알고리즘이 대표적인 캐쉬 교체 방법인 LRU 방법에 비해 높은 캐쉬 적중률을 보였으며, 이동 호스트가 지나가는 대상 지역의 데이터 밀집 정도에 따라서도 캐쉬 교체 방법들의 성능이 변화하였다^[6].

본 연구는 동적으로 변화하는 이동 컴퓨팅 환경 하에서 데이터의 위치와 영역 속성을 고려한 두 알고리즘 CR_LOC과 CR_REG의 성능에 영향을 미치는 요소들이 무엇인지 살펴보고, 실험을 통하여 각 요소들과 캐쉬 적중률간의 관계를 분석하여 이를 기반으로 서로 다른 특성을 갖는 사용자 각각에게 최적의 캐쉬 교체를 수행할 수 있는 캐쉬 교체 선택 알고리즘을 제시한다. 이러한 결과는 캐쉬의 적중률 향상과 이동 호스트와 이동 지구국 사이의 데이터 요청 빈도의 감소, 질의에 대한 응답 시간의

단축 효과를 기대할 수 있다.

본 논문의 구성은 다음과 같다. 제2장은 이동 호스트의 위치 변화와 데이터의 공간 속성을 고려한 캐쉬 교체 방법인 CR_LOC와 CR_REG에 대해 소개하고^[6], 제3장에서는 이동 호스트의 캐쉬 적중률에 영향을 미치는 요인을 분석하며, 4장에서는 성능 평가를 수행하여 이를 바탕으로 최적의 캐쉬 교체를 수행하기 위한 캐쉬 교체 선택 알고리즘을 제시한다. 마지막으로 결론과 함께 앞으로의 연구 방향에 대해 기술한다.

II. 데이터의 공간 속성을 이용한 캐쉬 교체 기법

알고리즘에 기술에 필요한 용어를 다음과 같이 정의한다.

[정의 1] 데이터 위치 DL_i (Data Location)는 데이터 D_i 의 공간상의 위치로 한 점의 좌표로 표시한다.

$$DL_i = (X_i, Y_i) \blacksquare.$$

각각의 데이터는 공간상의 위치와 더불어 해당 데이터가 의미를 갖는 논리적 범위를 갖는데 이러한 공간적 범위를 데이터 영역이라고 정의한다. 예를 들어 주식 값의 범위는 한 국가 전체이며, 날씨 정보는 한 지역에 국한되고, 교통 상황 정보는 좀더 좁은 범위에서 의미를 갖는다^[3,4]. 동일하게 셀 내의 다양한 데이터도 각각 영향을 미치는 공간적 범위를 갖는다. 예를 들면 1km마다 설치되어 있는 공중 전화는 1km²의 영역을 갖고, 10m마다 설치되어 있는 주유소는 100m²의 영역을 갖는다.

[정의 2] 데이터 영역 DR_i (Data Region)은 데이터 D_i 가 영향을 미치는 공간적 범위로 좌측 하단과 우측 상단 두 점의 좌표로 표시한다.

$$DR_i = (X_1, Y_1, X_2, Y_2) \blacksquare.$$

[정의 3] 이동 호스트(H_i)의 속성은 이동 호스트의 현재 위치를 나타내는 중심 좌표(X_h, Y_h)와 이동 호스트의 캐쉬 영역 (CR: Cache Region)으로 구성된다. 캐쉬 영역은 캐쉬가 처리하는 대상지역의 범위를 나타내는 버퍼 영역(Buffer zone)의 개념으로 이동 호스트의 현재 위치를 중심으로 한 반지름 CR_i 의 원 내부이다.

$$H_i = (X_h, Y_h, CR_i) \blacksquare.$$

[정의 4] 이동 호스트 H_i 의 캐쉬 크기(CS)는 이동 호스트에 따라 결정되는 고유한 값으로 캐쉬 내에 저장 가능한 데이터의 개수로 정의한다.

이때 데이터베이스 내의 모든 데이터 레코드의 크기(size)는 동일하다고 가정한다.

이동 호스트는 시간에 따라 위치를 변화하는 속성을 갖는데, “현재 위치에서 가장 가까운 지하철역의 위치는 어디인가?”와 같은 질의는 이동 호스트의 위치에 따라 질의 결과 값이 달라진다. 즉 이동 호스트의 위치에 따라 셀 내에 존재하는 지리 데이터가 갖는 의미와 중요도가 달라진다. 그러므로 사용자의 이동성과 데이터의 공간적 속성 (위치와 영역)을 고려한 새로운 캐쉬 교체 알고리즘이 필요하다.

[정의 5] CR_LOC(Cache Replacement by Location of data)는 이동 호스트의 위치와 데이터 위치 사이의 거리를 기준으로 캐쉬 교체를 수행하는 알고리즘이다. 이동 호스트의 캐쉬 교체 시 현재의 이동 호스트의 위치와 가장 거리가 먼 데이터가 캐쉬에서 삭제된다.

그림1은 이동 호스트의 캐쉬 크기가 4일 때 이동 호스트가 위치를 변화하면서 CR_LOC 알고리즘을 적용한 예이다.

[정의 6] CR_REG(Cache Replacement by Region of data)는 이동 호스트의 위치와 데이터 영역 사이의 거리를 기준으로 캐쉬 교체를 수행하는 알고리즘이다. 이동 호스트의 캐쉬 교체 시 현재의 이동 호스트의 위치로부터 데이터의 영역이 가장 멀리 위치한 데이터가 캐쉬에서 삭제된다.

그림 2는 이동 호스트의 캐쉬 크기가 4일 때 CR_REG를 적용할 때의 단계별 캐쉬 내용 변화를 나타낸다.

그림 2의 (C)에서 데이터의 위치로만 비교하면 1번 데이터가 2번 데이터보다 이동 호스트의 위치(7, 3)에 더 가깝지만 데이터의 영역이 큰 2번 데이터가 이동 호스트의 캐쉬에 오랫동안 남아있게 된다. 또한 5번 데이터의 영역은 (a)부터 (c)까지 지속적으로 캐쉬 영역 내에 남아 있는데 이는 데이터 영역의 크기가 작더라도 이동 호스트의 이동 경로 상에 존재하는 경우 계속 의미 있는 데이터로 남겨지는 것을 보여주는 예이다. 자세한 CR_LOC와 CR_

REG의 내용은 [6]에 기술되어 있다.

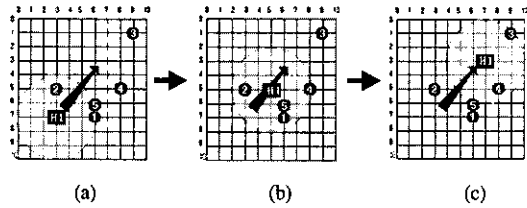


그림 1. CR_LOC의 적용 예

- H1: 이동호스트, ① ~ ⑤: 데이터, ● : 캐쉬영역(CR)
- (a) H_1 의 위치가 (3, 7)일 때 초기 캐쉬에는 데이터 1, 2가 저장되어 있다.
 - (b) H_1 의 위치가 (5, 5)일 때 캐쉬에는 데이터 1, 2, 4, 5가 저장된다.
 - (c) H_1 의 위치가 (7, 3)일 때 캐쉬 크기가 4이므로 2번 데이터가 삭제되고 3번 데이터가 삽입되어 캐쉬에는 데이터 1, 3, 4, 5가 저장된다.

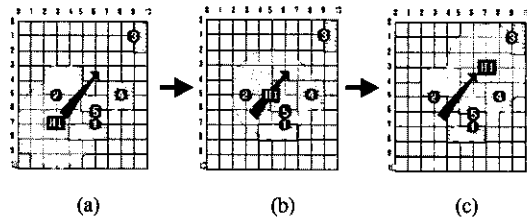


그림 2. CR_REG의 적용 예 :

▨ : 데이터 영역

- (a) H_1 의 위치가 (3, 7)일 때 초기 캐쉬에 저장된 데이터는 1, 2, 5이다.
- (b) H_1 의 위치가 (5, 5)일 때 캐쉬에는 데이터 1, 2, 4, 5이 저장된다.
- (c) H_1 의 위치가 (7, 3)일 때 캐쉬 크기가 4이므로 1번 데이터가 삭제되고 3번 데이터가 삽입되어 캐쉬에 저장된 데이터는 2, 3, 4, 5이다.

III. 이동 호스트의 캐쉬 적중률에 영향을 미치는 요인 분석

동적으로 변화하는 이동 컴퓨팅 환경하에서 이동 호스트의 캐쉬 적중률은 이동 호스트 자체의 성능과 사용자의 질의 특성 그리고 이동 호스트가 위치하는 지역의 특성 등 여러 가지 요인에 의해 영향을 받는다. 그러므로 최적의 캐쉬 교체를 수행하기 위해 이동 호스트의 캐쉬 적중률에 영향을 미치는 요인들을 다각도로 분석하여 이를 성능평가의 기준으로 사용하고자 한다.

이를 위해 다음의 몇 가지를 정의한다. 먼저 이동 호스트가 특정 서버가 관리하는 대상 지역을 이동할 때 최적의 캐쉬 교체 수행하기 위하여 자신의 캐쉬 크기에 적절한 캐쉬 영역 (CR)을 다음과 같이 결정한다.

$$CR = \sqrt{\frac{S_{area} \times CS}{S_{count} \times H}} \quad \blacksquare$$

이때 S_{area} 는 서버에서 관리하는 지역의 넓이고, S_{count} 는 서버의 데이터베이스 내에 포함된 데이터의 개수로 이동 호스트의 캐쉬 크기(CS)와 같이 미리 정해진 값이다.

그림3은 서버에서 담당하는 10 x 10 = 100 단위 (unit)크기의 지역 내에 20개의 데이터(점)가 분포하는 예를 표현한 것이다. 이때 (5,5) 위치에 있는 이동 호스트의 캐쉬 크기를 5 라고 가정하고 캐쉬 크기에 적당한 캐쉬 영역을 구해보자. S_{area} 는 100이고, S_{count} 는 20 이며, 캐쉬 크기가 5 이므로 CR은 약 2.82 이다. 즉 그림3에서 캐쉬가 처리하는 적절한 대상 지역의 범위는 캐쉬 영역(CR)을 반지름으로 하는 안쪽 원 내부를 가리킨다.

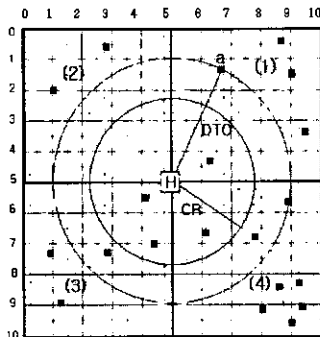


그림 3. 서버의 데이터 분포

[정의 7] 검색 대상 데이터의 분산도 (DTO: Distribution degree of query Target Objects)는 한 위치에서 이동 호스트가 수행한 질의들에 대한 검색 결과의 대상 데이터가 지리적으로 분산된 정도를 나타내는 것으로 질의의 검색 대상 데이터 중에서 현재 이동 호스트의 위치로부터 가장 멀리에 위치하는 데이터까지의 거리로 나타낸다.■

그림 3의 예에서 현재 이동 호스트 위치에서 요구한 데이터 중 가장 먼 거리에 위치한 데이터가 a 라 하면 모든 검색 대상 데이터들은 바깥쪽 원 내

부에 존재한다고 볼 수 있고 이때 DTO는 4이다.

[정의 8] 지역 밀집도 (Density of Region)는 서버에서 관리하는 전 지역의 데이터 밀집 정도에 대한 현재 이동 호스트가 위치하는 대상 지역의 데이터 밀집 정도를 나타내는 것으로 최소지역, 평균 지역, 밀집지역의 3단계로 구분한다.■

그림 3의 예에서 서버가 관리하는 지역을 4개의 사분면으로 나누어 지역 밀집도를 살펴보면, 1사분면과 3사분면은 평균 밀집도를 갖고, 제2사분면은 최소지역으로 구분할 수 있으며, 제 4사분면은 밀집 지역으로 구분 할 수 있다.

[정의 9] 사용자의 질의 유형(Query Pattern)은 이동 호스트의 사용자가 수행하는 질의가 데이터의 어떤 속성을 중심으로 수행되는가에 따라서 구분되며, 데이터의 위치 속성 중심 질의와 데이터의 영역 속성 중심 질의 2가지로 나뉜다.■

1. 이동 호스트의 캐쉬 크기

이동 호스트의 캐쉬 크기는 이동 호스트의 성능에 따라 결정되는 고유한 값이다. 현재 이동 장비들의 메모리 크기가 수백 바이트(byte)부터 수 기가 바이트(GB)까지 다양하고, 점차적으로 그 크기가 커질 것으로 기대되지만 서버의 데이터베이스 크기에 비해 이동 단말기의 메모리 크기는 매우 작으므로 이동 단말기의 캐쉬 교체 정책의 중요성은 점차로 증대되고 있으며 이때 사용자의 성능을 고려한 캐쉬 관리를 위해 캐쉬 크기는 반드시 고려되어야 하는 요인이다. CR_LOC와 CR_REG 방식의 캐쉬 교체 시 캐쉬 크기의 증가에 따라 캐쉬 적중률이 증가하지만 그 기울기는 점차 줄어드는 결과를 보인다⁶⁾.

2. 검색 대상 데이터의 분산도 (DTO)

응용 프로그램의 종류나 사용자의 질의 성향 또는 지역에 따라 검색 대상 데이터의 분산도는 달라질 수 있다. 예를 들어 특정 지역의 지리 정보를 알려주는 응용 프로그램을 도보 여행자를 위한 디바이스에 장착하는 경우에 비해 이동하는 운전자에게 서비스하기 위해 차량에 장착하는 경우에는 서비스 되는 데이터의 지리적 분산 정도가 커지게 된다. 또한 캐쉬 교체 시 DTO의 결정은 사용자마다 고유한 값을 적용할 수도 있고, 서버에서 일정 값을 지정하여 해당 지역을 통과하는 모든 사용자에게 동일하게 적용할 수도 있는데, 전자의 경우는 각각의 이동

호스트가 지금까지의 질의에 의한 DTO의 평균값에 대한 분석을 통해서 고유한 값을 사용하는 경우로 사용자의 질의 형태가 이동 호스트의 위치와 근접한 데이터를 주로 접근하는 경우에는 DTO값이 작고, 이와 반대로 질의의 형태가 이동 호스트의 위치와 무관하거나 멀리 위치한 데이터까지 접근하는 경우에는 DTO값이 클 것이다. 후자의 경우는 서버에서 해당 지역에서 발생하는 질의들을 분석하여 DTO의 평균값을 계산하고 이를 서버의 영역 내의 모든 이동 호스트에게 적용하는 경우로 서버에서 관리하는 데이터의 특성에 따라 특정 형태의 질의가 많이 발생하는 환경에서 적용 가능한 방법이다.

3. 지역의 밀집도

서버에서 관리하는 데이터는 지역의 특색에 따라 서로 다른 밀집도를 갖는다. 예를 들어 한 도시 내의 상점들에 대한 지리 정보를 갖는 데이터베이스인 경우 상점이 밀집되어 있는 도심 한가운데에 비해 공원이나 번두라 지역의 데이터는 희소한 분포를 갖게 된다. 즉 같은 넓이의 지역이라 할지라도 지역의 특성에 따라 대상 데이터의 수가 달라지므로 캐쉬 교체의 적중률을 높이기 위해서는 이동 호스트가 지나가는 대상 지역의 데이터 밀집도를 고려해야 한다.

4. 사용자의 질의 유형

캐쉬 교체 정책에서 가장 우선적으로 고려해야 하는 것이 사용자의 질의 유형이며 응용 프로그램의 특징에 따라 다르게 나타난다. 접근되는 데이터가 시간적으로 가까운 것을 많이 접근하는 응용 프로그램의 질의는 데이터의 시간속성을 중심으로 처리되는 질의이다. 이에 반해서 이동 컴퓨팅 환경 하에서 움직이는 사용자가 “현재 위치에서 A 지점까지의 거리는 얼마인가?” 와 같이 대상 데이터의 위치(Location) 속성을 기준으로 하는 질의를 주로 많이 사용하는 경우는 데이터의 위치 속성 중심 질의라 하고, “현재 위치에 백화점을 신설하는 경우 영향을 받는 재래 시장들은 어떤 것이 있는가?” 와 같이 대상 데이터의 영역(Region) 속성을 기준으로 하는 질의를 주로 많이 사용하는 경우를 데이터의 영역 속성 중심 질의라 한다.

IV. 성능 평가 결과

본 장에서는 3장에서 제시한 4가지 기준을 바탕

으로 데이터의 공간 속성을 고려한 두 가지 캐쉬 교체 방법 CR_LOC와 CR_REG 방법을 비교하여 성능 평가를 수행하고 그 결과를 바탕으로 이동 호스트의 성능과 환경에 따른 최적의 캐쉬 교체를 수행할 수 있는 캐쉬 교체 선택 알고리즘 (Cache replacement selection algorithm)을 제시한다. 논문에서 수행한 캐쉬 교체 알고리즘에 대한 성능평가는 서대문구 지역의 공간 데이터를 대상으로 Informix Universal Server와 Informix Spatial Data blade Module을 사용하여 구현하였고^[7,8], 이동 호스트의 위치를 변경하면서 다양한 조건하에서 캐쉬 교체 알고리즘의 성능을 평가하였다.

성능 평가의 기준인 적중률(hit ratio)은 발생한 전체 질의의 수에 대하여 이동 호스트의 캐쉬에 질의에 해당하는 데이터가 존재하는 비율로 측정한다.

[정의 10] 캐쉬 적중률

$$= \frac{\text{캐쉬에 데이터가 존재하는 질의의 수}}{\text{전체질의수}} * 100(\%) \blacksquare$$

1. 실험1: 검색 대상 데이터의 분산도의 변화에 따른 캐쉬 적중률 변화

동일한 캐쉬 크기를 갖는 이동 호스트가 동일한 지역 밀집도를 갖는 지역을 이동하더라도 검색 대상 데이터의 분산도가 변화함에 따라서 캐쉬 적중률이 변화하는 것을 살펴본다. 이때 사용자의 질의 유형은 데이터의 영역 속성 중심 질의를 기준으로 한다. 그림 4는 캐쉬 영역은 9일 때 DTO를 변화하면서 CR_REG을 이용하여 캐쉬 교체를 수행한 결과이다. 실험 결과 사용자의 질의에 대한 DTO가 캐쉬 영역보다 작은 경우는 이동 호스트의 캐쉬 크기가 질의에서 처리해야 할 데이터 수보다 많기 때문에 높은 캐쉬 적중률을 보이는 반면 DTO의 크기가 캐쉬 영역보다 커지는 DTO 9를 기준으로 캐쉬 적중률이 급격하게 감소한다. 이는 이동 호스트의 캐쉬 크기가 질의에서 처리해야 할 데이터 수보다 작은 경우 지속적인 캐쉬 교체가 발생하면서 캐쉬 적중률이 낮아지기 때문이다. 실험1의 결과는 CR_REG의 적용이 사용자의 위치에서 일정한 범위 내의 질의가 주로 발생한다는 가정에 기반을 두기 때문에 질의 대상 데이터의 분산도가 일정한 수준 이하인 경우에 적합한 방법임을 보여준다.

2. 실험2: 지역의 밀집도와 사용자 질의 유형에 따른 캐쉬 교체 알고리즘별 캐쉬 적중률 변화

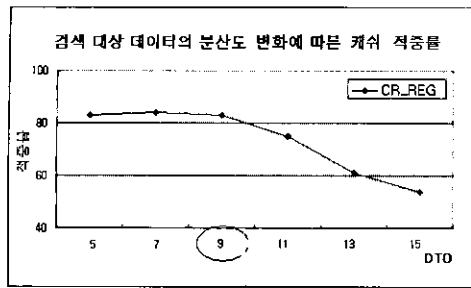


그림 4. 캐쉬 영역이 9 인 경우 DTO의 변화에 따른 캐쉬 적중률 변화

실험 2에서는 이동 호스트가 이동하는 대상 지역의 밀집도와 사용자의 질의 유형별로 캐쉬 교체 알고리즘들의 캐쉬 적중률 변화를 실험하였다. 즉 대상 지역의 밀집도 (Density of Region)가 최소한 곳과 밀집된 곳 각각에 대해 캐쉬 교체 알고리즘별로 캐쉬 적중률을 비교하는데 사용자의 질의 유형을 영역 중심 질의와 위치 중심 질의로 구분하여 차이를 살펴본다. 사용자의 질의 유형이 데이터의 영역 중심인 것의 결과는 그림5와 같고 질의 유형이 데이터의 위치 중심인 것은 그림6과 같다.

두 가지 질의 유형 모두 지역의 밀집도가 최소한 경우에는 CR_REG 방법이 CR_LOC 방법에 비해 우수한 캐쉬 적중률을 나타내고, 지역의 밀집도가 밀집한 경우에는 CR_LOC 방법이 CR_REG 방법에 비해 우수한 캐쉬 적중률을 나타내었다. 이러한 결과는 일정한 범위의 영역을 처리할 때 캐쉬에서 담당하는 데이터의 양이 CR_LOC 방법에 비해 CR_REG 방법이 더 많기 때문이다. 그러므로 이동 호스트가 평균 데이터 밀집도보다 더 많은 데이터가 존재하는 지역을 통과하는 경우에는 이동 호스트의 영역과 겹쳐지는 데이터를 캐쉬 내에 모두 저장하지 못하는 상황이 발생하고, 이에 따른 지속적인 캐쉬 교체 발생에 의해 CR_REG 방법의 성능이

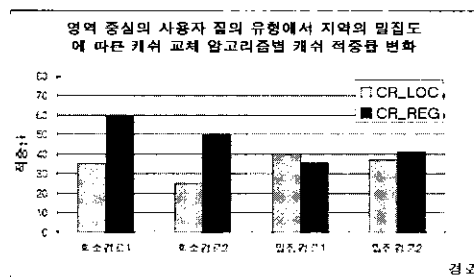


그림 5. 영역 중심의 사용자 질의 유형에서 지역의 밀집도에 따른 캐쉬 교체 알고리즘별 캐쉬 적중률 변화

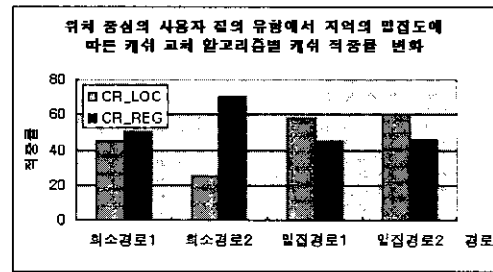


그림 6. 위치 중심의 사용자 질의 유형에서 지역의 밀집도에 따른 캐쉬 교체 알고리즘별 캐쉬 적중률의 변화

떨어지는 결과를 나타낸다. 그러나 같은 경로를 지나고 같은 크기의 캐쉬를 갖는 경우라 할지라도 질의 유형에 따라 서로 다른 적중률을 나타내는 것을 볼 수 있다. 이는 질의 유형에 따라서 캐쉬 교체 알고리즘의 선택이 달라져야 하는 것을 보여준다. 즉 그림5와 같이 사용자의 질의 유형이 데이터의 영역을 중심으로 하는 경우에는 밀집 경로를 지나는 경우에서도 CR_REG 방법이 CR_LOC방법에 비해 거의 같거나 오히려 높은 적중률을 보여주며 이와는 반대로 데이터의 위치를 중심으로 하는 질의가 많은 그림 6의 경우 밀집 경로에서는 CR_LOC 방법이 CR_REG 방법에 비해 월등히 높은 적중률을 보인다. 이러한 결과는 이동 호스트들이 서로 다른 지역 밀집도를 갖는 지역을 이동할 때 더 우수한 캐쉬 적중률을 위해서 사용자의 질의 유형을 고려해야 한다는 것을 보여준다.

3. 실험3: 캐쉬 요구 비율(CRR) 변화에 따른 캐쉬 교체 알고리즘별 캐쉬 적중률

실험 1,2의 결과에 의하면 이동 호스트의 캐쉬 성능은 캐쉬 크기, 검색 대상 데이터의 분산도, 지역의 밀집도, 사용자의 질의 유형, 캐쉬 교체 알고리즘 등 다양한 요인에 따라 변화한다. 그러므로 좀 더 효과적인 캐쉬 교체 방법의 제안을 위하여 캐쉬 크기, DTO 및 지역의 밀집도를 반영하는 새로운 파라미터인 캐쉬 요구 비율(CRR : Cache Requirement Rate)을 다음과 같이 정의하고 CRR 변화에 따른 캐쉬 교체 알고리즘별 캐쉬 적중률을 실험한다.

[정의 11] 캐쉬 요구 비율 (CRR : Cache Requirement Rate)은 이동 호스트가 사용자의 질의에 대해서 100 % 응답하기 위해서 요구되는 캐쉬 크기에 대한 비율로 DTO 범위 내에 존재하는 데이터의 수를 캐쉬 크기로 나눈 값의 %로 나타낸다.

캐쉬 요구 비율 =

$$\frac{\text{DTO 범위 내에 존재하는 데이터의 수}}{\text{캐쉬 크기}} * 100 (\%) \blacksquare$$

그림3의 예에서 DTO범위 내에 존재하는 데이터의 수는 8이고 캐쉬크기는 5이므로 이때의 CRR은 $8/5 * 100 = 160\%$ 로 해당 위치에서 이동 호스트가 사용자의 질의에 대해서 100 % 응답하기 위해서는 자신의 캐쉬 크기의 1.6배에 해당하는 크기의 캐쉬가 요구된다는 의미이다.

실험3은 다양한 크기의 캐쉬를 갖는 이동 호스트들을 대상으로 서로 다른 밀집도를 갖는 지역을 통과하면서 CR_LOC와 CR_REG 두 가지의 캐쉬 교체 알고리즘을 적용하여 캐쉬 요구 비율(CRR)의 변화에 따른 캐쉬 적중률의 변화를 살펴보고 그 결과는 그림 7과 같다.

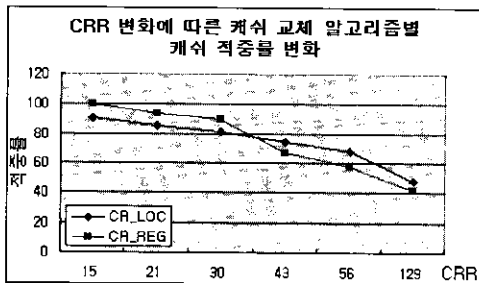


그림 7. CRR변화에 따른 캐쉬 교체 알고리즘별 캐쉬 적중률 변화

CRR값이 작은 경우에는 CR_REG이 더 우수한 캐쉬 교체 성능을 나타내고 반대로 CRR의 값이 큰 경우에는 CR_LOC이 우수한 성능을 나타내는 것을 알 수 있다. 그림 7의 경우 CRR 값 40을 기준으로 40 이하에서는 CR_REG 방법이 더 우수하고 40 이상에서는 CR_LOG 방법이 더 우수한 것을 볼 수 있는데 이와 같이 캐쉬 교체 알고리즘의 선택 기준이 변경되는 CRR 값을 CRR의 Cross Point (CP)라고 정의한다. 즉 실험3에서 CP 값은 40이다.

[정의12] CRR의 Cross Point (CP): CRR의 크기가 증가함에 따라 캐쉬 적중률이 더 우수한 캐쉬 교체 알고리즘이 CR_REG에서 CR_LOC로 변경되는 데 이렇게 캐쉬 교체 알고리즘의 선택이 변경되는 시점의 CRR값을 Cross Point (CP)라고 한다. ■

4. 실험4: 사용자 질의 유형 변화에 따른 CP 값의

변화

실험 4는 CP가 사용자의 질의 유형에 따라 변화하는 것을 실험한다. 그 결과 질의 유형이 데이터의 위치를 중심으로 이루어진 경우는 CP 값이 작고 질의 유형이 데이터의 영역을 중심으로 이루어 질수록 CP 값이 커진다. 이러한 CP 값의 변화는 서버에서 관리하는 데이터의 특징에 따라 결정되는 값으로 서버에서 데이터 방송 시 함께 전송할 수 있는 값이다. 그림8은 실험에서 사용하고 있는 데이터 집합의 경우 사용자 질의 유형 변화에 따른 CP 변화 결과를 나타낸다.

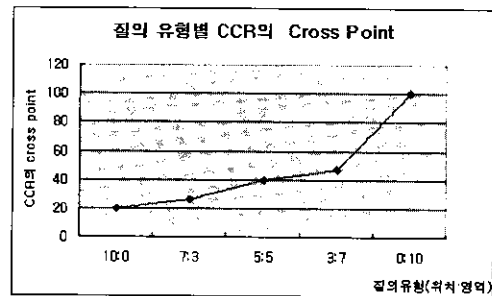


그림 8. 질의 유형별 CP 변화

5. 캐쉬 교체 선택 알고리즘

이동 호스트의 수가 증가하고 이동 호스트의 성능이 증가함에 따라서 점차적으로 캐쉬 관리에 대한 권한은 서버에서 이동 호스트로 이전되는 것이 바람직하다^[1]. 본 연구에서는 캐쉬 교체에 대한 권한을 이동 호스트가 갖는 것으로 가정하며, 동적으로 변화하는 이동 컴퓨팅 환경하에서 사용자의 성능과 대상 지역의 데이터 특색 그리고 응용 프로그램의 특징에 따르는 질의 유형에 따라 사용자 개개인에게 최적의 캐쉬 교체를 수행하기 위한 캐쉬 교체 선택 알고리즘 (Cache replacement selection algorithm)을 제시한다. 이때 캐쉬의 크기, 검색 대상 데이터의 분산도, 사용자의 질의 유형 등의 정보는 이동 호스트가 갖고 있고 서버가 관리하는 지역의 넓이 및 데이터 집합의 개수와 분포, 질의 유형에 따른 해당 지역의 CP값들 등의 정보는 서버가 관리한다고 가정하자.

캐쉬 교체 선택 알고리즘의 시나리오는 다음과 같다. 이동 호스트가 새로운 서버가 관리하는 지역으로 진입하면서 서버와 연결되면(①) 첫째, 이동 호스트는 서버가 방송하는 서버 관리 지역의 크기와 데이터의 전체 개수, 지역의 CP 값들에 대한 정

보를 듣는다(㉑). 둘째, 이동 호스트는 자신의 캐쉬 크기를 이용하여 해당 지역의 평균 밀집도와 자신의 캐쉬 크기에 적합한 캐쉬 영역(CR)의 크기를 구하고(㉒), 서버에서 방송된 CP 값들 중 자신의 질의 유형에 맞는 기준 CP 값을 결정한다(㉓). 셋째, 이동 호스트는 서버에서 주기적으로 방송하는 현재 위치에 대한 데이터 정보를 듣고 있다(㉔), 이동 호스트의 위치와 고유한 DTO 값에 따라 해당 위치에서의 CRR을 결정한다(㉕). 넷째, 이동 호스트의 현재 위치에서의 CRR이 ㉓에서 결정한 이동 호스트의 기준 CP 보다 작은 경우는 CR_REG 방법에 의한 캐쉬 교체를 수행하고(㉖), 더 큰 경우는 CR_LOC 방법에 의한 캐쉬 교체를 수행한다(㉗). 이를 요약한 것이 그림9와 같다.

```

cache_replacement_selection()
{
    Connect Server ;㉑
    Get Server_region , Data_count, CRR_Cross_Point ;㉒
    Compute CR ;㉓
    Set base_CP;㉔
    While Current_Position is included in Server_region
    { Get Broadcast Data ;㉕
      Compute Current_CRR; ㉖
      IF( Current_CRR < base_CP )
          Cache_Replacement_Algorithm = CR_REG ;㉗
      ELSE Cache_Replacement_Algorithm = CR_LOC;
    }
}

```

그림 9. 캐쉬 교체 선택 알고리즘

V. 결론

최근 차세대 고속 무선 통신망의 기반 구축과 이동 컴퓨팅 단말기의 보급 확대를 기반으로 위치의 제약에서 벗어나 좀 더 다양한 정보 서비스를 시간으로 제공받으려 하는 사용자들의 요구가 늘어나고 있고 이를 지원하기 위한 많은 연구가 수행되고 있다. 그러나 이동 컴퓨팅 환경은 좁은 대역폭과 잦은 단절, 전력 소모의 제한과 시간에 따른 이동성이라는 특징으로 인하여 기존의 데이터베이스를 기반으로 하는 응용 시스템과는 다른 많은 문제점을 갖고 있다. 본 논문에서는 이러한 문제들을 해결하기 위해 제안된 캐쉬 교체 알고리즘들 중 시간에 따른 사용자의 이동성과 데이터의 공간 속성을 지원하면

서 다양한 밀집도를 갖는 지역을 이동하는 이동 호스트에 대하여 효율적으로 데이터베이스 서비스를 지원하기 위한 캐쉬 교체 방법인 CR_LOC와 CR_REG을 대상으로 다양하게 변화하는 이동 컴퓨팅 환경 하에서 캐쉬 교체 알고리즘들의 성능에 영향을 미치는 요인들을 캐쉬 크기, 검색 대상 데이터의 분산도, 지역의 데이터 밀집도, 사용자의 질의 유형으로 분류하고, 각각의 기준에 대해 이동 호스트의 위치와 연관된 질의에 대한 처리를 수행하는 성능 평가를 수행하였다. 본 연구를 통하여 캐쉬 적중률에 영향을 미치는 다양한 요소들 간의 관계를 밝히고 이를 이용하여 서로 다른 성능과 질의 유형을 갖는 각각의 이동 호스트가 최적의 캐쉬 교체를 수행할 수 있도록 캐쉬 교체 선택 알고리즘을 제시하였다. 향후 연구 과제로는 이동 호스트가 현재 위치를 기반으로 데이터의 위치와 영역 특성을 고려한 캐쉬 교체 알고리즘을 수행하는데 적합한 서버의 데이터 방송구조 즉 어떤 directory 구조로 데이터를 방송하는 것이 적합한지에 대한 연구가 필요할 것이다.

참 고 문 헌

- [1] J. Jing, A. Helal, A. Elmagarmid, "Client-Server Computing in Mobile Environments", ACM Computing Survey, Vol.31, No.2, pages 117-157, 1999.
- [2] Daniel Babara, Tomasz Imielinski, "Sleepers and Workaholics: Caching Strategies in Mobile Environments," Proceedings of the ACM SIGMOD, pages 1-12, 1994.
- [3] Tomasz Imielinski, Henry F. Korth, "Introduction to Mobile Computing," Mobile Computing, Kluwer Academic Publishers, pages 1-43, 1996.
- [4] Tomasz Imielinski, S. Viswanthan, "Wireless Publishing: Issues and Solutions," Mobile Computing, Kluwer Academic Publishers, pages 299-329, 1996.
- [5] Boris Y. Chan, Antonio Si, Hong V. Leong, "Cache Management for Mobile Databases: Design and Evaluation," IEEE CS Data Engineering, pages 54-63, 1998.
- [6] 김호숙, 용환승, "이동 데이터베이스 시스템에서 데이터의 위치와 영역 특성을 고려한 캐쉬 교체

- 기법”, 정보과학회 논문지 27권1호, pages 53-63, 2000.
- [7] Informix, *Informix Universal Server Guide to SQL: Tutorial Version 9.1*, Informix Press, 1997.
- [8] Informix, *Informix Spatial Datablade Module: User's Guide*, Informix Press, 1997.

김 호 숙(Ho-sook Kim)

정회원



1993년 : 이화여자대학교
전자계산학과 졸업
1993년~1997 : 삼성 SDS
전임연구원
1999년 : 이화여자대학교
컴퓨터학과 공학석사

1999~현재 : 이화여자대학교 컴퓨터학과 박사과정
2001~현재 : 동의공업대학 컴퓨터정보계열 전임강사
<주관심 분야> 공간 데이터베이스, 데이터 마이닝

용 환 승(Hwan-seung Yong)

정회원



1983년 : 서울대학교 컴퓨터학과
학사
1985년 : 서울대학교 컴퓨터학과
공학석사
1985년~1989년 : 한국 전자통신
연구소 연구원

1994년 : 서울대학교 컴퓨터학과 공학박사
1995년~현재 : 이화여자대학교 컴퓨터학과 부교수
<주관심 분야> 웹기반 멀티미디어 데이터베이스
스텝 데이터마이닝, XML 데이터베이스