

# Pull 기반의 데이터 방송 환경에서 최근성을 고려한 캐쉬 대체 전략

신동천

중앙대학교 정보시스템학과

dcshin@cau.ac.kr

## A Cache Replacement Strategy Considering the Recentness in Pull-Based Data Broadcasting Environments

Dongcheon Shin

Dept. of Information Systems, Chung-Ang University

### 요약

pull 기반의 데이터 방송 환경에서는 클라이언트의 데이터 요청을 최소화하기 위하여 캐쉬의 도입이 특히 필요하다. 이러한 환경에서 캐쉬의 도입 효과를 극대화하기 위해서는 히트율 뿐만 아니라 미스 비용도 고려한 효율적인 캐쉬 대체 전략의 수립이 무엇보다도 시스템 성능에 중요한 영향을 미친다. 본 논문에서는 클라이언트의 데이터 액세스 빈도수 및 대기 시간 등을 고려하면서 최신성을 함께 반영하는 캐쉬 대체 전략을 제안한다.

### 1. 서론

통신 기술의 발달은 이동 컴퓨팅이라는 새로운 컴퓨팅 환경의 도래를 촉진시키고 있다. 이동 컴퓨팅 환경의 특징은 상대적으로 낮은 대역폭(bandwidth), 제한된 컴퓨팅 능력, 낮은 배터리 수명, 제한된 인터페이스 화면 등 여러 가지 환경적인 제약이 있다 [1, 2, 3]. 따라서 제한된 대역폭에서 다수의 클라이언트의 데이터 요구를 원활하게 수용하기 위한 데이터 분산 기반 시스템(data dissemination-based system)에 대한 관심이 높아지고 있다. 데이터 방송(data broadcasting) 기법은 대표적인 방법중의 하나이다.

데이터 방송 기법이 갖는 가장 큰 장점중의 하나는 클라이언트의 수에 관계없이 데이터 서비스를 할 수 있다는 점이다. 방송되는 데이터 페이지의 선정과 관련하여 pull, push, hybrid 의 3가지 형태로 구분 할 수 있다[4]. pull 기반의 데이터 방송에서는 낮은 대역폭으로 인하여 많은 클라이언트의 데이터 요청이 이루어지는 경우 시스템 성능이 급격히 저하될 수 있다. 캐쉬의 도입은 필요한 데이터가 방송되기를 기다리지 않고 캐쉬를 액세스하여 효율성을 높일 수 있다. 캐쉬를 고려하는 방법에서 해결해

야 될 중요한 문제중의 하나는 캐쉬내의 데이터 페이지를 새로운 페이지와 대체시키는 페이지 대체 전략이다.

궁극적으로 캐쉬 전략은 방송 알고리즘과 긴밀한 상호 의존성을 갖는다. 예를 들어, 자주 방송되는 데이터는 상대적으로 드물게 방송되는 데이터보다 캐쉬에 존재할 가치가 떨어지며 캐쉬에 존재하지 않는 데이터는 존재하는 데이터보다 상대적으로 자주 방송해야 대기 시간을 줄일 수 있다.

이동 컴퓨팅 시스템에서 데이터 방송 환경을 고려한 많은 전략들이 push 환경에서 제안되었다[5, 6, 7]. 이러한 전략들의 대부분은 데이터 방송 환경에서는 캐쉬의 히트율을 높이는 것뿐만 아니라 캐쉬 미스가 발생하는 경우 서버로부터 요구하는 데이터가 방송될 때까지 기다려야 하므로 미스에 따른 비용까지 고려해야 한다는 관점에서 제안되었다. 지금까지 제안된 대부분의 전략들은 방송 디스크[6, 8]를 가정하였기 때문에, 서버의 방송 스케줄링에 종속적이다. 그러나, pull 기반 환경에서는 이러한 방송 디스크를 적용하기 어렵다. 즉, 데이터를 주기적이며 반복적으로 보내지 않고 요구에 따라 비주기적인 방송

형태를 취하는 pull 환경에서 기존에 제안된 미스 비용을 고려한 캐쉬 대체 전략은 사용될 수 없다.

본 논문에서는 이동 컴퓨팅 환경에서 pull 기반의 데이터 방송 시스템에 적합한 새로운 캐쉬 대체 전략을 제안한다. 기존의 캐쉬 대체 전략과 달리 클라이언트의 데이터 액세스 빈도수와 캐쉬 미스에 따른 방송 대기 시간을 기본적으로 고려한다. 뿐만 아니라, 데이터의 액세스 빈도수는 계속 증가하다가 일정 시간이 지나면서부터 저하한다는 장기 인기도의 변화 패턴[9]을 반영하기 위해 데이터 액세스의 최근성을 함께 고려한다.

본 논문의 구성은 다음과 같다. 2장에서는 캐쉬 대체 전략을 위해 고려한 요소들을 소개하고 새로운 대체 전략을 제안한다. 끝으로, 3장에서 결론을 맺는다.

## 2. 캐쉬 대체 전략

### 2.1 고려 요소

액세스 빈도수는 클라이언트가 특정 페이지를 얼마나 많이 액세스하는 가를 나타낸다. 클라이언트가 많이 액세스하는 데이터를 캐쉬에 두어 서버로부터 방송되기를 기다리는 시간을 줄이므로써 전체적으로 클라이언트의 응답 시간을 줄일 수 있다. 결국, 캐쉬 히트율을 높이기 위해서는 액세스 빈도수의 고려가 필요하다.

대기 시간은 페이지에 대한 요구가 있는 후 실제 서버로부터 요구한 페이지를 받을 때까지 걸린 시간을 의미한다. 일반적으로 push 환경에서 서버는 액세스 빈도수가 높은 페이지를 그렇지 않은 페이지보다 자주 방송하고, pull 환경에서도 액세스 빈도수가 높은 페이지는 자주 요청되기 때문에 자주 방송하게 된다. 따라서, 요청수가 높으면 이에 반비례하여 대기시간은 감소하게 된다. 데이터 요청은 캐쉬 미스가 있는 경우에 이루어지므로 미스에 따른 비용은 대기 시간에 달려 있다. pull 방송 환경에서 특정 데이터의 방송 대기 시간은 서버의 데이터 방송에 따라 가변적이므로 전체적으로 시스템의 성능을 개선하기 위해서는 대기 시간을 고려할 필요가 있다.

기존의 캐쉬 대체 전략에서는 최근에 액세스된 데이터는 향후에도 액세스 될 수 있다는 직관으로 데이터 액세스의 최근성을 전통적으로 고려하고 있다.

### 2.2 캐쉬 대체 전략

pull 기반의 데이터 방송 환경에서 제안하는 효율적인 캐쉬 대체 전략은 다음과 같은 직관에서 출발한다. 아래의 모든 직관에는 최근성을 고려하고 있다.

#### II. 클라이언트가 요구를 자주 하는 데이터 페이지는 캐쉬에 있어야 한다.

클라이언트가 필요로 하는 데이터는 캐쉬에 있거나(cache hit) 방송중이거나(air hit), 혹은 서버에게 요청하는(pull request) 3 가지뿐이다. air 히트인 경우 원하는 데이터가 방송되기를 기다려야 하며 요청하는 경우도 최소한 다음 방송까지 기다려야 한다. 따라서, 자주 요구되는 데이터는 캐쉬에 있어야 가장 빠른 응답시간을 얻을 수 있다. 이를 뒷받침하기 위해 고려하는 요소는 데이터 액세스 빈도수이다.

같은 빈도수를 갖는 페이지라도 긴 시간간격을 갖는 페이지는 짧은 시간간격을 갖는 페이지에 비해 장기 인기도에 따르면 인기가 하락하고 있다고 생각할 수 있다. 반대로 같은 시간간격을 갖는 페이지라면 액세스 빈도수가 많을수록 인기가 상승중이라 판단할 수 있다. 즉, 이러한 직관을 반영하는 액세스 빈도수의 고려는 인기도(P)로 다음과 같이 표현한다.

$$P = R_n / (R_{it} - R_{ft}) \quad (1)$$

여기서  $R_n$ ,  $R_{it}$ ,  $R_{ft}$  는 각각 데이터에 대한 누적된 액세스 수, 가장 최근에 액세스 한 시간, 처음에 액세스 한 시간을 의미한다.

#### II. 클라이언트가 최근에 요구한 페이지일수록 캐쉬에 있어야 한다.

II에서 단위시간에 같은 수의 액세스 빈도수를 갖는 페이지는 동일한 인기도 P를 유지하게 된다. 즉, 같은 시간 간격이면 예전의 빈도수와 최근의 빈도수가 동일한 인기도를 갖게 되어 액세스의 최근성 반영을 하지 못하는 경우가 있다. 따라서, 전통적인 시스템에서 가장 많이 고려한 요소로 최근에 액세스된 데이터가 향후에도 일정기간 많이 액세스될 것이라는 직관을 기반으로 할 필요성이 있다.

한편, 간헐적인 요청으로 최근에 액세스된 데이터가 필요 이상으로 최신성을 인정받는 것을 배제하기 위해 처음에 액세스된 시간을 고려하여 가장 최근에 액세스된

시간을 처음과 가장 최근에 액세스 된 평균 시점으로 조정하여 다음과 같이 최신성(R)을 표현한다.

$$R = R_{ct} - (R_n + (R_{lt} - R_n)/2) = R_{ct} - (R_{lt} + R_n)/2 \quad (2)$$

여기서  $R_{ct}$ ,  $R_{lt}$ ,  $R_n$  각각 현재 시간, 가장 최근에 액세스 한 시간, 처음에 액세스 한 시간을 의미한다.

### 13. 미스비용을 최소화 하기 위해 대기시간이 큰 페이지일수록 캐쉬에 있어야 한다.

최근의 대기시간이 과거의 평균 대기시간보다 크다면 인기가 떨어지고 있어 방송 횟수가 줄게 되므로 캐쉬에 존재할 가치가 높아지며 그렇지 않다면 방송 횟수가 늘게 되어 캐쉬에 존재할 가치가 낮아진다고 판단할 수 있다. 따라서, 과거의 대기시간도 고려하여야 하지만 최근의 대기시간도 같은 비중으로 고려하여 최신성을 반영한다. 이를 위해 대기시간(W)은 단순한 대기시간의 평균을 사용하지 않고 과거의 대기시간의 평균과 가장 최근의 대기시간을 다음과 같이 표현한다.

$$W = (WT_{avg} + WT_1)/2 \quad (3)$$

여기서  $WT_{avg}$ ,  $WT_1$  은 각각 평균 대기시간과 가장 최근의 대기시간을 의미한다.

이상으로부터 인기도(P)와 대기시간(W)가 클수록 그리고 최신성(R)이 작을수록 캐쉬에 존재할 가치가 높아진다고 할 수 있다. 따라서, 캐쉬 대체를 위한 전략은 식 (4)로부터 유도된 PW/R 값이 제일 작은 데이터 페이지를 희생자로 선택한다.

$$PW/R = P \cdot W/R \quad (4)$$

PW/R 값을 구하기 위해 클라이언트는 아래와 같은 정보를 갖는 테이블을 유지한다. PID는 데이터 페이지의 식별자를 의미한다.

$$(PID, R_n, R_{lt}, R_{ft}, WT_{avg}, WT_1)$$

힙프(min heap) 구조나 정렬 리스트로 유지하여 효율성을 향상시킬 수 있다.

### 3. 결론

본 논문에서는 pull 기반의 데이터 방송 환경에서 클라이언트의 데이터 액세스에 대한 최신성을 반영하면서 여러가지 요소들을 함께 고려한 캐쉬 대체 전략을 제안하였다. 고려한 요소 중에서 인기도와 최신성은 클라이언트의 요구 상황을 반영한 것이며 대기시간은 서버의 방송 상황을 고려한 것이라 할 수 있다. 따라서, 제안한 대체 전략은 방송 시스템의 모든 주체의 입장을 고려한 전략이라 할 수 있다. 본 논문에서 제안한 전략의 효율성 검증을 위해서는 기존의 대체전략과의 성능분석을 위한 연구가 필요하다. 아울러, push나 hybrid 방송 환경에도 적용할 수 있도록 확장 혹은 개선되어야 한다.

### 참고문헌

- [1] R. Alonso and H. F. Korth, "Database System Issues in Nomadic Computing," ACM SIGMOD, pp. 388-392, May 1993.
- [2] D. Barbara, "Mobile Computing and Databases - A Survey", IEEE Transactions on Knowledge Engineering, Vol. 11, No. 1, pp. 108-117, January/February 1999.
- [3] G. Forman and J. Zahorjan, "The Challenges of Mobile Computing," IEEE Computer, 27(6), April 1994.
- [4] M. Frnaklin, S. Zdonik, "Dissemination-Based Information Systems", IEEE Data Engineering Bulletin, Vol. 19, No.3, pp. 20-30, Sept. 1996.
- [5] D. barbaba and T. Imielinski, "Sleepers and Workaholics: Caching Strategies in Mobile Environments," ACM SIGMOD, pp. 1-12, May 1994.
- [6] S. Acharya, M. Franklin, S. Zdonik, "Balancing Push and Pull for Data Broadcast", Proc. of ACM SIGMOD, Tuscon, Arizona, pp. 183-194, May 1997.
- [7] V. Liberatore, "Caching and Scheduling for Broadcast Diks Systems", Technical Report 98-71, UMIACS, 1998.
- [8] S. Acharya, R. Alonso, M. Franklin, S. Zdonik, "Broadcast Disks: Data Management for