The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 18, No. 5, pp.131-140, Oct. 31, 2018. pISSN 2289-0238, eISSN 2289-0246

https://doi.org/10.7236/JIIBC.2018.18.5.131
JIIBC 2018-5-17

데이터 액세스 확률의 이동 평균을 이용한 하이브리드 데이터 방송 스케줄링 기법

A Hybrid Data Broadcast Scheduling Scheme Using Moving Average of Data Access Probability

권혁민*

Hyeokmin Kwon*

요 약 방송 스케줄링 기법은 데이터 방송 시스템의 성능을 향상시키기 위해서 필수적인 기술이다. 본 논문은 질의의 응답시간을 줄이기 위하여 다중 채널을 통한 데이터 방송의 문제를 연구하여 HDAMA로 명명된 새로운 하이브리드 데이터 방송 스케줄링 기법을 제안한다. 제안된 기법은 푸쉬기반 방송과 풀기반 방송의 장점을 결합하는 전략을 채택한다. HDAMA는 다중 데이터 질의의 특성을 반영하여 액세스 확률의 영향력을 적절하게 제어할 수 있고, 풀 데이터를 비교적 적시에 방송할 수 있기 때문에 질의 응답시간의 성능을 향상시킬 수 있다. 제안된 기법의 성능 평가를 위해 시뮬레이션이 수행되었다. 실험 결과에 따르면, 평균 응답시간의 성능에서 HDAMA는 다른 기법보다 우수한 성능을 보인다.

Abstract A broadcast scheduling scheme is an essential technique to improve the performance of data broadcast systems. This paper explores the problem of data broadcast over multiple channels to reduce query response time, and proposes a new hybrid data broadcast scheduling scheme named HDAMA. The proposed scheme employs the strategy that combines the advantages of push-based broadcast and pull-based broadcast. HDAMA could enhance the performance of query response time since it is capable of controlling the influence of access probability properly reflecting the characteristics of multi-data queries and broadcasting pull data items relatively on time. Simulation is performed to evaluate the performance of the proposed scheme. The simulation results show that the performance of HDAMA is superior to other schemes in terms of the average response time.

Key Words : broadcast schedule, hybrid broadcast, mobile computing, simulation

Ⅰ. 서 론

인터넷의 폭발적인 인기 및 무선 네트워크의 활성화로 인하여 전 세계는 온라인 데이터 서비스에 대한 요구가 크게 증가하고 있고, 이에 따라 시스템의 확장성(scalability)이 어느 때보다 중요해 지고 있다. 특히, 뉴스 정보, 주식 정보, 일기 정보 및 교통 정보 서비스 시스템과 같은 정보 소비(information consumption) 모델의 형태를 띠고 있는 정보 시스템에서는 응답시간의 성능 저하 없이 어떻게 방대한 클라이언트들을 지원할 것이가가 주된 관심사였다[1-10]. 이와 같이 수많은 클라이언트들을 지원해야 하는 시스템에서는 전통적인 유니캐스트

*종신회원, 세명대학교 정보통신학부
접수일자 : 2018년 8월 15일, 수정완료 : 2018년 9월 15일
게재확정일자 : 2018년 10월 5일

Received: 15 August, 2018 / Revised: 15 September, 2018 / Accepted: 5 October, 2018
*Corresponding Author: hmkwon@semyung.ac.kr
School of Information and Communication, Semyung University, Korea

(unicast) 방식의 데이터 전달 기법은 동일 데이터에 대한 중복 요청 및 중복 응답으로 인하여 서버 및 네트워크 자원을 심각하게 낭비하는 경향이 있기 때문에 확장성 측면에서 큰 문제를 야기한다. 반면, 방송 방식의 데이터 전달 기법을 채택한 정보 시스템은 어떤 데이터를 방송하면, 그 데이터를 원하는 모든 클라이언트가 이를 액세스할 수 있기 때문에 시스템 확장성이 매우 우수하다.

데이터 방송 시스템에서 가장 중요한 성능 지수는 질의(query)의 평균 응답시간(response time)인데, 이는 클라이언트가 특정 질의를 제기한 순간부터 자신이 원하는 모든 데이터들을 수신할 때까지 걸린 대기 시간을 의미한다. 이 응답시간 성능은 방송 스케줄링 기법에 크게 영향을 받는데, 방송 스케줄링 기법은 각 방송채널에 어떤 데이터들을 할당하며, 이들을 어떤 순서로 방송할 것인가의 문제를 다룬다[1-10]. 방송 스케줄링 기법은 방송 데이터의 선정 방법에 따라 푸쉬(Push) 기반과 풀(Pull) 기반 기법으로 구분할 수 있는데[1-3], 푸쉬기반 기법에서는 방송 서버가 클라이언트들이 액세스하는 데이터들의 통계 정보를 바탕으로 미리 방송 스케줄을 구성하여 정해진 순서대로 데이터들을 방송한다. 반면, 풀기반 방송 스케줄링 기법에서는 방송 서버가 클라이언트들의 데이터 요청 정보들을 바탕으로 자신의 방송 스케줄링 알고리즘에 따라 적절한 데이터들을 선정하여 방송한다.

푸쉬기반 시스템은 클라이언트의 수가 크게 증가하더라도 성능 저하 없이 이를 지원할 수 있는 장점이 있지만, 서버의 모든 데이터들을 방송 스케줄에 포함시켜야 하므로 방대한 데이터를 방송해야 하는 경우에는 방송 주기가 매우 길어져 평균 응답시간의 성능이 저하된다. 그리고 푸쉬기반 시스템은 클라이언트들의 액세스 패턴의 변화를 실시간으로 방송 스케줄에 반영하기가 어렵고, 동적으로 변화하는 질의 요청률에 적절히 대응할 수 없다는 단점이 있다. 예를 들어, 방대한 클라이언트들을 지원해야 하는 환경에서도 질의 요청률이 낮은 시간대가 있기 마련인데, 푸쉬기반 시스템은 질의 요청률에 관계없이 항상 동일한 응답시간의 성능을 보인다.

풀기반 시스템은 질의 요청률이 낮은 환경에서 매우 바람직한 성능 특성을 보인다[1-2]. 그러나 풀기반 시스템에서 클라이언트의 수가 어느 이상으로 증가하게 되면 방송 서버가 클라이언트의 데이터 액세스 요청을 취합하여 실시간으로 방송 데이터를 선정하는 것이 불가능하게 된다. 단일 데이터를 액세스하는 질의가 제기되는 환경에서는 각 데이터의 요청 횟수 및 요청 시간 정보만을 이용하여 다음에 방송할 데이터를 선정하므로 클라이언트의 수가 많더라도 적시에 방송 데이터를 선정하는 것이 어느 정도 가능하다. 그러나 다중 데이터 질의 환경에서는 각 질의의 상황 및 질의에 포함된 데이터들을 고려하여 다음에 방송할 데이터를 선정해야 하므로 질의 요청률이 높을 경우에는 적시에 방송 데이터를 선정하는 것이 거의 불가능해진다. 그리고 방대한 클라이언트를 지원해야 하는 환경에서는 상향링크(uplink)의 병목 현상으로 인한 성능 저하를 야기할 수도 있다[1].

하이브리드(Hybrid) 방송 스케줄링 기법은 푸쉬기반과 풀기반 기법의 단점을 보완하기 위하여 액세스 확률이 높은 데이터들은 푸쉬 방식으로 방송하고 인기도가 낮은 데이터들은 클라이언트가 명시적으로 요청하여 온디맨드(on-demand) 형태로 방송하는 정책을 채택한다[1]. 하이브리드 방송 스케줄링 기법은 클라이언트의 액세스 패턴의 동적 변화에 적절하게 대응할 수 있을 뿐만 아니라, 풀기반 기법이 야기할 수 있는 상향 링크의 병목 현상과 다음에 방송할 데이터 선정을 위한 실행부담을 완화할 수 있다는 장점이 있다. 그러나 하이브리드 기법은 푸쉬 방식으로 방송되는 푸쉬 데이터와 풀 방식으로 방송되는 풀 데이터를 적절하게 결정해야 한다.

단일 데이터 질의 환경에서는 각 데이터의 액세스 확률에 기초하여 푸쉬 데이터와 풀 데이터를 결정하면 되는데, 다중 데이터 질의 환경에서는 액세스 확률이 방송 빈도에 미치는 영향력을 적절하게 제어해서 푸쉬 데이터와 풀 데이터를 결정해야 바람직한 성능 특성을 보일 수 있다. 본 논문은 다중 데이터 질의 환경에 한하여 연구를 수행하는데, 각 데이터의 방송 빈도를 그들의 액세스 확률의 이동 평균(Moving Average) 값의 제곱근에 비례하게 설정할 때 최적이라는 휴리스틱을 적용하여 푸쉬 데이터와 풀 데이터를 결정하여, 질의 응답시간의 성능을 개선할 수 있는 새로운 하이브리드 방송 스케줄링 기법을 제안한다. 본 논문의 구성은 다음과 같다. 2 장에서 기존에 개발된 기법에 대하여 기술하고, 3 장에서는 본 논문이 제안하는 새로운 방송 스케줄링 기법에 대하여 기술한다. 4장에서는 제안된 기법의 성능을 평가, 분석하고, 마지막으로 5 장에서 결론을 맺는다.

II. 관련 연구

이 장에서는 다중 방송채널 환경을 위해 제안된 방송 스케줄링 기법들에 대하여 기술한다. 데이터 방송 스케줄링 기법은 푸쉬기반과 풀기반 기법으로 구분되는데, 푸쉬기반 기법에는 단일 데이터 질의 환경을 위해 FLAT[4], DP(Dynamic Programming)[5], NODA(Near Optimal Data Allocation)[6]기법이 제안되었고, 다중 데이터 질의 환경을 위하여 PBA(Placement-Based Allocation)[7]와 DAMA(Data Allocation using Moving Average)[8] 기법 등이 제안되었다. 단일 데이터 질의 환경을 위해 제안된 대부분의 기법들은 데이터의 액세스 확률에 기초하여 데이터를 분할하여 방송채널에 할당했다. 가장 기초가 되는 FLAT 기법은 각 방송채널에 동일 개수의 데이터를 배정하고, 각 채널은 이들을 라운드-로빈 형태로 평형 방송한다.

DP 기법은 동적 프로그래밍 알고리즘을 사용하여 각 방송채널에 할당될 데이터들을 결정했다. 이 기법은 각 방송채널의 데이터들이 FLAT 방식으로 평형 방송되는 시스템 모델에서는 응답시간 성능 측면에서 최적(optimal) 알고리즘이다. 그러나 DP 기법은 같은 채널에 할당된 데이터들을 동일 빈도로 방송하기 때문에 응답시간의 성능을 개선하는 데는 한계가 있었다. 이 한계를 극복하기 위하여 NODA 기법은 각 방송채널이 동일 대역폭을 갖는 다수 개의 서브채널로 구성된다는 개념을 도입했다. 그리고 각 서브채널별로 평형 방송을 실시하는 시스템 모델을 정립하여, 단일 데이터 질의의 응답시간 측면에서 이론적 최적에 거의 필적하는 성능을 실현했다.

PBA 기법은 질의에 포함된 데이터들이 동시에 같은 시간대에 방송됨으로써 질의 응답시간이 커지는 문제점을 해결하기 위한 연구를 하였다. PBA는 더 높은 빈도로 요청되는 질의에 속한 데이터들을 충돌이 발생하지 않도록 우선 할당하여 해당 질의의 데이터 충돌 가능성을 줄임으로써 평균 응답시간 성능을 향상시켰다. DAMA 기법은 질의 프로파일과 요청 확률을 정형화하기 어려운 환경에서 각 데이터의 액세스 확률만을 이용하여 다중 데이터 질의를 처리하기 위한 연구를 수행했다. 다중 데이터 질의에서는 액세스 확률이 높을수록 더 빈번하게 방송되어야 하지만 질의가 액세스하는 데이터의 수가 많아질수록 액세스 확률이 방송 빈도에 미치는 영향력이 작아져야 한다. DAMA 기법은 이 점에 착안하여 각 데이터의 액세스 확률의 이동 평균(Moving Average) 값을 식 (1)과 같이 계산한 후 이를 이용하여 각 방송채널에 데이터를 할당했다. DAMA 기법은 각 데이터의 방송 빈도를 $\sqrt{MA(p_i)}$ 값에 비례하게 설정하는 정책을 사용하여 응답시간의 성능을 개선했다. 식 (1)에서 Q는 한 질의에 포함되는 평균 데이터 항목의 수를 의미하며, R은 Q가 방송빈도에 미치는 영향력을 의미한다.

$$MA(p_i) = \left(\sum_{j=i}^{i+Q\times R-1} p_j\right) / (Q \times R) \qquad (1)$$

기본적으로 푸쉬기반의 방송 스케줄링 기법들은 각 데이터의 액세스 확률이나 요청률을 미리 파악할 수 있어야 한다는 제약이 있고, 동적으로 클라이언트의 액세스 패턴이 변할 경우에는 적용하기가 어렵다는 단점이 있다. 이와 같이 데이터의 액세스 확률을 미리 파악하기가 어렵거나 데이터의 액세스 패턴이 동적으로 변경되는 환경에서는 클라이언트가 전송하는 데이터 요청 정보를 바탕으로 다음에 방송할 데이터를 선정하는 풀기반 스케줄링 기법을 적용하는 것이 바람직하다. 단일 데이터 질의 환경을 위한 풀기반 스케줄링 기법으로는 FCFS(First Come First Served), MRF(Most Requests First), LWL(Longest Wait First), RxW 기법들이 있는데[2, 9], 이들 중에서 RxW 기법은 적절한 실행부담을 보이면서도 매우 바람직한 성능 특성을 보인다[2]. RxW 기법은 서비스 큐에서 클라이언트에서 요청된 각 데이터들에 대한 요청 횟수 및 그 요청의 처음 도착시간 정보를 유지 관리한다. 그리고 데이터의 요청 횟수 R과 그 데이터가 처음 요청된 시간에서 현재까지 흐른 시간 W의 곱으로 계산되는 RxW 값에 근거하여 다음에 방송할 데이터를 선정한다. 가장 큰 RxW 값을 가진 데이터 항목이 다음에 방송할 데이터로 선정된다. RxW 기법은 자주 요청되었거나 요청된 시간이 오래 경과된 데이터를 다음에 방송할 데이터로 선정할 가능성이 높은 기법이다.

PBSA(Profit-Based Scheduling Algorithm)은[3] 다중 데이터 질의 환경을 위해 개발된 풀기반 방송 스케줄링 기법이다. PBSA는 요청된 질의에 포함된 데이터들을 방송함으로써 얻는 이익(Profit)과 각 질의의 기회 비용(Opportunity Cost)를 계산하여 다음에 방송할 데이터를 결정한다. 비록 이 기법이 매우 바람직한 성능 특성을 보이기는 하지만 스케줄링을 위한 실행 부담이 매우 크다

는 단점이 있다. 따라서 클라이언트들의 데이터 요청이 급격하게 증가하게 되면 적시에 방송 데이터를 선정하는 불가능해 질 수도 있다는 단점이 있다.

Ⅲ. 새로운 방송 스케줄링 기법

이 장에서는 본 논문이 제안한 새로운 하이브리드 방송 스케줄링 기법에 대하여 기술한다. 푸쉬기반 방송 스케줄링 기법은 질의 요청률의 변화에 적절히 대응할 수 없고, 풀기반 기법은 질의 요청률이 높아지면 적시에 방송할 데이터를 선정하는 것이 불가능해지는 단점이 있다. 이 단점은 극복하기 위하여 본 논문은 DAMA에서[8] 제안된 식 (1)을 이용하여 다중 데이터 질의의 특성을 고려하여 푸쉬 데이터와 풀 데이터를 구분하여 방송하는 새로운 하이브리드 방송 스케줄링 기법을 제안한다. 본 논문은 DAMA의 연구처럼 질의 패턴을 취합하는 것이 어려운 환경을 고려하여 데이터들의 액세스 확률과 질의의 평균 크기만이 주어진 환경에서 연구를 수행한다.

1. 기본 시스템 모델 및 용어

본 논문은 다음과 같은 시스템 모델과 용어를 사용한다.

1) 방송 데이터들은 $d_k(1 \le k \le N)$로 표현되는 동일 크기의 N 개의 데이터로 구성되며, d_k의 액세스 확률을 p_k로 표현한다. 그리고 데이터는 액세스 확률의 내림차순으로 정렬되어 있다고 가정한다. $d_1 \sim d_{NoOfPushData}$는 푸쉬 방식으로 방송되며 $d_{NoOfPushData+1} \sim d_N$은 풀 방식으로 방송된다.
2) 방송 서버에는 동일 대역폭을 갖는 K 개의 방송채널이 있는데, 이를 $CH_i(1 \le i \le K)$로 표현한다. $CH_1 \sim CH_{NoOfpushCh}$는 푸쉬 방식으로, $CH_{NoOfPushCh+1} \sim CH_K$는 풀 방식으로 데이터들을 방송하는데, 이들을 각각 푸쉬채널, 풀채널이라 명명한다.
3) 푸쉬 데이터들은 그림 1과 같이 $G_1 \sim G_{NoOfPushCh}$의 $NoOfPushCh$ 개의 그룹으로 분할되어 각 푸쉬채널에 할당되는데, 한 데이터는 한 채널에만 할당될 수 있다. $CH_i(1 \le i \le NoOfPushCH)$는 G_i에 속한 데이터들을 라운드-로빈 방식으로 평형 방송하며, 각 G_i의 마지막 데이터의 인덱스를 $ccut_i$로 표현한다.

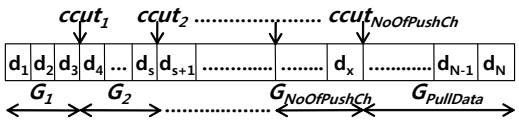

그림 1. 푸쉬 데이터를 *NoOfPushCh* 그룹으로 분할하는 문제
Fig. 1. Problem for partitioning push data in *NoOfPushCh* groups

4) 클라이언트가 풀 데이터를 액세스하려면 방송 서버로 해당 데이터의 방송을 명시적으로 요청하며, 방송 서버는 클라이언트들의 요청을 토대로 적절한 데이터를 선정하여 풀채널을 통하여 방송한다.
5) 한 방송채널에서 하나의 데이터를 방송하는데 걸리는 논리적인 시간을 1 tick으로 정의하고, 응답시간 성능의 단위로서 tick을 사용한다.

2. 푸쉬채널을 위한 데이터 분할 및 할당 알고리즘

단일 데이터 질의에서는 각 데이터들의 방송빈도를 $\sqrt{p_i}$ 값에 비례하게 설정할 때 가장 최적의 응답시간 성능을 발휘한다[10]. 그러나 다중 데이터 질의에서는 가장 늦게 수신된 데이터에 의해 응답 시간이 결정되므로 액세스 확률이 방송빈도에 미치는 영향력을 적절하게 제어해야 한다. 이를 위해 제안된 기법은 DAMA 기법처럼 식 (1)을 이용하여 각 데이터의 액세스 확률 p_i의 이동 평균값을 계산한다. 그리고 각 데이터의 방송 빈도를 $\sqrt{MA(p_i)}$ 값에 비례하게 설정할 때 우수한 성능을 발휘할 것이라는 휴리스틱을 적용하여 각 푸쉬채널에 데이터를 할당한다. 제안된 기법은 DAMA 기법에서 개발된 방식에 기초하여 푸쉬채널에 데이터를 할당하기 때문에 HDAMA(Hybrid DAMA)라 명명한다. HDAMA 기법에서 전체 데이터들을 분할하여 이들을 푸쉬채널에 할당하는 알고리즘이 그림 2에 제시되어 있다.

이 알고리즘은 각 데이터의 $MA(p_i)$ 값을 계산한 후, $\sum_{i=1}^{N} \sqrt{MA(p_i)}$ 값을 계산하여 $SumSrMAForAllData$에 저장한다. 각 데이터의 방송빈도가 $\sqrt{MA(p_i)}$에 비례하기 위해서는 K 개의 방송채널 모두가 푸쉬채널이라는 가정하에 $SumSrMAForAllData$를 각 방송채널에 균등하게 배분해야 한다. 이를 위하여, 그림 2의 ① 부위의 반복문에서는 각 푸쉬채널 CH_i에 할당된 데이터들의 $\sum_{d_j \in CH_i} \sqrt{MA(p_j)}$가 $SumSrMAForAllData/K$에 근접

```
compute MA(p_i) for each data according to Eq. (1)
SumSrMAForAllData = Σ_{i=1}^{N} √MA(p_i)
//determine ccut_0 ~ ccut_NoOfPushCh
AvgSrMAPerCh = SumSrMAForAllData / K
ccut_0 = 0
①for each cut point i from 1 to NoOfPushCh
    ChSumSrMA = 0
    for each data index j from ccut_{i-1}+1 to N
        ChSumSrMA = ChSumSrMA + √MA(p_j)
        if (ChSumSrMA >= AvgSrMAPerCh)
            ccut_i = j
            break;
        end if
    end for
end for
for each cut point index i from 1 to NoOfPushCh
    allocate d_{ccut_{i-1}+1} ~ d_{ccut_i} into CH_i
end for
The remaining data items(d_{ccut_NoOfPushCh+1} ~ d_N) will
be broadcast on demand
```

그림 2. 푸쉬 채널에 데이터 할당 알고리즘
Fig. 2. Data allocating algorithm on push channel

하도록 데이터들의 분할점 $ccut_i$(i=1~$NoOfPushCH$)를 결정한다. 각 $ccut_i$가 결정되면, CH_i(i=1~$NoOfPushCH$)에는 $d_{ccut_{i-1}+1} \sim d_{ccut_i}$(단, $ccut_0$=0임)의 데이터들이 할당되고, $ccut_{NoOfPushCH}$은 $NoOfPushData$가 된다. 각 푸쉬 채널은 자신에게 할당된 데이터들을 라운드-로빈 방식으로 평형 방송한다.

3. 풀채널을 위한 방송 데이터 선정 알고리즘

푸쉬채널에 할당되지 않은 나머지 데이터들 $d_{NoOfPushData+1} \sim d_N$은 풀 데이터로 결정되는데, 클라이언트가 이들을 액세스하려면 방송 서버로 이 데이터의 방송을 요청해야 한다. 풀기반 스케줄링 기법에서 방송 서버가 이용할 수 있는 주된 정보는 데이터에 대한 요청 횟수와 그 요청이 도착한 시간이다. HDAMA 기법은 이 두 가지 정보만을 이용하여 풀 데이터를 스케줄링하는데, RxW[2, 9] 기법에 기초하고 있다. RxW 기법은 R과 W에 동일한 가중치를 주고 있는데 비하여, HDAMA 기법은 R과 W에 서로 다른 가중치를 주는 RxW^a로 명명된 새로운 기법을 제안하여 풀 데이터를 스케줄링하는데 채택한다. RxW^a 기법은 식 (2)로 표현되는 RxW^a 값이 큰 데이터에 우선권을 주어 방송 데이터를 스케줄링한다.

$$RxW^a = R * (Q - first_arrival_time)^a \qquad (2)$$

여기서 R은 해당 데이터의 요청 횟수를 의미하며, Q는 현재 시간을, $first_arrival_time$은 그 데이터가 처음 요청된 시간을 의미한다. 어떤 데이터 d_i가 방송되고 나면, d_i의 $first_arrival_time$은 d_i가 다시 처음으로 요청될 때의 시간으로 재설정된다. R은 해당 데이터가 방송되면 0으로 초기화된다. a 값은 W의 가중치를 의미하는 성능 튜닝을 위한 조정 변수로서 본 논문에서는 a 값을 변화시키면서 a 값이 응답시간의 성능에 미치는 영향을 파악할 것이다. HDAMA 기법은 각 tick마다 풀 데이터들 중에서 RxW^a 값이 가장 큰 (K-$NoOfPushCH$) 개의 데이터를 선정하여 풀채널을 통하여 방송한다. 방송 서버는 이와 같은 방식으로 방송 데이터를 선정하기 위하여 서비스 큐를 운영하는데, 이 큐에는 요청되는 데이터의 아이디, 요청 횟수 및 $first_arrival_time$ 정보가 유지 관리된다.

Ⅳ. 성능 평가 모델 및 성능 평가

이 장에서는 제안된 HDAMA 기법의 성능을 평가하고 그 결과를 기술한다. 비교 대상으로는 FLAT[4], DP[5], DAMA[8], 그리고, RxW[2, 9] 기법을 선정하였다.

1. 성능평가 모델

본 논문의 성능평가 모델은 크게 K 개의 방송채널을 가진 방송 서버와 모바일 클라이언트들로 구성되는데, K 개의 방송채널중 $PullPct$ 만큼의 채널은 풀채널이 되며, 나머지는 푸쉬채널이 된다. 본 논문에서 클라이언트들은 이들의 전체 질의 요청시간 간격이 지수분포(exponential distribution) 형태를 갖는 단일 요청 스트림으로 모델링되었다. 질의는 Q 개의 데이터로 구성되는데, 클라이언트는 현재 질의가 필요로 하는 모든 데이터를 수신한 후 다음 질의를 제기할 수 있다. 클라이언트는 한 순간에는 한 채널만을 수신할 수 있으며, 동일 질의에 포함된 데이터들이 서로 다른 방송채널을 통하여 동시에 방송되면 가장 방송 빈도가 낮은 데이터를 우선 수신한다. 본 성능평가 모델은 MCC에서 개발한 CSIM[11]으로 구현하였으며, 본 논문에서 사용된 입력 변수들은 표 1과 같다.

표 1. 성능 평가를 위한 입력 변수
Table 1. Input parameters for performance evaluation

입력변수	의미	기본값	설정
N	데이터 항목의 수	5000	5000
K	방송 채널의 수	10	1 ~ 20
PullPct	K 대비 풀채널의 퍼센트	50%	20 ~ 80
Θ	데이터 액세스 zipf 분포의 Θ 값	0.95	0.0 ~ 1.0
QRT	tick 구간당 질의 요청횟수	50	1 ~ 80
Q	질의에 포함된 데이터 수	10	1 ~ 15
α	RxW^{α}에서 W의 가중치	2	1 ~ 5
R	Q가 방송빈도에 미치는 영향력	100	100

서버가 방송하는 데이터들은 동일한 크기를 가진 N개의 데이터로 구성되며, 각 데이터의 액세스 확률은 불균등한 액세스 형태를 모델링하는데 많이 사용되는 zipf 분포 모델을[4, 6] 따른다고 가정한다. zipf 분포 모델에서 각 데이터의 액세스 확률은 식 (3)으로 표현되는데, 각 질의는 이 액세스 확률에 근거하여 질의에 포함시킬 데이터를 결정한다. 여기서 θ 값은 각 데이터들의 액세스 확률이 어느 정도 편향되어 있는지를 나타내는 인수로서, θ 값이 커질수록 불균등한 액세스의 정도가 심화된다.

$$p_i = \frac{(1/i)^{\theta}}{\sum_{j=1}^{N}(1/j)^{\theta}} \quad (1 \le i \le N) \qquad (3)$$

HDAMA 기법에서 a 값을 1~5로 변경시키면서 성능 추이를 살펴보았는데, a가 2일 경우에 가장 바람직한 성능 특성을 보이기 때문에 이 값을 2로 설정하고 실험을 실시했다. 본 논문에서는 *PullPct*를 20~80으로 변화시키면서 실험을 실시했는데, 통계적으로 유의미한 큰 차이가 없기 때문에 실험 결과는 *PullPct*를 50으로 설정한 성능 결과만을 제시한다. 방송 스케줄링 기법의 주요 성능지수는 질의의 평균 응답시간이므로 본 논문은 이를 중심으로 성능을 평가한다. 성능평가 실험에서 실험 시작시의 초기 편향(initial bias)을 제거하기 위하여 초기 십만 개의 질의 결과는 무시하였다. 성능 결과에서 제시된 값은 5 개의 서로 다른 임의 수를 사용하여 실시된 실험 결과의 평균값으로, 각 실험은 천만 개의 질의가 처리될 때까지 실시하였다.

2. 질의 요청률 변화에 따른 평균 응답시간의 성능

이 절은 매 tick 구간에서 요청되는 질의의 개수를 의미하는 질의 요청률을 1~80으로 변경시키면서 응답시간의 성능 추이를 살펴보았다. 방송채널의 수를 10, 질의 크기를 10, 그리고 θ 값은 데이터의 편향적인 액세스를 모델링하기 위한 대표 값으로 많이 사용되는 0.95로 설정하고[4, 6] 실시한 실험의 평균 응답시간의 성능이 그림 3에 제시되어 있다.

본 논문이 제안한 HDAMA 기법은 질의 요청률이 극히 작을 경우 이외에는 다른 기법보다 우수한 성능을 보인다. 푸쉬기반 기법인 DAMA, DP, 그리고 FLAT 기법은 방송채널에 미리 데이터를 할당해 놓고 이를 주기적으로 방송하므로 질의 요청률에 관계없이 항상 일정한 성능을 보인다. 그러나 이들은 데이터 할당 정책의 차이로 인하여 서로 다른 성능을 보인다. DP는 다른 기법에 비해 낮은 성능을 보이는데, 이의 주된 원인은 액세스 확률이 낮은 데이터가 질의에 포함되면 이 데이터를 액세스하는데 시간이 많이 걸리기 때문이다. DAMA 기법도 액세스 확률이 낮은 데이터를 액세스하려면 다른 데이터보다 시간이 더 걸리긴 하지만, 이 기법은 다중 데이터 질의의 특성을 반영하여 액세스 확률이 방송 빈도에 미치는 영향력을 적절하게 제어하기 때문에 DP 기법보다 더 우수한 성능을 보인다. HDAMA 기법에서 액세스 확률이 낮은 데이터들은 풀채널을 통하여 방송되는데, 질의 요청률이 낮은 경우에는 이들을 빠르게 스케줄링하는

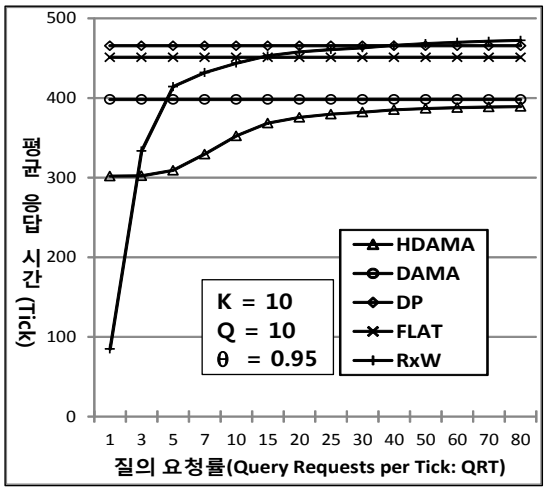

그림 3. 질의 요청률의 변화에 따른 평균 응답시간
Fig. 3. Query requests vs. average response time

것이 가능하다. 따라서 HDAMA 기법은 질의 요청률이 낮을 경우에 DAMA와 DP 기법보다 훨씬 우수한 성능을 보인다.

순수 풀기반 기법인 RxW는 클라이언트가 요청한 데이터만을 방송하기 때문에 질의 요청률이 매우 낮을 경우에는 가장 우수한 성능을 보인다. 그러나 질의 요청률이 증가함에 따라 응답시간의 성능이 급격하게 저하되기 시작해, *QRT*가 40 이상이 되면 가장 낮은 성능을 보인다. 질의 요청률이 낮은 경우에 HDAMA 기법은 RxW보다 낮은 성능을 보이는데, 이는 *QRT*가 낮은 경우에도 푸쉬채널에 할당된 데이터를 액세스하기 위해서는 일정한 시간을 필요로 하기 때문이다.

3. 질의 크기의 변화에 따른 평균 응답시간의 성능

이 절에서는 질의 크기를 1~15까지 변화시키면서 각 기법의 성능추이를 살펴보았다. 이 실험에서 질의 요청률은 HDAMA 기법의 평균 응답시간이 최대에 근접하는 50으로 설정하였다. K와 Θ 값을 각각 10과 0.95로 설정하고 실시한 실험의 평균 응답시간 성능이 그림 4에 제시되어 있는데, HDAMA는 대체로 다른 기법보다 우수한 성능을 보인다.

질의 크기가 커질수록 모든 기법들은 성능이 저하되는데, 이는 질의가 완료되기 위해서는 더 많은 수의 데이터를 액세스해야 하기 때문이다. Q가 증가하면, 액세스 확률이 낮은 데이터도 질의에 포함될 가능성이 커지기 때문에 액세스 확률이 방송빈도에 미치는 영향력이 작아져야 한다. HDAMA와 DAMA 기법은 $Q \times R$ 개 데이터의 액세스 확률의 이동평균 값에 기초하여 푸쉬채널에 데이터를 할당하기 때문에 Q가 커짐에 따라 액세스 확률이 방송 빈도에 미치는 영향력을 감소시킬 수 있다. 따라서 이들은 다른 기법에 비하여 Q가 증가하더라도 성능저하가 상대적으로 작은 편이다. HDAMA 기법은 Q가 4 이상이 되면 다른 기법보다 우수한 성능을 보이기 시작하는데, 참고로 질의 크기가 10이 되면 HDAMA는 DP와 RxW 기법에 비해 각각 20%, 21% 이상의 우수한 성능을 보이는데, 질의 크기가 커질수록 그 성능 차이는 점점 커진다.

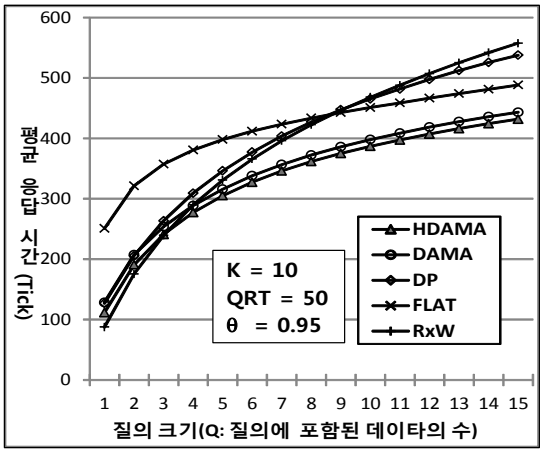

그림 4. 질의 크기에 따른 평균 응답시간
Fig. 4. Query size vs. average response time

HDAMA 기법은 DAMA와 유사한 성능 특성을 보이는데, 이는 HDAMA에서 푸쉬채널에 할당된 데이터들을 액세스하기 위해서는 DAMA와 동일한 시간을 필요로 하기 때문이다. 그러나 HDAMA 기법은 풀 데이터들을 그들의 요청횟수 및 요청시간 정보에 따라 스케줄링하므로 DAMA 기법보다 더 적시에 방송하는 것이 가능하다. 따라서 HDAMA 기법은 전 질의 크기 구간에서 DAMA보다 약간 더 우수한 성능을 보인다.

질의 크기가 작을 경우에는 요청횟수가 많거나 오래 대기한 데이터를 선정하여 방송하면 응답시간의 성능을 개선할 수 있기 때문에 RxW는 다른 기법보다 우수한 성능을 보인다. 그러나 RxW 기법은 요청횟수의 가중치가 너무 높은 경향이 있기 때문에 액세스 확률이 낮은 데이터들은 다음번 방송 데이터로 선정될 가능성이 낮아지게 된다. 따라서 RxW 기법은 질의 크기가 증가하게 되면 다른 기법에 비해 급격하게 성능이 저하되기 시작하여 Q가 10 이상으로 증가하게 되면 가장 낮은 성능을 보이기 시작한다. FLAT 기법은 액세스 확률에 관계없이 모든 데이터들을 동일 빈도로 방송하므로 질의 크기가 증가하더라도 성능 저하는 상대적으로 작은 편이다. 그러나 이 기법은 Q가 작을 경우에는 다른 기법에 비해 매우 낮은 응답시간의 성능을 보인다. DP 기법은 액세스 확률이 낮은 데이터의 방송빈도를 낮게 하는 정책을 사용함으로써 액세스 확률이 높은 데이터들을 빈번하게 방송할 수 있었다. 그러나 질의 크기가 커지면 액세스 확률이 낮은 데이터들도 질의에 포함될 가능성이 높아지기 때문에 Q가 증가하면 DP 기법의 응답시간 성능은 급격하게 저하된다.

4. 채널 수의 변화에 따른 평균 응답시간의 성능

이 절에서는 방송채널의 수를 1~20으로 변화시키면서 각 기법의 성능 추이를 살펴보았다. 이 실험에서 Q는 10으로, θ 값은 0.95로 설정하였다. 질의 요청률은 시스템에 트랜잭션의 포화상태가 시작되는 10, 그리고 트랜잭션이 포화상태에 이르는 50으로 설정하고 실시한 실험의 평균 응답시간 성능이 각각 그림 5와 그림 6에 제시되어 있다.

HDAMA 기법은 풀채널 수를 $\lfloor K \times PullPct/100 \rfloor$으로 계산하므로, K가 1일 경우에는 풀채널이 존재하지 않는다. 그리고 HDAMA는 다른 푸쉬기반 기법과 동일하게 방송채널 별로 평형 방송을 실시하는 시스템 모델을 사용한다. 따라서 K가 1일 경우에는 HDAMA, DAMA, DP, 그리고 FLAT 기법은 모든 데이터들을 같은 빈도로 방송하게 되므로 이들은 모두 동일한 성능을 보인다. RxW 기법은 클라이언트의 요청 정보를 바탕으로 적절한 데이터를 선정하여 방송하므로 타 기법들과는 다른 성능 결과를 보이는데, 다중 데이터 질의의 특성을 스케줄링 정책에 제대로 반영하고 있지 않기 때문에 낮은 성능을 보인다.

그림 5와 그림 6를 살펴보면, 방송채널의 수가 증가함에 따라 모든 기법들의 질의 응답시간이 감소하는 것을 확인할 수 있다. 이는 K가 증가함에 따라, 이에 비례하여 각 tick당 방송할 수 있는 데이터의 수가 증가하기 때문이다. 이 실험에서 HDAMA 기법은 단일 방송채널인 경우를 제외한 모든 채널 구간에서 다른 기법보다 우수한

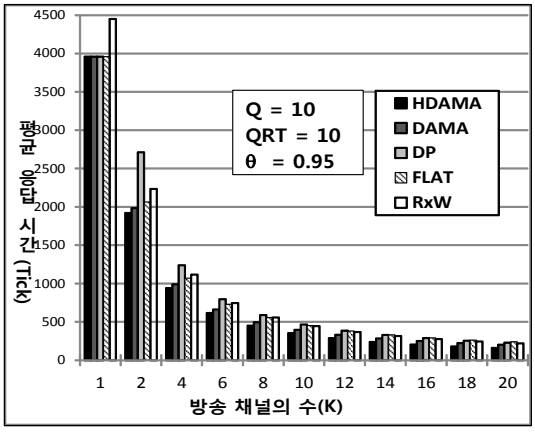

그림 5. 방송 채널의 수에 따른 평균 응답시간(QRT=10)
Fig. 5. No. of channel vs. average response time (QRT=10)

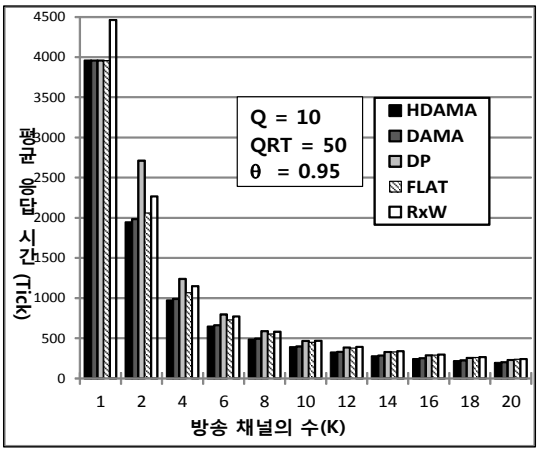

그림 6. 방송 채널의 수에 따른 평균 응답시간(QRT=50)
Fig. 6. No. of channel vs. average response time (QRT=50)

응답시간의 성능을 보이는데, 채널의 수가 증가하면 대체로 더 큰 상대적인 성능 차이를 보인다. 이는 HDAMA 기법은 다중 데이터 질의의 특성을 반영하여 액세스 확률이 방송 빈도에 미치는 영향력을 적절하게 제어하고 있을 뿐만 아니라, 액세스 확률이 낮은 풀 데이터들은 클라이언트의 요청 정보를 받아 RxW^{α} 값에 기초하여 온디맨드 형태로 방송하므로 다른 기법보다는 더 적시에 방송하는 것이 가능하기 때문이다. 참고로 그림 5에서 K가 10일 경우에 HDAMA 기법은 DAMA, DP, FLAT, 그리고 RxW에 비해 각각 13%, 32%, 28%, 26% 정도 더 우수한 성능을 보이는데, K가 20으로 증가하면 26%, 44%, 49%, 36%로 성능 차이가 더 커진다. 그림 6에서 K가 10인 경우의 성능을 살펴보면, 각 기법들은 대략 400 tick 이상의 응답시간을 보인다. 그리고 tick당 50 개의 트랜잭션들이 제기되므로 시스템 내에는 대략 20,000 개 이상의 트랜잭션들이 동시에 실행되고 있을 것이다. 이와 같이 시스템에 트랜잭션이 포화 상태에 이르는 환경에서도 HDAMA는 다른 기법에 비하여 일관되게 우수한 성능을 발휘한다.

본 논문에서 그 결과는 제시하지 않지만, θ 값을 0~1.0으로 변화시키면서 평균 응답시간의 성능추이를 살펴보았다. θ 값이 0일 경우에는, 각 데이터의 액세스 확률이 동일하게 되므로 스케줄링 정책에 관계없이 각 기법들은 거의 유사한 성능을 보인다. 그러나 θ가 증가함에 따라 각 데이터간의 액세스 확률의 차이는 점점 커지게 되며, 이에 따라 모든 기법들의 응답시간 성능은 향상되

는데, HDAMA와 DAMA는 다른 기법들에 비하여 급격하게 성능이 향상되는 편이다.

V. 결 론

본 논문은 방송 방식의 데이터 전달 기법을 채택한 다중 방송채널 환경의 정보시스템에서 다수의 데이터를 검색하는 다중 데이터 질의를 효율적으로 처리하기 위한 연구를 수행했다. 이 연구는 다중 데이터 질의의 응답시간을 최소화시키는 것을 목표로 하여 푸쉬기반 방송과 풀기반 방송의 장점을 결합할 수 있는 HDAMA라 명명된 새로운 하이브리드 데이터 방송 스케줄링 기법을 제안했다. 본 논문은 다중 데이터 질의를 위한 시뮬레이션 모델을 설계 구현하여 성능평가를 실시하였다. 제안된 기법은 다중 데이터의 특성을 반영하여 액세스 확률의 영향력을 적절하게 제어할 수 있을 뿐만 아니라, 액세스 확률이 낮은 풀 데이터들은 온디맨드 방식으로 전송함으로써 비교적 적시에 방송하는 것이 가능하기 때문에 대부분의 환경에서 다른 기법보다 우수한 평균 응답시간의 성능을 보였다.

본 논문은 각 데이터의 액세스 확률과 질의 크기만이 주어진 환경에서 다중 데이터 질의의 응답시간을 개선하기 위한 연구를 진행하였다. 본 논문의 미래 연구로서는 질의 패턴 및 질의의 요청 빈도를 파악할 수 있는 환경에서 다중 데이터 질의를 처리하기 위한 연구를 진행할 것이다.

References

[1] M. Narvekar, H. Mukadam, and S.S. Mantha, "Comparison of Broadcast Scheduling Techniques in Mobile Computing Environment," Procedia Computer Sciences Vol. 45, pp. 706-711, 2015.
DOI: https://doi.org/10.1016/j.procs.2015.03.136

[2] D. Aksoy and M.S. Leung, "Pull vs Push: A Quantitative Comparison for Data Broadcast," in Proc. of the 47th IEEE Global Telecommunications Conference (GLOBECOM'04), pp. 1464 - 1468, 2004.
DOI: https://doi.org/10.1109/GLOCOM.2004.1378225

[3] J. Lv, V.C.S. Lee, M. Li, and E. Chen, "Profit-based scheduling and channel allocation for multi-item requests in real-time on-demand data broadcast systems," Data & Knowledge Engineering, Vol. 73, pp. 23-42, 2012.
DOI: https://doi.org/10.1016/j.datak.2011.09.005

[4] S. Acharya, "Broadcast Disks: Dissemination-based Data Management for Asymmetric Communication Environments," Ph.D. thesis, Brown University, 1998.

[5] W.G. Yee, S. Navathe, E. Omiecinski, and C. Jermaine, "Efficient Data Allocation over Multiple Channels at Broadcast Servers," IEEE Trans. on Computers, Vol. 51, No. 10, pp. 1231-1236, Oct. 2002.
DOI: https://doi.org/10.1109/TC.2002.1039849.

[6] H. M. Kwon, "A Near Optimal Data Allocation Scheme for Multiple Broadcast-Channel Environments," The Journal of the Institute of Webcasting, Internet and Telecommunication, Vol. 12, No. 1, pp. 17-27, 2012.
DOI: https://doi.org/10.7236/JIWIT.2012.12.1.17.

[7] H.P. Hung, J.W. Huang, J.L. Huang, and M.S. Chen, "Scheduling dependent items in data broadcasting environments," ACM SAC 2006.
DOI: https://doi.org/10.1145/1141277.1141555

[8] H.M. Kwon, "A Broadcast Data Allocation Scheme for Multiple-Data Queries Using Moving Average of Data Access Probability," The Journal of the Institute of Internet, Broadcasting and Communication(JIIBC), Vol. 14, No. 5, pp. 35-43, 2014.
DOI: https://doi.org/10.7236/JIIBC.2014.14.5.35.

[9] D. Aksoy and M. Franklin, "R×W: a scheduling approach for large-scale on-demand data broadcast," IEEE/ACM Trans. on Networking, Vol. 7, No. 6, pp. 846-860, 1999.
DOI: https://doi.org/10.1109/90.811450

[10] C. Hsu, G. Lee, A.L.P. Chen, "A near optimal algorithm for generating broadcast programs on multiple channels," Proc. ACM 10th Int'l Conf. CIKM, Atlanta, Georgia, pp. 303-309, 2001.

DOI: https://doi.org/10.1145/502585.502637.
[11] H. Schwetman, 'CSIM Users' Guide for Use with CSIM Revision 16', Microelectronics and Computer Technology Corporation, 1992.

저자 소개

권 혁 민(종신회원)

- 1984년 : 서울대학교 제어계측공학과 학사
- 1994년 : 한국과학기술원 정보및통신공학과 석사
- 1998년 : 한국과학기술원 정보및통신공학과 박사
- 1984년 ~ 1991년 : 대우전자 중앙연구소 컴퓨터개발부 선임연구원
- 1999년 ~ 현재 : 세명대학교 정보통신학부 교수

<주관심분야 : 분산/병렬데이터베이스, 모바일 컴퓨팅, 데이터 방송시스템>

※ 이 논문은 2017학년도 세명대학교 연구년 지원에 의한 연구임.