

이동컴퓨팅환경에서의 효율적인 캐시 일관성을 위한 방송기법 연구

강무경⁰ 정병수 김대호
경희대학교 컴퓨터공학과
{mydogzia,jeong,tonkey}@dmlab.khu.ac.kr

Broadcasting study for Efficient Cache Consistency in Mobile Computing environments.

Kang Moo Kyung⁰ Jung Byung Soo, Kim Dae Ho
Dept. of Computer Engineering, Kyung Hee Univ.

요 약

최근 이동컴퓨팅(Mobile Computing) 환경에서의 잦은 접속 단절로 인한 클라이언트 캐시 일관성(consistency) 문제를 해결하기 위한 방법과 함께 통신 대역과 에너지의 제약을 고려한 연구가 진행되고 있다. 이러한 캐시 유지 방법의 한 분야로서, 방송(Broadcasting)에 대한 연구가 진행 중이다. 그러나 기존의 무효화 보고에 대한 연구는 서버에서의 갱신 정도와 상관없이 전송되기 이전 일정 시간동안의 서버 갱신에 대한 정보만을 포함한다. 이러한 전송기법은 다수의 클라이언트가 필요로 하면서 갱신정도가 빈번한 데이터에는 적합한 방송 방법이 아니다. 본 논문은 이 같은 단점을 보완 하기 위해 자주 갱신되는 데이터의 빈도수에 따라 데이터를 방송하고 클라이언트의 요청에 의해서 데이터 방송빈도수를 변경하여 클라이언트가 특정 데이터의 응답시간을 최소화 할수 있도록 하는 방안을 제시하였다.

1. 서 론

무선통신 기술의 발달과 함께 고성능 휴대용 컴퓨터의 등장은 이동 컴퓨팅(Mobile Computing) 환경을 현실화 시켰다. 이동 컴퓨팅 환경은 클라이언트들이 고정 네트워크로 연결되어 있는 것이 아닌 언제 어디서든 자유로운 이동을 하면서 네트워크로 연결되어 있는 환경을 말한다. 그러나 이 같은 환경은 고정네트워크보다 많은 제약조건들이 발생한다. 이동 컴퓨터(노트북, PDA, PALM 등)은 에너지 제한, 무선 통신 대역폭(Wireless channel bandwidth)의 협소함, 무선 통신망과 기기의 안전성 결여 단말기의 화면 크기 등으로 뺄 수 있다[1, 2]. 또한 클라이언트와 서버간의 데이터 양 및 통신 능력을 고려해 볼 때, 서버 측의 가용 대역폭에 비하여, 클라이언트가 가진 가용 대역폭은 상대적으로 적을 뿐만 아니라, 서버 측에 요구하는 클라이언트의 요구 데이터도 서버 측에서 클라이언트 측으로 전송하는 데이터 양 보다 작다. 이러한 무선 통신의 송신 수신 특성에 기인한 통신 환경을 비대칭(Asymmetric)통신환경이라고 한다[3]. 이러한 환경에서는 클라이언트가 서버 측으로 직접 데이터를 요청하는 기존의 방법보다 클라이언트의 요청 없이도 서버 측에서 데이터를 클라이언트에게 전달하는 방식인 데이터 방송(Data Broadcasting) 기법이 제기되었다[4, 5, 6]. 그러나 이 같은 방법기법은 다수의 클라이언트를 그 대상으로 하고 있기 때문에 특정 클라이언트가 원하

는 데이터를 수신하기 위해서는 상당시간 동안 방송채널(Broadcast Channel)에서 기다려야 한다는 문제점이 발생한다. 이러한 문제점을 해결하기 위하여 클라이언트는 필요한 데이터를 서버에 요청을 하고, 서버는 클라이언트가 필요로 하는 데이터를 클라이언트에 전송하여 서버에서 갱신된 데이터에 대한 일관성(Consistency)을 보장한다.

최근 이러한 방법을 기반으로 클라이언트 캐시의 일관성을 유지하는 많은 방법들이 제안되었으나 기존의 방법들은 클라이언트의 접속 단절이 일정 시간 이상 지속될 경우 서버에서의 갱신 정도와 상관없이 캐시의 모든 데이터를 버려야 한다는 문제점을 가지고 있다. 이에 따라 본 논문에서는 클라이언트의 접속 단절 시간 보다는 서버에서의 갱신정도에 따라 클라이언트 캐시의 일관성을 유지할 수 있는 방안을 제시한다.

2. 관련연구

이동 컴퓨팅 환경의 일반적인 구조는 그림 1과 같다. 이동 컴퓨팅 환경은 크게 네트워크 안에서 이동하는 이동 호스트와 이동하지 않는 고정 호스트(fixed host)의 두 가지 요소로 구성된다. 그리고 고정 호스트 중에서 무선 통신 인터페이스를 가지고 이동 호스트와 통신하며 서비스를 제공해주는 호스트를 지구국이라 한다. 하나의 지구국이 서비스해줄 수 있는 영역을 셀이라 하며, 같은

셀 안에 있는 이동 호스트들은 무선 채널(wireless channel)을 통하여 동일한 지구국과 통신을 한다.

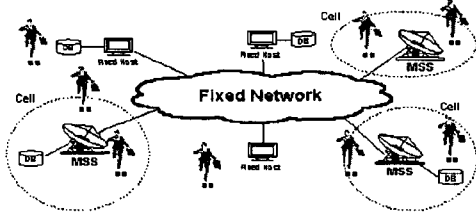


그림 1. 이동 컴퓨팅 환경

호스트의 넓은 대역폭과 풍부한 에너지를 가지고 있으며 유선 망(wired network)에 비해 클라이언트가 가진 가용 대역폭은 상대적으로 적을 뿐만 아니라 서버 측에 요구하는 클라이언트의 요구 데이터도 서버 측에서 클라이언트 측으로 전송하는 데이터양 보다 작다. 이러한 환경에서는 클라이언트가 서버 측으로 직접 데이터를 요청하는 기존의 방법보다, 클라이언트의 요청 없이도 서버 측에서 데이터를 클라이언트에게 전달하는 방식인 데이터 방송(Data broadcast)기법이 제기되었다[4, 5, 6].

기존의 스케줄링을 이용한 클라이언트 캐시 일관성 유지 기법은 크게 stateless 서버와 stateful 서버를 가정한 방법 두 가지로 구분할 수 있다.

stateful 서버 기술은 클라이언트의 캐시 상태를 알고 있는 서버가 클라이언트의 임의의 접속 단절에 대하여 클라이언트 캐시를 유지 할 수 있도록 하였다. 그러나 이러한 기법은 stateful 서버를 유지하기 위해 클라이언트가 특정 데이터를 캐시에 저장할 때마다 서버에 그 정보를 전송해야 한다는 면에서 대역폭의 낭비를 초래하게 된다.

stateless 서버를 가정한 방법은 크게 다시 두 부분으로 구분된다. 첫 번째는 클라이언트가 질의 처리를 할 때마다 서버에 캐시 데이터의 일관성 확인 요청 메시지를 전송하는 방법이다. 그러나 이 방법은 질의 수행을 할 때마다 네트워크에 상당량의 데이터 전송을 요구하게 된다. 두 번째는 서버가 주기적 혹은 비 주기적으로 서버에서 일정 시간동안 갱신된 데이터들에 대한 정보를 방송하는 방법이다. 이 방법은 위의 두 방법에 비하여 대역폭의 오버헤드를 줄일 수 있어 클라이언트 캐시 일관성 유지를 위한 방법으로 가장 널리 사용되는 방법이다.

2.1 기존 방송기법

이 절에서는 이전 데이터 방송 스케줄링기법 들인 FCFS, Broadcast Disk(BD)를 소개한다.

2.1.1 FCFS

FCFS 스케줄링 기법은 도착순서에 의한 데이터 방송 스케줄링이 이루어지는 순차 데이터 서비스 스케줄링에 속한다. 즉 클라이언트에 의하여 요구되는 데이터의 요

구율이나 데이터 요구 빈번성에 대한 고려 없이, 직관적으로 전체 요구 데이터 중 측정시점에서 요구시점이 가장 우선인 데이터를 선정하게된다. 따라서, 요구된 데이터는 서버의 임시 저장소에 요구순서로 배열하게 되며 배열순서에 따른 순차적인 데이터 방송이 이루어진다.

2.1.2

2.1.3 BD 기법

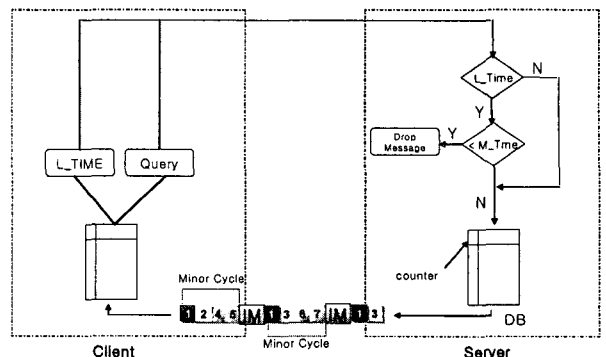
Acharya에 의하여 제기된 방송 데이터 스케줄링 기법이다. Acharya는 데이터 방송시 일정한 주기로 반복되어 데이터 전송이 이루어지는 것이 클라이언트의 응답시간 개선에 효율적임을 설명하였으며, 평균적인 클라이언트 응답시간을 위해서 많은 클라이언트가 접근 할것으로 예상되는 데이터를 상대적으로 자주 방송할 필요가 있음을 지적하였다. 따라서 방송을 위한 데이터는 선호도를 통하여 Hot,Cold 의 여부를 선별한 후 Hot 데이터는 Cold 데이터보다 높은 비율로 방송하여 빈번히 요구되는 데이터의 액세스 확률을 높이는 방법이다.

본 논문은 비대칭통신환경인 모바일 환경에서 Hot 데이터와 Cold 데이터를 모바일 유닛을 사용하는 사용자들 통해서 정확히 판정 할 수 있고 클라이언트가 서버로 요청하는 질의를 줄여 대역폭의 낭비를 줄 일수 있는 Inter-Message 방송 스케줄링 방식을 제안하였다.

3. 제안사항

3.1 Inter Message

FCFS 방법처럼 클라이언트가 접속단절로 인하여 하나 이상의 방송을 수신하지 못하는 경우 캐시의 데이터 전체를 버려야 하는 문제를 해결하고 BD 방법처럼 데이터 정적인 스케줄링을 가지지 않고 서버로의 업링크를 줄일 수 있는 방법으로 Inter_message를 제안 한다. Inter Message 방법은 단절된 클라이언트가 재접속할시 연결된 클라이언트의 캐시 상태 까지 고려한 방법이다. 그림 2는 Inter Message 방법의 전체 시스템의 구조를 보여 주고 있다.



<그림2 Inter-Message System>

3.1.1 클라이언트 작업모델

클라이언트는 서버에서 잦은 갱신이 이루어지는 Hot 데이터와 갱신이 거의 이루어지지 않은 Cold 데이터를 분류하여 한 주기 동안에 다른 빈도수로 데이터를 보내는 것을 받아 클라이언트의 캐쉬 안에 있는 각 데이터를 갱신 하여 캐쉬 일관성을 유지한다. 아래 표는 클라이언트 캐쉬 구조를 나타낸다.

L_TS	Data_ID	Data_value

<표1. 클라이언트 캐쉬 구조>

L_TS: 마지막에 받은 데이터의 타임스탬프 값

Data_ID: 데이터 식별자

Data_value: 실제 데이터 값

서버에서 방송되는 데이터 중 받지 못한 데이터가 있거나 모바일환경의 특징인 잦은 단절로 인해 방송을 듣지 못할 경우에는 클라이언트는 서버로 자신이 필요로 하는 데이터를 요청(Query)을 하게 된다. 그리고 접속 단절되었다가 재접속된 클라이언트는 재접속 당시 자신의 캐쉬안에 있는 데이터중 가장 최근에 받은 타임스탬프 값(L_TS) 을 서버로 전송하여 자신이 접속단절된 시간의 값인 타임스탬프를 보내게 된다.

3.1.2 서버 작업 모델

서버 프로세스는 데이터 방송 서버를 하나로 가정하고 하나의 서비스 서버가 존재하며, 다수 클라이언트 프로세스가 요청하는 데이터를 방송을 통하여 전송 한다. 아래 표2는 서버의 데이터베이스 구조를 나타내고 있다.

L_TS	M_TS	Data_ID	Data_Val	CH_TS	counter

<표2 서버의 데이터베이스 구조>

L_TS : 최근에 보낸 타임스탬프 값

M_TS : 최근에 갱신된 타임스탬프 값

CH_TS : 특정데이터(Data_ID)가 갱신된 타임스탬프 값

counter : 클라이언트가 요청(Query)한 빈도

클라이언트로부터 요청된 데이터와 접속 단절된 클라이언트가 재접속시 보내어지는 L_TIME 스탬프 값을 통하여 서버는 클라이언트로부터 오는 메시지를 구분하여 작동하게 된다. 클라이언트로부터 특정 데이터를 요청하는 메시지가 온다면 서버는 요청된 데이터의 카운터 값을 올리고 요청한 클라이언트에게 데이터 값과 타임스탬프 값을 보낸다. 그리고 재접속 메시지인 L_Time 메시지가 서버로 온다면 서버는 L_Time 값이 서버의 데이터베이스에서 낮은 갱신타임스탬프 값인 M_Time 값보다 작다면 클라이언트가 가지고 있는 모든 캐쉬는 무효화 하여야 하기 때문에 클라이언트에게 Drop_Message를 보내어 클라이언트 캐쉬의 내용을 모두 삭제하게 하고 방송되는 데이터를 받게 한다. L_Time 값이 M_Time 값보

다 크다면 서버는 L_TIME 보다 큰 CH_TS 데이터의 카운터 값을 높여 주어 다음 주기에 이루어질 방송에서 서버는 카운트 값에 따라 동적으로 다수의 클라이언트가 필요로 하는 데이터 값을 알아내어 Hot 데이터와 Cold 데이터를 분류 할수 있고 그에 맞는 빈도수로 데이터를 방송하여 클라이언트가 자신이 필요로 하는 데이터의 대기시간을 최소화 할 수 있다.

4. 결론

이동컴퓨팅 환경에서의 잦은 접속 단절로 인한 클라이언트 캐시 일관성(consistency) 문제를 해결하기 위한 방법과 함께 캐쉬 히트율을 높이면서 통신 대역과 에너지의 제약을 줄이고 클라이언트로부터 오는 데이터를 통해서 서버의 방송 스케줄링이 정적에서 동적으로 해주어 항상 높은 히트율을 유지 할수 있는 Inter_Message 방법을 제안하였다. 향후과제로 Drop Message를 보내기 위한 M_Time의 최적의 범위와 클라이언트의 캐쉬 크기에 따른 성능 평가를 필요로 한다.

5. 참고문헌

- [1] G.Forman and J. Zahorjan, " The Challenges of Mobile Computing," IEEE Computer, Vol. 27, No.6, pp.38-47, April 1994.
- [2] D.Barbara,"Mobile Computing and Database : A Survey", IEEE Trans. On Knowledge and Dta Eng., pages 108-117,1999.
- [3] S. Acharyaya, R. Alonso, M.J Franklin and S.Zdonik. "Broadcast Disks: Data Management for Asymmetric Communications Environments" In Proc. ACM SIGMOD Intl Conf, on Management of Data, Phoenix, Ariz, pp.183-194,May 1997.
- [4] K. Wu, P. S. Yu, and M. Chen, "Energy-Efficient Caching for Wireless Mobile Computing", Proc.12th Int'l Conf. Data Eng., pages 336-343, 1996.
- [5] D. Barbara and T. Imielinski, "Sleepers and Workaholics : Caching Strategies in Mobile Environments", Proc. ACM SIGMOD Int'l Conf. Managements of data, pages 1-12, May 1994
- [6] G. Cao, "A Scalable Low-Latency Cache Invalidation Strategy for Mobile Environments," Int.Conf, on Mobile Computing and Networking, pp.200-209, 2000