

다중 무선 방송채널에서의 상호관련 데이터들을 위한 효율적인 데이터할당 방법

박성욱, 정성원

서강대학교 컴퓨터학과

e-mail : psw0405@sogang.ac.kr, jungsung@sogang.ac.kr

Efficient data allocation scheme for dependent items over multiple wireless broadcast channels

Sung-wook Park, Sungwon Jung
Dept. of Computer Science, Sogang University

요 약

무선방송환경은 모바일 클라이언트의 수에 상관없이 다수의 클라이언트에게 데이터를 보낼 수 있다는 특징으로 인하여 많은 관심을 받아왔다. 그러나 기존 대부분의 연구는 한 개의 독립적인 데이터를 최단시간에 획득하는 것을 목표로 하였고, 질의어와 같이 상호 관련된 여러 데이터를 동시에 획득해야 하는 경우에 대한 연구는 미비했다. 또한 그러한 논문들에서도 질의어 내의 데이터들이 다른 채널상의 동일시간에 나타나지 않도록 할당하는 문제에 대해서만 연구되어 왔고 각 데이터의 요청빈도에 따른 데이터의 접근확률은 반영되지 않았다. 우리는 본 논문에서 이러한 데이터들을 채널에 할당하는 문제뿐만 아니라 여러 질의어에 포함되어 상대적으로 요청빈도가 높은 데이터들을 한 방송 사이클 동안 다중 채널상에 여러 번 할당되게 하여 해당 질의어에 속한 데이터들을 모두 획득하는데 요구되는 평균응답시간을 줄이는 새로운 방법을 제안한다.

1. 서론

최근 무선 이동통신 기술과 휴대장치의 발전으로 인하여 서버에서 불특정 다수의 클라이언트에게 무선 채널을 통해 주기적으로 데이터들을 전송해주는 방송 기법이 많은 관심을 받아왔다.

그러나 무선방송환경에서의 기존 논문들은 클라이언트가 한 개의 독립적인 데이터만을 요청하고 획득하는데 걸리는 시간을 최소로 만드는 것을 목표로 하였다. 그러나 실제 환경에서는 이렇게 개개의 데이터만을 요구하기 보다는 관련된 여러 데이터들을 동시에 요구하는 경우가 더 많다. 예를 들어 주식정보를 원하는 사용자가 모토로라와 같은 회사의 주식을 원한다면, 관련 회사인 노키아나 삼성과 같은 관련된 회사의 주식정보 또한 요청할 확률이 높다. 본 논문에서의 궁극적인 목표는 이러한 상호관련 있는 데이터들로 구성된 질의어의 평균응답시간을 줄이는 것이다.

이러한 질의어의 평균응답시간을 최소화 시키고자 하는 방법은 [1][2][7]에서 제시되었다. Y.D.Chung 이 제시한 QEM(Query Expansion Method)은 상대적으로 요청될 확률이 높은 질의어에 속한 데이터들에게 먼저 채널에 할당될 수 있는 기회를 주었다[2]. G.Lee 는 [2]에서 제시한 방법의 엄격한 제한을 완화함으로써

개선된 방안을 제시하였다[7]. 그러나 [2][7]이 단일 채널환경만을 고려한 반면 Hung 은 다중 채널에서 적용될 수 있는 새로운 방법인 PBA(Placement-Based Allocation)를 제시하였다[1]. 다중 채널에서의 데이터들은 임의의 다른 채널들에 할당될 수 있다. 이때 채널에 나타난 데이터의 위치를 오프셋이라고 하는데 한 질의어에 속한 여러 데이터들이 같은 오프셋에 위치하는 상황을 가리키는 데이터충돌의 발생 가능성이 가장 큰 문제점이 된다. 데이터충돌이 발생하면 한 순간에 하나의 채널에만 접속할 수 있는 무선환경의 특징으로 인하여 해당 방송 사이클에서 하나의 데이터만을 획득하고 남은 데이터는 다음 방송 사이클까지 기다려야 하므로 응답시간이 상당히 증가하게 되는 문제점이 생긴다. 이러한 문제점을 줄이기 위해서 PBA 는 더 높은 접근빈도를 가지는 질의어에 속한 데이터들을 먼저 우선 할당하여 해당 질의어에서의 데이터충돌을 없앴으로써 전체적으로 볼 때 데이터충돌의 발생 가능성을 줄일 수 있었다.

그러나 이러한 방식은 질의어의 접근빈도를 반영한 것이지 각 데이터의 접근빈도를 반영한 것이 아니다. 또한 모든 데이터가 한 방송 사이클 내에서 오직 한번만 할당되기 때문에 그림 1 과 같은 문제점이 생길 수 있다.

Query	Prob.	Items
Q1	0.4	9, 12, 13
Q2	0.3	2 18
Q3	0.2	2 12
Q4	0.1	2 3, 6

(그림 1) 질의어 프로파일 내에서 데이터의 중요도

그림 1 과 같은 질의어 프로파일 내에서 데이터 2 는 Q1 에 속한 9, 12, 13 번 데이터 보다 더 자주 요청되는 데이터이다. 그러나 Q1 의 접근빈도가 가장 높기 때문에 할당 시에 9, 12, 13 번 데이터보다 우선순위에 서로 밀리고 데이터의 높은 접근빈도 또한 반영되지 못했다.

우리는 본 논문에서 이러한 문제점을 해결하기 위하여 기존 방법에서처럼 할당 시에 질의어 할당 우선순위를 유지하여 데이터충돌의 영향[1]을 줄이면서, 질의어 프로파일 내에서 상대적으로 많이 요청되는 데이터는 한 방송 사이클 동안 다중 채널상에 여러 번 할당하여 해당 데이터를 획득하는데 요구되는 응답시간을 줄일 것이다. 그러나 이 논문에서 제안하는 방법은 각각의 데이터들이 적절한 반복횟수로 다중 채널상에 나타나기 때문에 새로운 데이터 탐색기법이 요구된다. 우리는 이 논문에서 튜닝타임을 최소로 하는 인덱싱기법 또한 제시할 것이다.

이 논문의 남은 구성은 다음과 같다. 2 장에서는 상호관련 데이터를 다중 무선방송채널상에 효율적으로 할당하는 새로운 방법에 대하여 살펴볼 것이고 3 장에서는 다중 채널에 할당된 데이터들을 찾을 수 있는 인덱싱 기법에 대하여 설명하겠다. 마지막으로 4 장에서는 본 논문의 결론과 향후 연구 방향을 기술할 것이다.

2. 상호관련 데이터들을 위한 효율적인 데이터할당

이번 장에서는 각 데이터의 요청빈도에 따른 적절한 할당횟수를 구하고 다중 채널상에 효율적으로 할당하는 방법에 대하여 설명하겠다.

2.1 데이터 할당 횟수를 구하는 방법

우리는 클라이언트로부터 요청되는 데이터의 요청 빈도에 따라 각 데이터 마다 다중 무선채널상에 할당되는 횟수를 정할 것이다.

Query	Prob.	Items
Q1	0.310	d9, d12, d18, d19
Q2	0.178	d10, d11, d13, d17, d18
Q3	0.128	d3, d14, d16, d17, d18
Q4	0.102	d2, d5, d6, d11, d14, d15
Q5	0.085	d2, d13, d17, d20
Q6	0.074	d1, d11, d17, d20
Q7	0.065	d3, d4, d15, d19
Q8	0.058	d7, d8, d14, d18

(그림 2) 질의어 프로파일예제

그림 2 와 같이 클라이언트들로부터 받은 질의어 프로파일 이 있다고 하자. 이때 Q1 은 요청될 확률이 0.310 이다. 따라서 데이터 9, 12, 18, 19 는 모두 Q1 에

의해 0.310 의 확률만큼 요청된다고 볼 수 있다. 질의어 Q1 의 0.310 이라는 확률의 의미는 질의어가 1000 번 요구 되어질 때 Q1 이 310 번 요청된다는 것을 의미한다. 이런 식으로 질의어 1000 번 요청 시 Q2 는 대략 178 번, Q3 는 128 번 요청될 것이다. 이러한 방식을 통하면 각 데이터는 자신을 포함하고 있는 질의어의 요청횟수를 모두 더한 값만큼 요청될 것이다. 즉, 18 번 데이터는 Q1, Q2, Q3 그리고 Q8 에 포함되어 있기 때문에 이들의 요청횟수의 합인 $310+178+128+58=674$ 번 만큼 요청될 것이다. 이렇게 만들어진 요청횟수를 가지고 우리는 각 데이터에 대한 적절한 할당횟수를 정할 것이다.

Data	18	17	19	11	9	12	14	13	3	2
회수	674	465	375	354	310	310	288	263	193	187

Data	10	15	20	16	5	6	1	4	7	8
회수	178	167	159	128	102	102	74	65	58	58

(그림 3) 질의어 요청에 따른 각 데이터에 대한 상대적인 요청빈도수

그림 3 을 보면 데이터 18 번이 674 번 요청될 때 데이터 8 번은 58 번 요청되는 것을 알 수 있다. 따라서 8 번 데이터의 요청횟수인 58 로 18 번의 요청횟수 674 를 나누면 상대적인 할당횟수를 구할 수 있다. 그러나 이러한 방식은 가장 적은 요청횟수가 가장 큰 요청횟수에 비하여 상대적으로 상당히 작다면 전체 할당회수의 크기가 커져 전체 방송사이클이 커지는 단점이 있다. 이러한 문제를 해결하기 위하여 각 데이터의 요청횟수를 정규화하여 타당한 할당횟수를 구하는 방법이 필요하다. 그림 3 의 예제에서는 1, 4, 7, 8 번 데이터는 비교적 비슷한 요청횟수를 갖는다. 따라서 1 번 데이터의 요청횟수 74 보다 작은 요청횟수에는 할당횟수를 한번으로 하여 모든 데이터의 상대적인 할당횟수를 구하였다. 이러한 방식을 통하여 데이터 18 번은 그림 4 와 같이 할당횟수가 9 가 될 것이고, 17 번 데이터는 6 회 할당횟수를 가질 것이다.

Data	18	17	19	11	9	12	14	13	3	2
할당회수	9	6	5	5	4	4	4	4	3	3

Data	10	15	20	16	5	6	1	4	7	8
할당회수	2	2	2	2	1	1	1	1	1	1

(그림 4) 각 데이터의 할당회수

2.2 데이터 할당횟수를 이용하여 할당하는 방법

2.1 절에서 질의어 프로파일 에 속한 질의어의 평균 응답시간을 최소화하기 위한 각 데이터의 최적의 할당횟수를 정하였다. 이번 절에서는 이러한 각각의 데이터의 할당횟수를 이용하여 다중채널상에 어떻게 데이터를 할당할지를 설명하겠다. 먼저 모든 데이터의 할당횟수의 총합을 구한다. 이때 가장 작은 할당횟수는 1 이다. 즉 모든 데이터가 하나의 방송사이클 동안 최소 한번은 할당 되어진다.

데이터를 채널에 할당하는 순서는 질의어 프로파일

상에서 질의어 요청이 높은 첫 번째 질의어의 데이터들부터 할당된다. 할당순서는 PBA[1]에서 제시하는 방법과 같다. 할당방법을 설명하기에 앞서 다음과 같은 정의가 요구된다.

정의 1. *Cycle_length* 는 각 데이터의 할당횟수의 총합을 채널수로 나눈 값이다. 즉 방송사이클의 크기를 말한다.

정의 2. *Diff(D_i)*는 방송될 데이터 집합의 *i* 번째 데이터 *D_i* 가 채널상에 주기적으로 나타나는 일정한 간격이다. *Diff(D_i)*는 *Cycle_length / t_i* 이다. 여기서 *t_i* 는 *D_i* 의 할당횟수이다.

데이터 *i* 를 지정된 위치에 처음 할당하고 직전 할당된 위치로부터 *Diff(D_i)*크기만큼 떨어진 위치의 다음 채널에 할당하는 것을 할당회수 *t_i* 만큼 반복한다. 여러 할당회수를 가진 데이터는 일정한 간격으로 채널들에 할당됨으로써 획득하는데 요구되는 시간을 최소화 시킬 수 있다. 만약 할당되어야 할 위치의 모든 채널에 이미 다른 데이터가 할당되어 있다면 오른쪽으로 1 만큼 이동하여 할당한다. 그림 5 는 이러한 할당 알고리즘을 보여준다.

```

Input : 질의어 프로파일 Q, 데이터채널의 개수 K
Output : 데이터가 할당된 다중 방송사이클

Begin
diff[N] : N 개의 데이터가 diff[i-1]의 크기만큼의 간격마다 할당된다.
current_offset : 현재 데이터를 할당할 위치
allocation count[N] : data 1 ~ data N 까지 각 데이터가 한 방송사이클 내에 다중채널상에 할당되는 횟수

for( each data i in query profile ){
    current_offset = 다중채널상에서 첫 번째 빈 위치
    for( allocation count[i] of data i ){
        if( current_offset 이 빈 slot 이라면 ){
            current_offset 에 data i 를 할당한다.
            current_offset 을 다음 채널로 이동시키고, diff[i-1]만큼 offset 을 right shift 시킨다.
        } else {
            // 현재 위치에 데이터가 할당되어 있다면
            if( 다른 채널의 current_offset 과 같은 오프셋에 빈 slot 이 있다면 ){
                current_offset = 찾은 빈 slot
                current_offset 에 data i 를 할당한다
                current_offset 을 다음 채널로 이동시키고, diff[i-1]만큼 offset 을 right shift 시킨다.
            }
        }
        else{
            if( current_offset 의 1 오른쪽 slot 이 비어있다면 )
                해당 slot 에 data i 를 할당한다
                current_offset 을 다음채널로 이동시키고 diff[i-1]만큼 offset 을 right shift 시킨다.
            else if( current_offset 의 1 왼쪽 slot 이 비어있다면 )
                해당 slot 에 data i 를 할당한다
                current_offset 을 다음채널로 이동시키고 diff[i-1]만큼 offset 을 right shift 시킨다.
        }
    }
}
end
    
```

(그림 5) 다중 채널에 데이터를 할당하는 알고리즘

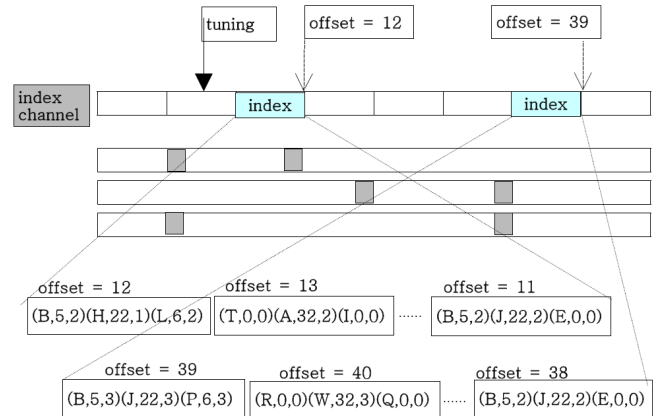
채널에 데이터를 할당하는 예제를 보이겠다. 그림 4 에서 모든 데이터의 할당횟수의 총합은 61 이고 데이터채널의 수가 4 라고 하면 방송사이클의 크기는 16 이 된다. 이때 남은 세 개의 버킷에는 임의의 데이터를 한번 더 할당한다. 할당 시에는 먼저 그림 2 의 질의어 프로파일에서 가장 높은 요청확률을 가지고 있는 Q1 에 속한 데이터 9, 12, 18, 19 번부터 채널에 할당한다. 9 번 데이터는 16/4=4 의 간격마다 할당횟수 4 회만큼 주기적으로 채널을 순환하면서 할당 되어지고, 12 번 데이터도 역시 16/4=4 의 간격마다 4 회를 채널을 순환하면서 할당된다. 그림 5 의 알고리즘을 사용하여 모든 데이터들을 채널에 할당하면 그림 6 과 같이 하나의 방송사이클을 구성하게 된다.

Ch1	9	12	18	19	10	11	13	17	18	3	14	16	18	17	2	5
Ch2	17	11	13	16	9	12	17	19	6	11	18	2	10	15	18	3
Ch3	18	20	14	17	18	3	14	15	9	12	13	19	1	11	14	4
Ch4	7	18	19	8	11	2	18	6	5	20	17	1	9	12	13	19

(그림 6) 데이터를 채널에 할당한 모습

3. 인덱싱 기법

2 장에서 우리는 질의어 기반의 방송기법에서 각 개별 데이터의 요청빈도를 반영하여 다중 채널상에 할당하는 방법을 알아보았다. 그러나 이 논문에서 제시한 할당방법은 기존의 논문과는 달리 각 데이터들이 서로 다른 수의 할당횟수를 가지게 되고 채널에 독립적으로 할당되기 때문에 적절한 인덱싱 기법이 없다면 튜닝시간이 매우 커질 것이다. 이번 장에서는 이 문제를 해결하기 위하여 N 개의 데이터로 구성된 질의어를 O(인덱스버킷의 크기 + N)의 시간만 채널을 들음으로써 원하는 데이터들을 모두 획득할 수 있는 새로운 인덱싱 기법을 제시할 것이다. 여기서 인덱스 버킷의 크기는 모든 데이터채널의 각 버킷마다 방송되는 데이터들의 관한 정보(데이터의 ID, 다음 번 방송될 때까지 남은 시간 그리고 채널번호)의 총합이다. 우선 채널은 K 개의 데이터 채널과 더불어 한 개의 인덱스 채널이 요구된다. 클라이언트는 처음 채널에 접속 시에 인덱스 채널을 검색하게 된다.

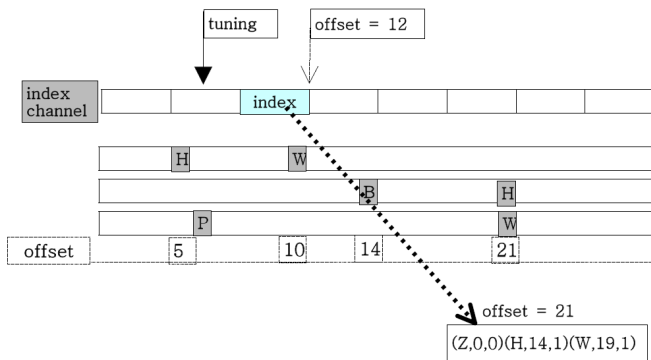


(그림 7) 인덱싱 기법

그림 4 의 각 데이터의 할당횟수를 이용하여 다중 하나의 인덱스 버킷안에는 해당 버킷이 끝나는 시점

부터 시작하여 버킷의 끝 시점까지 순환하면서 모든 데이터 채널상에 나타나는 데이터에 대한 정보가 나타난다. 그림 7에서 $offset = 12$ 의 (B,5,2)(H,22,1)(L,6,2)가 의미하는 것은 12 라는 타임오프셋에 세 개의 데이터채널에 나타나는 데이터 정보가 각각 (B,5,2), (H,22,1), (L,6,2)라는 것이다. 여기서 첫 번째 인자는 데이터 ID 를 의미하고, 두 번째 인자는 해당 데이터가 앞으로 다시 나오기까지 남은 타임오프셋을 의미한다. 마지막 세 번째 인자는 다음 번에 해당 데이터가 나오는 채널 번호를 의미한다. 여기서 두 번째와 세 번째 인자는 질의어에 속한 데이터들이 여러 채널상에 동시에 나왔을 경우 두 번째 인자를 보고 앞으로 더 오랜 시간 후에 나타날 데이터를 먼저 획득하기 위한 정보이다. 이러한 인덱스정보를 이용하여 클라이언트는 자신의 질의어에 속한 데이터가 나오는 시점과 채널번호에 대한 정보를 유지하면서 튜닝시간을 최소화하여 데이터를 얻을 수 있게 된다.

Query = { B, H, P, W }



(그림 8) 인덱싱 기법의 예제

그림 8의 예제를 보자. 클라이언트는 질의어={ B, H, P, W }를 획득하기 위하여 인덱스 채널에 튜닝한 후 인덱스로부터 $offset(14)$ 에 데이터 B를 획득할 수 있다는 것을 알 수 있다. 그리고 $offset(21)$ 에 데이터 H와 데이터 W가 동시에 방송된다는 것을 알게 되고 두 번째 인자를 통하여 다음에 방송되기까지 남은 시간이 큰 데이터 W를 1번 채널에서 획득해야 한다는 것을 알게 된다. 그리고 다음 방송사이클의 $offset(5)$ 과 $offset(6)$ 에서 데이터 H와 데이터 P를 각각 획득할 수 있다는 것을 알게 된다.

4. 결론

본 논문에서는 기존 연구에서는 시도되지 않았던 무선 다중채널환경하에서 상호관련 데이터들의 요청 빈도에 따른 적절한 할당횟수를 구하였다. 또한 이러한 할당횟수를 기반으로 각 데이터는 일정한 간격을 유지하면서 다중채널상에 할당하여 질의어에 포함된 모든 데이터를 획득하는데 요구되는 평균응답시간을 최소화 하는 방법을 제안하였다. 그리고 우리는 서로 다른 수의 할당횟수를 가지고 채널에 독립적으로 할당되는 데이터들을 작은 튜닝시간에 획득할 수 있는 새로운 인덱싱 기법을 제안하였다. 향후, 우리는 데이

터의 할당횟수를 효율적으로 정규화는 방법에 대한 연구와 본 논문에서 제안한 방법에 대한 구체적인 성능평가를 할 계획으로 있다.

참고문헌

- [1] Hao-Ping Hung, Jen-Wei Huang, Jung-Long Huang, Ming-Syan Chen, Scheduling Dependent Items in Data Broadcasting Environments, ACM SAC 2006.
- [2] Y. D. Chung and M. H. Kim. Effective Data Placement for Wireless Broadcast. Distributed and Parallel Databases, 9(2), 2001.
- [3] S. Acharya, R. Alonso, M. Franklin, and S. Zdonik, "Broadcast Disks: Data Management For Asymmetric Communications Environments," Proc. ACM SIGMOD Conf. Management of Data, pp. 199-210, May 1995.
- [4] T. Imielinski, S. Viswanathan, and B. Badrinath, "Data on Air: Organization and Access", IEEE Trans. Knowledge and Data Eng., vol. 9, no. 3, pp. 353-372, May 1997.
- [5] J. -L. Huang and M. -S. Chen. Dependent Data Broadcasting for Unordered Queries in Multiple Channel Mobile Environment. IEEE Trans. on Knowledge and Data Engineering, 16(6), Jun. 2004.
- [6] Narayanan Shivakumar, and Suresh Venkatasubramanian, "Efficient indexing for broadcast based wireless systems", Mobile Networks and Applications, vol. 1, no. 4, pp. 433-446, December 1996
- [7] G.Lee and S. C. Lo. Broadcast Data Allocation for Efficient Access of Multiple Data Items in Mobile Environments. Mobile Networks and Applications, 8, 2003.