이동 컴퓨터에서 사용되는 전속기는 벤더의 제약 và 이동통신망의 높은 장애율로 사용에 불편을 느낀다. 여러 mobile client 는 공통 관심이 있는 데이터를 서버로부터 받는 방법으로 broadcasting 을 많이 쓰는데, 이때 indexing 기법을 이용해 클라이언트는 원하는 데이터를 filtering 해서 수신함으로서 논리지는 효율적 사용을 기울 수 있다. index 를 사용하기 위해서 원하는 데이터 접근(access) 시간을 줄이고 우선 통신망의 장애에 따른 성능저하를 줄일 수 있다. 본 논문에서는 K-level indexing 기법을 위한 장애에 따른 최적의 중복회수를 구하고 데이터 수신시 데이터 접근 시간과 튜닝(tuning) 시간을 구한다.

1. 서론

데이터 브로드캐스트(broadcast)는 서버가 클라이언트에 의해 필요로 하는 데이터를 스케줄(schedule)에서 전송 채널 (communication channel)에 전송하는 것으로 여러 클라이언트 들이 원하는 공통의 데이터를 효과적으로 전송하는 방법이다. 그 때 클라이언트는 전송 채널에 들어오는 데이터를 filtering 함으로써 원하는 데이터만 읽어 들인다.

클라이언트는 휴면 모드(doze mode)와 동작 모드(active mode) 에서 수행되며(1), 휴면 모드로 채널의 전력은 사용하여 활성화하는 상태이고, 보내는 채널로 어떤 데이터가 들어오는지 확인하지 않는다. 동작 모드는 보낼 채널로 들어오는 데이터를 무시하지 않게 하기 위해 processor 가 작동하여 데이터를 사용하는 상태를 말한다. 현재의 클라이언트는 원하는 데이터를 보내면서 동작 모드이고 그렇지 않으면 휴면 모드가 있는데 이런 동작을 선택적 데이터 수신(selective tuning)이라고 한다. 그리고 전력 (energy)의 효율을 있으므로 가능하다. 선택적 데이터 수신이 가능하면 원하는 데이터의 위치를 알아야 한다. 데이터의 사이사이에 인덱스(index)를 둔으로써 원하는 데이터를 쉽게 찾을 수 있다. 기존의 인덱스 기법(1)에서는 데이터 접근(data access) 시간 혜증이 높아 인덱스를 효과적으로 복잡하고 중복시키는 방법을 고려하였다. 본 논문에서는 기존 연구(2)를 바탕으로 우선 통신망의 장애를 고려한 k-level 인덱스 기법의 데이터 접근 시간과 튜닝(tuning) 시간을 분석하였고 최적의 인덱스 중복이수를 구하였다.

2. 인덱스 브로드캐스트(Index Broadcast)

다음은 본 논문의 기반인 설명을 index broadcast 기법(1)을 간단히 요약한 것이다.

2.1 브로드캐스트를 위한 Data 구성(1)

브로드캐스트에 의해 전송된 데이터는 서버에서 파일(file)의 형태로 존재하며 파일은 primary key 에 의해 구분되는 레코드(record)들로 구성된다. 파일의 크기와 내용은 각각 변한다. 서버는 다수의 클라이언트들에게 주기적으로 브로드캐스트하며, 클라이언트는 전송되는 데이터를 받아서 primary key 값에 의해 구분되는 레코드를 저장한다. 그러나, 파일에 대한 갱신(update)은 연속된 전송사이에 반영된다. filtering은 primary key 값의 simple pattern matching에 의해 이루어지며, 대부분의 시간을 휴먼 모드로 정해진다. 데이터 item은 계속 바뀌면서 서버는 클라이언트에게 지속적으로 이것을 알려주기 때문에, directory 및 data item name 별 브로드캐스트한다. 다양한 채널(channel)에서 많은 데이터를 전송되므로 클라이언트는 원하는 데이터만 선택하는 데 선택적 데이터 수신기의 기능이 필요하다. 서버는 data 와 인덱스(index)를 얻어본다.

다음은 본 논문에 사용되는 용어에 대해 간략히 설명한다.

- Access time : 브로드캐스트가 primary key 값으로 원하는 데이터를 찾아 저장하는 데까지의 평균 검색 시간을 말한다. Probe Wait 및 Broadcast Wait의 합으로 나타낼 수 있다. probe wait은 지연 시간을 갖고 index 를 읽는 데 걸리는 평균 시간을 말하며 broadcast wait은 최초의 index 를 읽고 원하는 데이터를 얻는데 걸리는 평균 시간을 말한다.

- Tuning time : 네트워크를 설정하기 위해 검색 시간, 즉 채널을 결정하기 위해 다음은 인덱스 및 데이터를 읽는 시간을 의미한다. 이 시간은 동일한 모바일 사용자 모드인 active mode로 한다.

- Bucket : 브로드캐스트의 가장 작은 논리적인 단위이다. bucket의 크기는 모두 같다. access time 및 tuning time은 bucket 의 수로 측정된다.

- Index Bucket : 필요한 데이터를 찾기 위한 offset 등의 bucket.


- Tuning Optimal : Index Bucket 을 브로드캐스트의 시작부분에 만든 때 나오는 것이라. 클라인트의 최적 Index에 Index을 읽고 primary key로 원하는 데이터의 위치를 알아내어 그때까지 휴먼 모드로 돌아온다.

원하는 데이터의 위치가 되면, 동작 모드로 전환되어 필요한 데이터를 읽는다. Tuning time은 index level의 수와 최적의 index 위치를 얻기 위한 데이터를 읽는 것과 원하는 데이터를 읽는 것의 합으로 구한다. probe wait는 (Data + Index)/2이고 broadcast wait는 (Data + Index)/2이다. 따라서 access time은 (Data + Index)이므로, 이것은 access time이 최대가 된다.
3. 무선 통신망의 장애를 고려한 Indexing Broadcast

이동 computing 환경에서는 무선 통신망에 장애가 발생할 가능성이 높다. 이에 따라 인덱스를 배치하고 어느 정도 중복되야 하는 것을 고려하는 것이 중요하다. 기본적으로 원하는 데이터를 얻기 위해 데이터가 손상되었을 경우, 다음번 주기의 데이터를 얻는다.

무선 통신망의 장애를 고려한 index broadcast protocol은 다음과 같다.
1. 현재의 bucket을 얻는다. 여기서 데이터가 손상되면, 다음 bucket을 얻는다.
2. 다음의 가장 가까운 index segment의 주소를 알려주는 offset을 얻는다.
3. doze mode로 들어가고 index segment에서 index를 얻는다.
4. index segment에서 primary root를 원하는 데이터의 위치를 결정한다. multiindex index 기법에는 인접한 probe가 있으면, 이때 데이터 위치가 선정되어, offset을 얻는다.
5. index key에 의해 구현한 데이터의 위치에서 index에서 downを 얻는다. 실패하면, offset을 얻는다.

위의 protocol을 이용할 때 access 시간과 tuning 시간을 구하면 다음과 같다. 무선 통신망 장애는 저수포분포가(평균 1/p마다 발생)로 발생한다고 가정하였으며 이 때 한 bucket의 index 또는 데이터가 손상된다고 가정하였다.

또한, 인덱스의 위치에 접근하는 데 필요한 access 시간과 tuning 시간을 분석한다. 사용되는 parameter는
- \( n \) : 한 bucket의 index 개수
- \( n \) : bucket tree의 level 수
- \( m \) : bucket tree의 fully balanced 할 때

\[
K = \left\lfloor \log_2 \frac{Data}{\text{Index}} \right\rfloor
\]

\( \text{Index} = \text{index bucket 의 크기} \)
\( \text{Data} = \text{한 브로드캐스트 주기의 bucket 의 수} \)
\( m = \text{index bucket 의 중복도} \)

그림 2에와 같이 Markov model로 나타낼 때, state 0은 첫번 사람이 데이터를 얻을 상태를 나타내며, state 1은 root index를 얻는 상태를 나타내고, state 2는 index를 얻는 상태를 나타내며, state 3은 Data를 얻는 상태를 나타내며, state 4는 최종 상태를 나타낸다. state 5은 다시 state 0부터 다시 시작한다. C_{\text{idx}}는 state a에서 state b로의 상태전이에 따른 cost를 나타낸다.

access 시간을 계산하기 위하여 상태전이에 따른 cost를 구하면,

\[
C_{\text{idx}} = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{(1-p)^i}
\]

\[
C_{\text{data}} = \frac{D}{2} \sum_{i=1}^{n} \frac{1}{(1-p)^i}
\]

\[
C_{\text{tuning}} = \frac{D}{2} \sum_{i=1}^{n} \frac{1}{(1-p)^i}
\]

4. 파라미터 분석

다음은 파라미터 변화에 따른 성능을 측정하였다. 여기서 각각의 인덱스는 각 데이터 버켓의 트리에 잘 분산되어 있는 상태를 가정하고 파라미터를 분석하였다. 여기서 D는 한 브로드캐스트 주기당 데이터 버켓의 크기를 나타내며, k는 인덱스 개수를 나타내며, n은 인덱스 버켓의 개수를 나타내며, p는 인덱스 버켓이 인덱스 할 수 있는 데이터 버켓의 크기를 나타낸.
다.
그림 3과 그림 4는 무선통신판 장애율에 따른 액세스 시간의 변화를 나타낸다. 무선통신판 장애율에 따른 액세스 시간의 변화를 보면, 장애율이 높아짐수록 액세스 시간이 증가한다. 그림 5, 6, 7은 인덱스 중복회사에 따른 액세스 시간의 변화를 나타낸다. 액세스 시간은 테이터 크기, 인덱스 레벨, 무선통신 장애율, 인덱스 중복회사에 따라 영향을 받음을 알 수 있다. 장애율에 따라 최적의 인덱스 중복회사가 존재함을 알 수 있다.

그림 7. 인덱스 중복회사에 따른 액세스시간의 변화. (D=1000, k=1, l=1, n=1000)

그림 8. 무선통신판장애율에 따른 튜닝시간의 변화.

그림 8은 무선통신판 장애율에 따른 튜닝시간의 변화를 나타낸다. 튜닝 시간은 무선통신판 장애율과 인덱스 레벨에영향을 받지 인덱스 중복회사에는 영향을 받지 않는다. 인덱스 레벨이 증가함에 따라 인덱스 액세스 시간이 단락성기 높기 때문에 장애율의 변화에 더욱 민감함을 알 수 있다.

5. 결론 및 향후 연구
무선통신망의 장애를 고려하여 인덱스 기법을 이용한 방송기법의 액세스 시간과 튜닝 시간을 분석하였고 액세스 시간을 최소화시키기 위해 최적의 인덱스 중복회사를 구하였다. 분석결과에 따르면, 장애율이 높을수록 인덱스 중복회사를 높이는 것이 효과적임을 알 수 있다. 추가 연구로 인덱스 범위가 분산되어 있을 경우 어떻게 배치하고 어떻게 중복시키야 효율적인지가 고려된다.

참고자료