

센서 네트워크에서 블룸 필터를 이용한 하이브리드 인-네트워크 조인 기법

(A Hybrid In-network Join Strategy using Bloom Filter in Sensor Network)

송 임 영 [†] 김 경 창 ^{**}
 (Im Young Song) (Kyung Chang Kim)

요약 본 논문에서는 센서 네트워크에서 효율적인 조인 처리를 수행하여 통신비용을 최소화할 수 있는 인-네트워크 조인 기법인 SBJ(Semi & Bloom Join)를 제안하였다. SBJ는 하이브리드 기법으로 배터리 소모를 줄이기 위하여 송, 수신하는 데이터의 크기를 줄이는 방법으로 블룸 필터(Bloom Filter)를 이용하여 데이터의 내용을 필터링하도록 하였다. 기존 데이터 대신 작은 크기의 필터링 된 값을 전송함으로써 배터리 소모를 줄이게 된다.

SBJ는 조인 처리의 이른 단계에서 조인 결과에 해당되지 않는 데이터를 제거하여 불필요한 통신비용을 감소하여 통신비용을 최소화하기 위한 네트워크 내에서 효율적인 조인 질의 수행 방법이다. 시뮬레이션을 통하여 기존의 조인 기법들과 비교한 결과 SBJ 기법은 블룸 필터를 이용하여 데이터를 필터링 함으로써 센서 노드 배터리 소모량을 크게 줄였으며, 통신 반경에 대한 통신비용도 훨씬 효과적이었다.

키워드 : 센서 네트워크, 하이브리드 조인 전략, 통신비용, synopsis 조인, semi 조인, bloom 조인, 블룸 필터

Abstract This paper proposes an in-network join strategy SBJ(Semi & Bloom Join), an efficient join strategy for sensor networks, that minimizes communication cost. SBJ is a hybrid join strategy that can reduce energy consumption by using a bloom filter to reduce the size of data that needs to be sent or received in sensor network. The key to reducing the communication cost in SBJ is to eliminate data not involved in the join result in the early stages of join processing. Through simulation, the paper shows that compared to other join strategies in sensor network, SBJ join strategy is more efficient in reducing the communication cost resulting in a significant reduction in battery consumption.

Key words : sensor network, hybrid join strategy, communication cost, synopsis join, semi join, bloom join, bloom filter

1. 서론

센서 노드들은 제한적인 배터리 에너지를 갖기 때문에

센서 네트워크에서 분산 질의를 구현하기 위해서는 배터리 에너지의 주요 소비원인 통신비용을 최소화해야 한다.

본 논문에서는 센서 네트워크에서 효율적인 조인 처리를 수행하여 통신비용을 최소화할 수 있는 인-네트워크 조인 기법인 SBJ(semi & bloom join)를 제안하였다.

제안하는 조인 기법은 배터리 소모를 줄이기 위하여 송, 수신하는 데이터의 크기를 줄이는 방법으로 블룸 필터를 이용하여 데이터의 내용을 필터링(filtering)하도록 하였다. 이 방법은, 조인 처리의 이른 단계에서 조인 결과에 해당되지 않는 데이터를 제거하여 불필요한 통신비용을 감소시킨다. 배터리 소모 에너지에 대하여 기존 조인 전략들과 다양한 비교 실험을 수행함으로써, 제안한 조인 전략의 우월한 성능을 입증하였다.

본 논문의 구성은 2장에서 관련 연구로 센서 네트워크

· 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업인(No.2009-0076875)

† 학생회원 : 홍익대학교 컴퓨터공학과
 songimyoung@gmail.com

** 종신회원 : 홍익대학교 컴퓨터공학과 교수
 kckim@hongik.ac.kr

논문접수 : 2009년 9월 2일

심사완료 : 2010년 5월 25일

Copyright©2010 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 데이터베이스 제37권 제3호(2010.6)

크에서의 조인 전략에 대하여 설명하고 3장에서는 SBJ의 세부사항을 설명한다. 4장에서는 실험을 통해서 제안하는 조인 전략과 기존의 전략을 비교하고 마지막으로 5장에서는 본 논문의 결론을 정리한다.

2. 관련 연구

초기 센서 네트워크 연구인 Cougar[1], TinyDB[2]에서는 데이터베이스의 관점에서 센서 네트워크를 위한 질의 기반 데이터 관리 시스템을 구축하였다. 이들 연구는 단일 질의에 최적화가 되어있기 때문에 조인연산에 대한 연구는 부족하다. 종합적인 정보를 얻고자 할 때, 특정 시간이나 위치를 기반으로 조인 연산을 수행하게 된다.

조인의 예를 들기 위해 동물원에 설치된 센서 네트워크를 고려하자. 각 센서 노드는 동물의 ID(AnimalID)를 감지하고, 동물의 감지된 시간의 timestamp를 기록한다고 가정하자. N_R 과 N_S 가 각 질의 지역 Region R과 Region S에 대한 센서 노드의 세트이며 테이블 R과 S는 각 N_R 과 N_S 에 저장된다. 어떤 동물이 T1과 T2 시간 사이에 Region R과 S에 동시에 나타났는지 알아보는 조인 질의는 다음과 같다.

```
SELECT R.animalID
```

```
FROM R* as R, R* as S
```

```
WHERE R.loc IN Region R and S.loc IN Region S
and R.timestamp IN TimeRange [T1, T2]
and S.timestamp IN TimeRange [T1, T2]
and R.animalID=S.animalID
```

(Q1)

위의 Q1을 처리하기 위해서는 Region R과 Region S로부터 T1과 T2 시간 사이에 측정된 센서 값들을 수집하고 속성 animalID에 대해 조인해야 한다.

조인 전략의 주요 성능 기준은 조인을 구현할 때 발생하는 총 통신비용으로 총 통신비용을 최소화하기 위해서는 조인을 수행하는 사이트를 중앙집중식 전략인 외부 데이터베이스 서버가 아닌 센서 네트워크 내의 노드들을 지정하여 인-네트워크[3] 전략을 사용한다.

센서 네트워크에서 기존의 일반적인 조인 전략은 조인 지역의 위치에 따라 Naive join, Sequential join, Centroid join 등으로 분류할 수 있다[4]. 이런 일반적인 조인 전략의 주요 문제점은 낮은 조인 선택도의 질의에 대한 통신 오버헤드이다. 이것은 N_R 과 N_S 의 조인 후보가 아닌 튜플들이 조인 지역으로 전송되는 것이다.

낮은 조인 선택도의 질의에 대한 통신 오버헤드를 해결하기 위하여 제안된 시놉시스(Synopsis) 조인 전략(이하 SNJ)[4]은 센싱된 데이터의 시놉시스를 사용하여 조인 결과와 관련이 없는 데이터를 제거하는 것이다. 각 센서는 자체 로컬 테이블에 대한 시놉시스를 생성한다.

시놉시스의 의미는 조인을 수행하기 위한 테이블의 요약으로 하나의 튜플은 튜플 ID, 속성 값과 속성 값의 count로 구성된다. 각 센서는 자체 로컬 테이블에 대한 시놉시스를 생성한다. 각 노드에서 전송된 시놉시스 조인을 수행하고 조인 결과를 질의 지역에 통보하여 조인 후보가 아닌 튜플들을 필터링하고 오직 조인 후보 튜플들만 최종 조인한다.

본 논문에서 제안하는 통신비용을 최소화하는 하이브리드 조인 전략에서 이용되는 기존의 분산 데이터베이스를 위한 대표적인 조인 알고리즘들은 Semi 조인[5], Bloom 조인[6] 그리고 PERF 조인[7]이 있다.

Semi 조인 (SxR) 전략은 테이블 R과 S의 조인 속성들에 대해 테이블 R을 프로젝션(projection)하여 새 테이블 P_R 을 얻는다. P_R 을 테이블 S가 있는 사이트로 이동시켜 테이블 P_R 과 테이블 S의 조인을 수행하여 새 테이블 S' 을 얻는다.

Bloom 조인은 Semi 조인과 달리 조인 속성 값 대신 블룸 필터를 전송한다. 목적은 모든 컬럼 키를 보내는 대신에 네트워크 I/O를 줄이기 위해 적은 비트 리스트를 보내는 것이다. 블룸 필터는 조인 속성 프로젝션과 같은 조인 정보를 비트 벡터를 이용하여 인코딩(encoding) 한다. 테이블 R과 S를 조인할 때 R의 조인 속성 값들이 해당 비트가 1로 세트되는 블룸 필터의 어떤 주소로 해시 된다. 해싱 후 0비트는 조인 속성 값이 없음을 의미한다. Bloom 조인의 단점으로는 일치하지 않는 튜플이 포함되는 오탐(false positive)이 발생될 수 있다는 것이다.

PERF 조인은 양방향 Semi 조인의 백워드(backward) 단계의 비용을 최소화 하도록 설계되었다. [8]은 통신비용을 감소하기 위하여 데이터의 크기를 줄이기 위해 비트 벡터(Bit-vector)를 이용한다.

3. SBJ(Semi & Bloom Join) 전략

이번 장에서는 본 논문에서 제안하는 SBJ의 소개 및 세부사항을 설명한다.

SBJ 조인 알고리즘을 기술하는데 쓰이는 용어들은 다음과 같다. 이 용어들을 논문 전반에 걸쳐 사용하도록 한다.

- 지역 Region R과 Region S의 센서들의 집합은 각각 N_R 과 N_S 이며 각 지역에서 센싱된 데이터는 네트워크에 저장된 두 개의 분산 테이블 R과 S이다.
- P_R : R의 세미 테이블이다.
- $P_{R'}$: R의 세미 테이블 P_R 에서 최종조인에 참여하는 속성만 가지는 세미 테이블이다.
- $P_{S'}$: S에서 최종조인에 참여하는 속성만 가지는 세미 테이블이다.

- c_h : 예비 조인 노드이다.
- N_F : 최종 조인 노드들의 집합이다.
- R^v : 조인 속성 v 에 대한 R 의 조인 후보 튜플 집합이다.
- S^v : 조인 속성 v 에 대한 S 의 조인 후보 튜플 집합이다.

그림 1은 SBJ 전략의 전체 흐름을 나타내고 있다. 본 논문에서는 N_R 의 조인 속성 값으로 bloom 필터를 생성하는 것으로 조인 전략 단계를 설명한다.

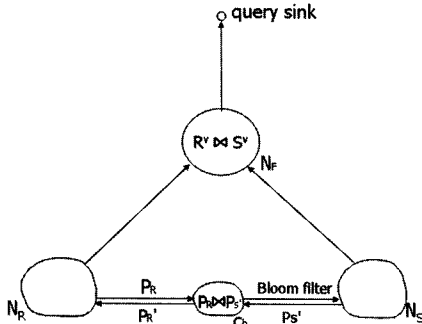


그림 1 SBJ 조인 전략

SBJ 조인 전략 단계는 다음과 같다:

1. 예비 조인

- 1.1 세미 테이블 P_R 을 예비 조인 노드 c_h 로 전송한다.
- 1.2 P_R 을 이용하여 해시 기반 bloom 필터 $H_h(P_R)$ 를 생성 후 N_S 에 뿌려진다.
- 1.3 P_S^v 를 생성하여 c_h 로 전송하고 P_R 과 P_S^v 조인 후 P_R^v 를 얻어 N_R 에 뿌린다.

2. 최종 조인

- 2.1 두 질의 지역의 조인 후보 튜플들을 N_F 로 전송하여 최종 조인을 한다.
- 2.2 조인 결과를 쿼리 싱크로 전송한다.

위와 같은 SBJ 조인 전략 단계의 세부사항을 이어지는 절들에서 구체적으로 설명한다.

3.1 예비 조인 수행

쿼리 싱크에서 N_R, N_S 의 중심에 해당하는 지역을 계산하여 그 지역에 가까운 노드를 두 지역의 중심 노드 c_r 과 c_s 로 선택한다. 그리고 generalized fermat's point[9]를 이용하여 c_r, c_s 와 쿼리 싱크의 중심점에 해당하는 c_t 를 구한다. 쿼리 싱크는 질의 Q와 c_t 에 대한 정보를 라우팅 프로토콜 GPSR[10]을 이용하여 c_r, c_s 에 전송한다. c_r 는 N_R 과 N_S 의 조인 후보 튜플들을 조인 속성 값 v 에 해당하는 최종 조인 노드로 전송하는 라우팅 노드로 이용될 노드이다. c_r 과 c_s 는 라우팅 트리를 형성하며 이웃 노드들에게 질의 Q와 c_t 의 정보를 뿌린다.

질의 Q를 전송 받으면 N_R 과 N_S 의 각 노드는 로컬

세미 테이블 $P(R_i)$ 와 $P(S_j)$ 를 생성한다. N_R 의 각 노드들이 $P(R_i)$ 를 c_r 로 전송하면, c_r 은 속성 값 중복을 제거하고 P_R 을 생성하여 c_h 로 전송한다.

예비 조인 노드 c_h 는 N_R 에서 전송되는 P_R 과 N_S 에서 전송되는 자격이 있는 조인 속성 값 세미 테이블 P_S^v 을 조인하여 N_R 에 자격이 있는 조인 속성 값을 통보해주는 노드이다. c_h 는 N_R 과 N_S 에서 송, 수신하는 데이터의 특성을 고려하여 통신비용이 최소가 되는 노드를 선택한다. 세미 테이블을 전송하는 것보다 bloom 필터를 전송하는 것이 더 통신비용 효율적이므로 본 논문에서는 그림 2와 같이 통신비용을 줄이기 위해 예비 조인 노드 c_h 를 N_S 에서 가장 가까운 N_R 에 인접되어 있는 노드들 중 c_r 과 c_s 를 연결한 선이 지나는 위치에 존재하는 노드를 c_h 로 선택하였다. 예비 조인 노드 c_h 가 선택되면 c_r 에 통보하고 c_r 은 P_R 을 c_h 에 전송한다.

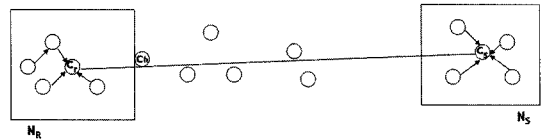


그림 2 예비 조인 노드 c_h 선택

P_R 로 bloom 필터 $H_h(P_R)$ 를 생성하는 이유는 N_R 의 조인 속성 값으로 N_S 의 자격이 없는 조인 속성 값을 필터링하기 위한 목적이다. P_R 을 전송하는 것 보다 적은 비트 리스트로 데이터를 해싱하여 표현할 수 있는 bloom 필터가 더 통신비용 효율적이다.

c_h 에 전송된 P_R 로 $H_h(P_R)$ 를 생성하여 N_S 에 bloom 필터를 뿌린다. 질의 Q를 만족하는 $P(S_j)$ 를 해싱한 후 $H_h(P_R)$ 와 비교하여 각 노드의 자격이 있는 조인 속성 값 세미 테이블 $P(S_j^v)$ 을 N_S 의 중심 노드 c_s 로 전송한다. N_S 의 각 노드들이 $P(S_j^v)$ 를 c_s 로 전송하면 c_s 는 속성 값의 중복을 제거하고 자격이 있는 조인 속성 값 세미 테이블 P_S^v 을 생성하여 c_h 로 전송한다.

c_h 에서 P_R 과 P_S^v 의 조인 결과인 P_R^v 을 N_R 에 뿌려 각 노드의 속성 값과 P_R 을 비교하여 조인 후보 튜플들을 선택한다.

3.2 최종 조인 수행

최종 조인 지역으로 전송될 조인 후보 튜플들이 선택 되면 c_t 로 전송한다.

테이블 R과 S는 각각 조인 속성 값의 범위를 갖는 몇 개의 독립적인 서브 테이블(sub-table)들로 분할될 수 있다. N_F 의 각 노드는 같은 조인 값 범위 v 를 갖는 두 개의 서브 테이블 R^v 와 S^v 를 조인한다. 같은 조인 속성 값을 갖는 튜플들은 항상 같은 조인 노드에서 조인된다. 따라서 테이블 S의 튜플들을 조인 노드 집합인

N_F 에 뿌리는 것을 막을 수 있다. N_F 는 N_R 과 N_S 의 조인 후보 튜플들을 계산하여 결정한다. c_r 는 전송되는 튜플들을 조인 속성 값 v 를 기반으로 $|N_F|$ 그리드로 해시하여 해당 노드로 튜플들을 전송한다. 하나의 센서 노드에 대해 조인 후보 튜플들 $C(R_i)$ 는 $|N_F|$ 개로 구성된다.

N_R 과 N_S 로부터 튜플들을 모두 전송 받으면 조인 수행 후 결과를 쿼리 싱크로 전송한다.

SNJ는 센싱된 튜플들의 시놉시스를 사용하여 조인에 참여하지 않는 튜플들을 필터링하는 조인 전략이다. 본 논문에서 제안하는 SBJ는 다음과 같은 특성으로 SNJ보다 통신비용 효율적이다.

- ① 블룸필터로 자격이 없는 조인 속성 값을 필터링하므로 세미 테이블 P_S 를 예비 조인 노드로 전송할 필요가 없다.
- ② 예비 조인 수행 후 N_R 에만 조인 결과를 통보한다.
- ③ N_S 의 시놉시스 대신에 자격이 있는 조인 속성 값을 예비 조인 노드로 전송한다.

그러나, 적은 비트 리스트를 이용하여 속성 값을 표현할 수 있는 블룸필터는 일치하지 않는 튜플이 포함되는 오탐이 발생될 수 있다. 하지만 실험 결과 블룸필터로 인해 오탐이 발생하더라도 블룸필터를 사용하는 것이 보다 더 통신비용 효율적이었다.

4. 실험

제안하는 SBJ 조인 전략을 테스트하고 객관적인 효율성을 검증하기 위해 기존의 SNJ와 최적의 조인 지역을 선택하여 조인을 수행하는 Centroid 조인 전략에 대해서 통신비용을 계산하여 비교하였다. 센서 네트워크에서 통신비용의 정의는 절의 배포부터 최종 조인 결과 수집까지의 메시지 전송이다. 본 논문에서의 통신비용은 전체 메시지 전송에 소모된 배터리 소모 에너지이다.

4.1 실험 환경

실험에 사용된 시스템은 CPU Intel(R) Core(TM)2 E8400 3.00 GHz이고 RAM은 2GB를 사용했다. 운영체제는 Windows XP, 개발 언어 및 플랫폼으로는 JAVA를 사용하였다.

본 논문의 실험환경은 객관성을 위해 SNJ[4] 실험 환경과 같게 시뮬레이션 하였다. 가로 세로 100×100 그리드에 균등하게 10,000개의 센서 노드 생성하였다. 각 그리드의 중심에 하나의 센서 노드가 위치하며 쿼리 싱크는 중앙 위에 위치한다. N_R 과 N_S 는 오른쪽 아래와 왼쪽 아래에 위치해 있고 각각 400개의 센서 노드들이 존재한다. 테이블 R과 S는 300가지의 조인 속성 값을 가지고 튜플을 랜덤으로 생성하여 각각 1000개의 튜플들로 구성되었으며 센서 노드들에 균등하게 분포되어 있다. 소스 노드에서 최종 노드로 하나의 메시지를 전송하는

데 요구되는 홉 수는 두 센서 노드 사이의 거리와 통신 반경을 사용한다. 메시지 전송 실패는 없다고 가정하였다. 튜플 사이즈는 40바이트이고 SNJ에 사용한 시놉시스는 10바이트이다. 센서 노드 메모리 용량은 250 × 40 바이트, 통신 반경은 4 유닛(units)이다.

제안하는 SBJ의 블룸필터는 해싱함수를 늘리는 방법으로 오탐 발생확률을 조절하였다[6]. 함수를 1회부터 8회까지 적용하여 SBJ의 오탐 발생확률을 계산하니 5회 적용하였을 때 오탐 발생확률이 가장 낮았다. SNJ와 비교하여 SBJ의 오탐 발생확률은 조인 선택도 0.0001부터 0.01까지 N_R 과 N_S 에서 N_F 로 전송되는 두 조인 전략의 전송 바이트 차이를 계산하였다. 실험에서 블룸필터는 P_R 을 32비트 시스템에서 잘 작동되도록 설계된 MD5를 5번 적용하여 생성하였다. 본 실험에서는 N_R 과 N_S 의 센싱 튜플 수가 동일하기 때문에 N_S 의 세미 테이블로 블룸필터를 생성하여도 성능에는 크게 영향을 주지 않는다. 센싱 튜플의 수가 더 많은 테이블의 세미 테이블을 이용하여 블룸필터를 생성한다면 통신비용을 더 줄일 수 있을 것이다.

4.2 실험 결과 및 분석

설명한 실험 환경을 바탕으로 성능평가 결과를 설명한다. 모든 실험의 최종 측정값은 동일한 파라미터 값들에 대해 10번씩 실험한 결과 값들의 평균으로 정하였다.

데이터를 전송하기 위하여 필요로 하는 노드의 배터리 소모 에너지는 [11]에서 사용한 값을 동일하게 적용하였다. 모든 노드들의 초기 배터리 에너지를 2J로 동일하다고 가정하였으며 거리 d 에 대해 1비트의 데이터를 송수신할 때 소모하는 에너지는 다음과 같다.

$$E_{TX}(K, d) = (E_{elec} \times K) + (E_{amp} \times K \times d^2) \quad (1)$$

$$E_{RX}(K) = E_{elec} \times K \quad (2)$$

노드별 소모 에너지를 산출하기 위해 필요한 각 요소들의 정의 및 해당 값들은 표 1과 같다.

먼저 조인 선택도를 0.0001부터 0.01까지 변화시켜 세 조인 전략의 배터리 소모 에너지를 구하여 결과를 그림 3에 정리하였다. 조인 후보가 아닌 튜플들도 최종 조인 노드로 전송되는 Centroid의 배터리 소모 에너지가 가장 많다. SBJ는 모든 조인 선택도에 대해 SNJ보다 적은 배터리 소모 에너지를 보였다. 조인 선택도가 0.0001일 때 SBJ는 SNJ 보다 소모 에너지가 49% 정도 감소

표 1 노드별 에너지 소모 요소

	정의	값
E_{elec}	송신회로 소모 에너지	50nJ/bit
E_{amp}	송신 앰프 소모 에너지	100pJ/bit/m ²
d	노드 전송 반경	4유닛
K	데이터 크기	

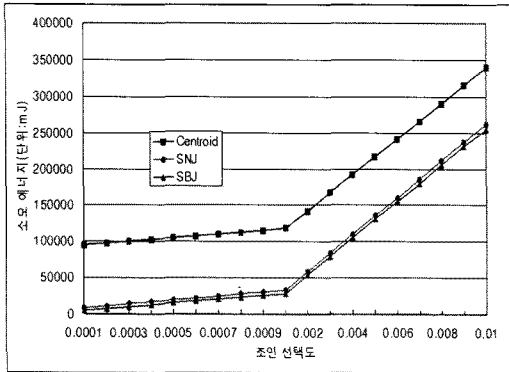


그림 3 조인 선택도에 대한 배터리 소모 에너지

하였다. 이는 4090mJ 만큼 덜 소비된 것으로 406876 바이트를 더 전송할 수 있는 배터리 에너지이다. SBJ는 조인에 필요한 최소한의 데이터 이동과 적은 비트 리스트로 데이터를 표현할 수 있는 블룸필터를 이용함으로써 시뮬시스를 조인하여 최종 조인 지역으로 전송되는 튜플 수를 줄이는 것 보다 더 통신비용 효율적이다. 이때 오탐 발생으로 N_s 에서 최종 조인 지역으로 1240바이트만큼 통신비용이 더 발생하였지만 블룸필터를 사용하여 데이터를 전송함으로써 13450바이트 정도의 통신비용을 줄일 수 있었다.

네트워크 밀도를 다양하게 하여 조인 전략의 확장성을 확인한 결과는 그림 4이다. 조인 선택도는 0.0001이다. 통신 반경이 커짐에 따라 단위 데이터를 전송하기 위하여 필요로 하는 에너지는 증대될 것이므로 배터리 소모 에너지는 커진다. SBJ는 SNJ 보다 4 유닛일 때 49% 정도 덜 소모하고 16 유닛일 때 55% 정도 덜 소모한다. 세미 테이블과 블룸필터 보다 전송 바이트가 더 큰 시뮬시스를 통신반경 내 노드들에게 전송하므로 통신반경이 커질수록 배터리 소모 에너지는 커진다.

다음으로 첫 번째 메시지가 쿼리 싱크에 도착하는데

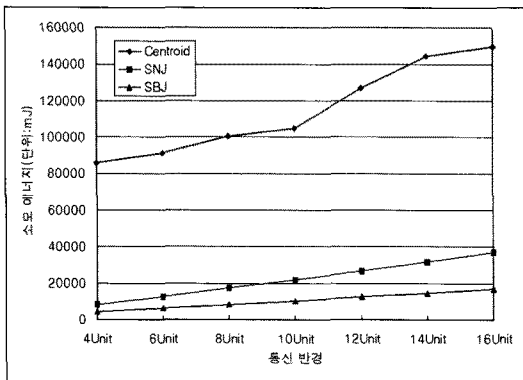


그림 4 통신 반경에 대한 배터리 소모 에너지

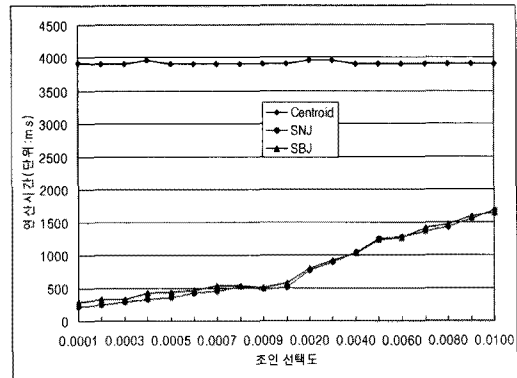


그림 5 쿼리 싱크로 전송되는 첫 번째 메시지 도착까지의 연산시간

드는 연산 시간을 실험하였다. 그림 5는 조인 선택도 변화에 대한 연산 시간 실험 결과이다. 최종 조인은 모든 조인 후보 튜플이 N_F 에 전송되었을 때 수행되므로 조인 선택도에 상관없이 모든 튜플을 최종 조인 지역으로 전송하는 Centroid의 연산 시간이 가장 오래 걸렸다. SBJ가 블룸필터를 생성하여 질의 지역의 노드들과 비교하는데 드는 연산시간으로 인해 조인 선택도 0.0001일 때 SNJ 보다 41ms만큼의 연산시간이 더 필요했지만 조인 선택도나 질의 지역 변화에 대해 두 조인 전략 간에 연산시간의 큰 차이는 없다.

5. 결론

본 논문에서는 제안하는 SBJ는 하이브리드 기법으로 배터리 소모를 줄이기 위하여 조인 시 송, 수신하는 데이터의 크기를 줄이는 방법으로 블룸 필터를 이용하여 데이터의 내용을 필터링하도록 하였다. 이 방법은 조인 처리의 이른 단계에서 조인 결과에 해당되지 않는 데이터를 제거하여 불필요한 통신비용을 감소하여 통신비용을 최소화하기 위한 네트워크 내에서 효율적인 조인 질의 수행 방법이다.

Centroid와, SNJ 등과 성능평가 결과 최소한의 데이터 이동과 블룸 필터를 이용하여 조인에 참여하지 않는 튜플을 제거시키는 SBJ의 배터리 소모 에너지가 가장 적었다. SBJ는 오탐 발생으로 인해 최종 조인 노드들의 전송 튜플 수는 SNJ에 비해 SBJ가 2% 정도 많았지만 통신비용을 감소시켜 센서 노드 배터리 소모량은 낮은 조인 선택도에서 50% 정도 줄일 수 있었으며 통신반경이 커지면 SNJ보다 상대적으로 더 적은 배터리 에너지를 소모하였다.

참고 문헌

[1] Y. Yao, and J. Gehrke, "The Cougar Approach to

- In-Network Query Processing in Sensor Networks," *SIGMOD Record*, vol.31, no.3, pp.9-10 2002.
- [2] S. Madden, M. J. Franklin, J. M. Hellerstein and W. Hong, "TinyDB : an acquisitional query processing system for sensor networks," *ACM Trans. Database Syst.*, vol.30, no.1, pp.122-173, 2005.
- [3] A. Coman, and M. A. Nascimento, "A distributed Algorithm for Joins in Sensor Networks," *Proc. of Intl' Conference on SSDBM*, pp.27-35, 2007.
- [4] H. Yu, E. Lim, and J. Zhang, "In-Network Join Processing for Sensor Networks," *LNCS*, vol.3841, pp.263-274, 2006.
- [5] P. A. Bernstein, and D. W. Chiu, "Using Semi-Joins to Solve Relational Queries," *Journal of th ACM*, 28(1), pp.25-40, 1981.
- [6] A. Broder, and M. Mitzenmacher, "Network applications of Bloom filters : a survey," *Internet Mathematics*, vol.1, pp.485-509, 2004.
- [7] Z. Li, and K. Ross, "PERF Join : An Alternative to Two-way Semijoin and Bloomjoin," *Proc. of Intl' CIKM*, pp.137-144, 1995.
- [8] I. Song, K. Kim "Efficient Record Filtering In-network Join Strategy using Bit-Vector in Sensor Networks," *Journal of KSCI*, vol.15, no.4, pp.27-36, April. 2010. (in Korean)
- [9] I. Greenberg, and R. A. Robertello, "The Three Factory Problem," *Mathematics Magazine*, 38-2, pp.67-72, 1965.
- [10] B. Karp, and H. Kung. "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," In *Proc. of ACM MobiCom '00*, pp.243-254, 2000.
- [11] W. B. Heinzelman, A. P. Chandarkasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *Wireless Comm, IEEE*, vol.1, no.4, pp.660-670, 2002.



송 임 영

2002년 서울산업대학교 전자계산학과 공학사. 2004년 서울산업대학교 컴퓨터공학과 공학석사. 현재 홍익대학교 컴퓨터공학과 재학. 관심분야는 객체지향DB, Ubiquitous 환경에서의 Data Mining, 센서 DB



김 경 창

1978년 홍익대학교 전자계산학과 공학사. 1980년 한국과학기술원 전산학과 공학석사. 1990년 University of Texas at Austin 전산학과 공학박사. 현재 홍익대학교 정보컴퓨터공학부 교수. 관심분야는 객체지향DB, 주기억DB, 센서DB