



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0068461
(43) 공개일자 2011년06월22일

(51) Int. Cl.

H04W 40/32 (2009.01) H04W 84/18 (2009.01)
H04L 12/28 (2006.01)

(21) 출원번호 10-2009-0125428

(22) 출원일자 2009년12월16일

심사청구일자 2009년12월16일

(71) 출원인

한국기술교육대학교 산학협력단

충남 천안시 동남구 병천면 가전리 307 한국기술
교육대학교 내

(72) 발명자

민준기

대전광역시 서구 월평동 266-1 샛별아파트 103동
1308호

양희정

충청남도 천안시 서북구 백석동 아이파크 104동
1301호

정진완

대전광역시 유성구 노은동 520-1 열매마을 8단지
814동 1302호

(74) 대리인

김희곤, 박용순, 김인한

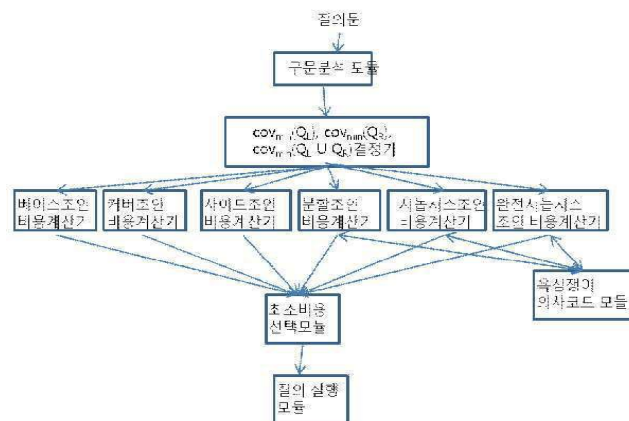
전체 청구항 수 : 총 6 항

(54) 비용기반 최적 조인 처리 방법 및 시스템

(57) 요약

본 발명은 센서 네트워크를 구성하는 각 센서들에 획득된 측정값들의 조인 연산을 효율적으로 처리하기 위한 비용기반 최적 조인 처리 방법 및 시스템에 대한 것이다. 본 발명에서는 센서 네트워크상에서 조인 연산을 수행할 수 있는 다양한 방법들을 파악하고 새로운 조인 방법을 제공하고 이를 위한 비용식을 도출하여 다양한 환경에 따른 최적의 조인 방법을 선택하는 것을 특징으로 함으로써 센서 네트워크상에서의 조인 연산에 따른 센서들의 에너지 소비를 최소화한다.

대표도 - 도5



특허청구의 범위

청구항 1

센서네트워크에서 사용되는 조인처리 방법에 있어서,

- (a) 조인 연산이 포함된 질의문을 수행할 센서 노드들을 선택하는 단계;
- (b) 상기 선택된 센서 노드들에 대한 조인방안들을 생성하고, 상기 조인방안들 각각에 대한 처리 비용을 계산하는 단계;
- (c) 상기 처리 비용 중 최소 비용에 해당하는 조인방안을 선택하여, 상기 선택된 조인방안에 따른 조인연산을 수행하는 단계를 포함하는 것을 특징으로 하는 비용기반 최적 조인 처리 방법.

청구항 2

제 1항에 있어서,

상기 (b) 단계 중 처리비용을 계산하는 것은 아래의 수학적식 1 을 사용하며,

{수학적식 1}

$$Cost(Q, d) = Cost_{gathering}(s) + Cost_{sending}(s, d)$$

여기서, Cost(Q,d)는 Q 영역 내의 센서들의 측정값을 s 노드를 거쳐 d 노드까지 보내는 비용이며,

상기 Cost_{gathering}(s) 는,

$$Cost_{gathering}(s) = \sum_{c \in child(s)} (Cost_{gathering}(c) + T(c))$$

로서,

여기서, Cost_{gathering}(c) 및 T(c) 는 각각 상기 Q 영역 내의 임의의 센서로부터의 측정값을 s 노드로 보내는 비용 및 상기 측정값의 크기이며,

상기 Cost_{sending}(S,d) 은,

$$Cost_{sending}(s, d) = hopDiff(s, d) \cdot T(s)$$

로서,

여기서, hopDiff(s,d)는 s 노드에서 d 노드까지 거치는 노드의 개수인 것을 특징으로 하는 비용기반 최적 조인 처리 방법.

청구항 3

제 1항에 있어서,

상기 (b) 단계의 조인방안들 각각은,

베이스 조인 방안, 커버 조인 방안, 사이드 조인 방안, 분할 조인 방안, 시놉시스 조인 방안, 및 완전 시놉시스 조인 방안 중에서 선택된 것을 특징으로 하는 비용기반 최적 조인 처리 방법.

청구항 4

조인 연산이 포함된 질의문을 분석하는 구문분석 모듈;

상기 분석한 질의문에 따라 센서 노드들을 결정하는 결정기;

상기 결정된 센서 노드들에 대한 조인방안들을 생성하고, 상기 조인방안들 각각에 대한 처리 비용을 계산하는

비용계산기; 및

상기 계산된 처리 비용 중 최소 비용의 조인방안을 선택하는 최소비용 선택모듈; 및

상기 선택된 조인방안에 따라 조인 연산을 수행하는 질의 실행 모듈을 포함하는 것을 특징으로 하는 비용기반 최적 조인 처리 시스템.

청구항 5

제 4항에 있어서,

상기 비용계산기는,

베이스 조인 비용계산기, 커버 조인 비용계산기, 사이드 조인 비용계산기, 분할 조인 비용계산기, 시놉시스 조인 비용계산기, 및 완전 시놉시스 조인 비용계산기 중 하나 이상을 포함하는 것을 특징으로 하는 비용기반 최적 조인 처리 시스템.

청구항 6

제 5항에 있어서,

상기 분할 조인 비용계산기, 시놉시스 조인 비용계산기, 및 완전 시놉시스 조인 비용계산기가 최소비용을 계산하도록, 넓이 우선 탐색 (breadth first traversal)에 기반한 욕심쟁이 접근 (greedy approach) 방법을 제공하는 욕심쟁이 의사코드 모듈을 더 포함하는 것을 특징으로 하는 비용기반 최적 조인 처리 시스템.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 센서 네트워크에서의 비용기반 최적 조인 처리 방법 및 시스템에 관한 것으로, 더욱 상세하게는 센서 네트워크를 구성하는 센서들로부터 획득된 측정값들에 대한 조인 연산에 있어서 효율적인 조인 처리 방안의 확장 및 이에 대한 처리 비용을 계산하여 다양한 상황에 따른 최적의 조인 처리 기법을 선택하도록 하여 센서들의 에너지 소비를 줄이는 방법에 관한 것이다.

배경 기술

[0002] 센서 네트워크는 다수의 센서와 사용자가 데이터에 접근할 수 있는 베이스 스테이션으로 구성되어 있다. 센서 데이터 모니터링은 특정지역에 분포되어 있는 센서들을 이용하여 해당 지역의 온도, 습도, 조도 등 물리적 환경을 수치적으로 측정하여 베이스스테이션 (base station)으로 데이터를 무선 통신을 통하여 전송하여 지속적으로 관찰하는 작업을 말한다. 이에 따라서, 모니터링을 위한 센서 네트워크는 지속적으로 측정값을 베이스스테이션으로 전송함으로써 네트워크 자원을 많이 소비한다. 따라서, 데이터의 무선 전송에 따른 에너지 소비를 최소화하는 방안에 대한 관심이 증대되고 있다.

[0003] 센서 모니터링에서 사용자가 서로 다른 두 지역의 센서 측정값들 간의 관계를 파악하고 싶을 경우가 있다. 예를 들어 기후학자가 특정지역의 강우량과 다른 지역의 온도 간의 상관관계를 알고 싶을 것이다. 이러한 상관관계를 두 지역의 센서 측정값에 대한 조인 연산으로 표현된다.

[0004] Cougar [참고문헌: A. Demers, J. Gehrke, R. Rajaraman, N. Trigoni, and Y. Yao. The cougar project: A work-in-progress report. *SIGMOD Record*, 32(3) 2003.] 나 TinyDB[참고문헌: S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. Tinydb: an acquisitional query processing system for sensor networks. *ACM Trans. Database Systems*, 30(1):122-173, 2005] 같은 센서 데이터 관리 시스템은 센서 네트워크에서 조인 연산을 수행하는 가장 단순한 방법인 두 지역의 센서들이 자신의 측정값을 베이스 스테이션으로 보내고 베이스 스테이션에서 조인 연산들 수행하는 방법을 사용하였다. 이러한 경우 각 센서들이 측정값을 전송하기 위하여 많은

에너지를 소모하게 되는 단점이 발생한다.

- [0005] 따라서, 가능한 빨리 관련 없는 센서 측정값을 제거하기 위하여 센서 네트워크를 구성하는 센서들에서 조인을 수행하는 네트워크 내 조인 처리 (in-network join processing) 방법들이 제안되었다.
- [0006] Alabdi 등[참고문헌: D. J. Abadi, S. Madden, and W. Lindner. Reed: Robust, efficient filtering and event detection in sensor networks. In *Proceedings of VLDB*, 2005.] 은 센서 측정값과 네트워크 외부에서 만들어진 데이터들간의 조인을 지원하는 REED라는 기법을 제안하였다. 그러나 이 기법은 서로 다른 지역의 센서 측정값 간의 조인 연산을 고려하지 않았다.
- [0007] 서로 다른 두 지역의 센서 측정값에 대한 네트워크 내 조인 기법들이 Cowdhary 와 Gupta [참고문헌: Communication-efficient implementation of join in sensor networks. In *Proceedings of DASFAA* 2005.] 의 하여 제안되었다. 그러나 이 기법은 단순히 두 지역의 중심 위치들과 베이스 스테이션 위치로 구성된 세 점의 중심점(centroid)을 조인 연산이 수행되는 위치로 선정함으로써 다양한 환경에 따른 최적의 조인 방법을 제시하지 못하고 있다.
- [0008] Yang 등 [참고문헌: X. Yang, H. B. Lim, M. T. Ozsu, and K. L. Tan. In-network execution of monitoring queries in sensor networks. In *Proceedings of ACM SIGMOD*, 2007.] 은 두 단계 자신 조인(Two-phase self join: TPSJ) 기법을 제안하였다. Stern등 [참고문헌: M. Stern, E. Buchmann, and K. B ohn. Towards efficient processing of general-purpose joins in sensor networks. In *Proceedings of ICDE*, 2009.]은 quad tree를 이용한 네트워크 내 조인 기법인 SENS-Join을 제안하였다. 이는 Yang등의 기법과 유사하게 두 단계로 구성된다. 그러나, 질의 최적화기에 필요한 SENS-Join에 대한 비용 모델을 제시하고 있지 않다.
- [0009] Yu 등은 [참고문헌: H. Yu, E.-P. Lim, and J. Zhang. On in-network synopsis join processing for sensor networks. In *Proceedings of the International Conference on Mobile Data Management*, 2006.] 데이터의 요약본인 시놉시스(synopsis)를 이용한 네트워크 내 조인 방법을 제안하였다. 그러나 여기서 사용한 방법은 두 영역의 중간 위치를 파악하여 이를 조인 수행위치로 선정하는 것으로 Cowdhar와 Gupta방법과 마찬가지로 다양한 상황에 따른 조인 최적화기의 방안으로 사용하기에는 너무 단순하다.

발명의 내용

해결 하고자하는 과제

- [0010] 본 발명은 상술한 문제를 해결하기 위하여 안출된 것으로, 본 발명의 목적은, 상기한 센서 네트워크에서 조인 연산을 처리할 때 다양한 상황에 따라서 최적의 실행 방안을 선택하도록 하여 센서들의 에너지 소비를 최소화하기 위하여 이루어진 것으로서, 센서들의 무선 데이터 전송의 특성을 고려하여 다양한 네트워크 내 조인 연산 처리 기법들을 분석하고 새로운 조인 연산 처리 기법들을 제공하고 각 기법들에 대한 처리 비용 모델을 제공하여 센서 네트워크에서의 비용 기반 최적 조인 처리 방법 및 시스템을 제공하는 데 있다.

과제 해결수단

- [0011] 본 발명의 일 실시형태에 따른, 비용 기반 최적 조인 처리 방법은, 센서네트워크에서 사용되는 조인처리 방법에 있어서, (a) 조인 연산이 포함된 질의문을 수행할 센서 노드들을 선택하는 단계; (b) 상기 선택된 센서 노드들에 대한 조인방안들을 생성하고, 상기 조인방안들 각각에 대한 처리 비용을 계산하는 단계; (c) 상기 처리 비용 중 최소 비용에 해당하는 조인방안을 선택하여, 상기 선택된 조인방안에 따른 조인연산을 수행하는 단계를 포함하는 것을 특징으로 한다.

- [0012] 여기서, 상기 (b) 단계 중 처리비용을 계산하는 것은 아래의 수학식 1 을 사용하며, {수학식 1}

- [0013]
$$Cost(Q, d) = Cost_{gathering}(s) + Cost_{sending}(s, d)$$

- [0014] 여기서, Cost(Q, d)는 Q 영역 내의 센서들의 측정값을 s 노드를 거쳐 d 노드까지 보내는 비용이며, $Cost_{gathering}(s)$ 는,

- [0015]
$$Cost_{gathering}(s) = \sum_{c \in child(s)} (Cost_{gathering}(c) + T(c))$$
로서,

- [0016] 여기서, $Cost_{gathering}(c)$ 및 $T(c)$ 는 각각 상기 Q 영역 내의 임의의 센서로부터의 측정값을 s 노드로 보내는 비용 및 상기 측정값의 크기이며, 상기 $Cost_{sending}(S, d)$ 은, $Cost_{sending}(s, d) = hopDiff(s, d) \cdot T(s)$ 로서, 여기서, $hopDiff(s, d)$ 는 s 노드에서 d 노드까지 거치는 노드의 개수인 것을 특징으로 한다.
- [0017] 또한, 상기 (b) 단계의 조인방안들 각각은, 베이스 조인 방안, 커버 조인 방안, 사이드 조인 방안, 분할 조인 방안, 시놉시스 조인 방안, 및 완전 시놉시스 조인 방안 중에서 선택된 것을 특징으로 한다.
- [0018] 또한, 본 발명의 또 다른 실시형태에 따른 비용 기반 최적 조인 처리 시스템은, 조인 연산이 포함된 질의문을 분석하는 구문분석 모듈; 상기 분석한 질의문에 따라 센서 노드들을 결정하는 결정기; 상기 결정된 센서 노드들에 대한 조인방안들을 생성하고, 상기 조인방안들 각각에 대한 처리 비용을 계산하는 비용계산기; 및 상기 계산된 처리 비용 중 최소 비용의 조인방안을 선택하는 최소비용 선택모듈; 및 상기 선택된 조인방안에 따라 조인 연산을 수행하는 질의 실행 모듈을 포함하는 것을 특징으로 한다.
- [0019] 여기서, 상기 비용계산기는, 베이스 조인 비용계산기, 커버 조인 비용계산기, 사이드 조인 비용계산기, 분할 조인 비용계산기, 시놉시스 조인 비용계산기, 및 완전 시놉시스 조인 비용계산기 중 하나 이상을 포함하는 것을 특징으로 한다.
- [0020] 또한, 상기 분할 조인 비용계산기, 시놉시스 조인 비용계산기, 및 완전 시놉시스 조인 비용계산기가 최소비용을 계산하도록, 넓이 우선 탐색 (breadth first traversal)에 기반한 욕심쟁이 접근 (greedy approach) 방법을 제공하는 욕심쟁이 의사코드 모듈을 더 포함하는 것을 특징으로 한다.

효 과

- [0021] 본 발명에 의한 센서 네트워크에서 조인 연산이 포함된 질의문 수행을 위한 비용 기반 최적 조인 처리 방법 및 시스템은 다양한 상황에 따라서 최적의 실행 방안을 선택하도록 하여 센서들의 에너지 소비를 최소화하도록 한다. 따라서, 본 발명은 센서 네트워크 상의 측정값들을 검색, 분석하는 등 센서 데이터관리 분야에 크게 기여할 수 있다.

발명의 실시를 위한 구체적인 내용

- [0022] 이하에서는 첨부한 도면을 참조하여 바람직한 일 실시형태에 따른 비용기반 최적 조인 처리 방법 및 시스템에 대해서 상세히 설명한다. 다만, 실시형태를 설명함에 있어서, 관련된 공지 기능 혹은 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그에 대한 상세한 설명은 생략한다.
- [0023] 기존의 센서 네트워크에서의 단일 조인 연산 처리 방법들은 한가지 방법만을 제공함으로써 상황에 따른 최적의 조인 방법을 제공하지 못한다. 이에 반하여 본 발명에서 제안하는 비용기반 조인 처리 최적화기는 상황에 따른 다양한 조인 연산 방법들의 비용을 계산하고 비교하여 최적의 조인 연산 방법을 제공한다. 더욱이 본 발명에서는 기존의 센서 네트워크에서의 조인 연산 처리 방법과는 다른 조인 연산 처리 방법들을 고안하고 이에 대한 비용 모델을 제시한다. 조인 연산 최적화기에서는 기존의 조인연산 처리 방법들 및 새롭게 고안된 조인 연산 처리 방법들을 고려하여 최적의 조인 연산 처리 방법을 선택하고 이를 수행하도록 한다.
- [0024] 관계형 데이터베이스의 질의언어인 SQL을 확장하여 센서네트워크상의 두 영역의 센서 측정값을 조인하는 질의언어 구문 규칙의 일부분을,
- ```
SELECT L.*, R.*
FROM Sensor L, Sensor R
WHERE L.location in Q_L
 AND R.location in Q_R
 AND $P_L(L)$
 AND $P_R(R)$
 AND $P_{JOIN}(L,R)$
{SAMPLE INTERVAL X} {FOR D}
```
- [0025]
- [0026] 와 같이 나타낸다.
- [0027] 이를 참조하면, FROM절의 L과 R은 특정 시간의 센서 네트워크 상의 센서들이 위치 및 측정한 측정값들을 나타내며, WHERE절의 영역  $Q_L$ 과  $Q_R$ 로 서로 다른 두 영역 상에 존재하는 센서들의 측정값으로 제한된다.  $P_L$ 과  $P_R$ 은 각각 L과 R 측정값들의 선택조건을 나타낸다. 예를 들어 측정값이 30 이상인 경우로 제한하고자 할 경우 사용된다.

$P_{JOIN}$ 은 L과 R의 조인 조건을 나타낸다. 센서 측정값들은 지속적으로 생성됨으로 이를 위하여 D 초 동안 X 초 간격으로 질의가 수행됨을 SAMPLE INTERVAL X FOR D 형태로 표현한다.

[0028] 우선, 본 발명에서 상기 제시된 형태의 질의의 수행 비용을 계산하기 위하여 필요한 기본 수식에 대하여 설명하도록 한다. 그 다음, 다양한 조인 연산 처리 방법을 제시하고 기본 수식을 적용하여 처리 비용 계산 방법을 설명하도록 한다.

[0029] 도 1은 특정 질의 영역에 속한 센서들이 무선 통신을 통하여 임의의 노드 d 까지 측정값을 전송하는 것을 나타낸다.

[0030] 도 1에서 보는 바와 같이 영역 Q에 속한 센서들은 무선 통신으로 노드 s를 경유하여 원하는 목적지 d까지 데이터를 멀티 홉 라우팅 (multi hop routing)을 이용하여 전송하게 된다. 즉, 센서 s에서 Q에 속한 센서들의 측정값을 모두 수집하여 목적지 d로 전송한다. 추가로 질의 영역 Q의 측정값에는 도면 1의 질의 구문에 나타난 바와 같이 선택 조건  $P_Q$ 가 적용된다. 이 선택 조건  $P_Q$ 을 만족하는 측정값들의 비율을 선택도(selectivity)  $sel(P_Q)$ 라고 하자. 여기서, 질의 영역 밖에서 측정된 측정값들의 선택도는 0이다. 또한,  $P_Q$ 가 적용되지 않았을 경우에 Q상의 측정값들의 선택도는 1이 된다. 센서 측정값의 크기를 r이라고 할 때, 노드 s가 전송할 측정값들의 총 크기  $T(s)$ 는 수학적 식 1과 같이 표현될 수 있다.

### 수학적 식 1

$$T(s) = r \cdot sel(P_Q) + \sum_{c \in child(s)} T(c)$$

[0031]

[0032] 따라서, Q 영역 안에 속해 있는 센서들의 측정값을 d까지 보내는 비용을  $Cost(Q, d)$ 라 할 때 다음과 같은 수학적 식2로 표현할 수 있다.

### 수학적 식 2

$$Cost(Q, d) = Cost_{gathering}(s) + Cost_{sending}(s, d)$$

[0033]

[0034] 노드 s에서 측정값들을 모으기 위해서는 s의 자식 센서들이 측정값을 모으고, 모아진 측정값을 s로 전송한다. 따라서,  $Cost_{gathering}(s)$ 는 수학적 식3과 같이 정의된다.

### 수학적 식 3

$$Cost_{gathering}(s) = \sum_{c \in child(s)} (Cost_{gathering}(c) + T(c))$$

[0035]

[0036] 노드 s에서 다른 노드로 측정값을 전송할 때 전송량  $T(s)$ 는 변화하지 않음으로 노드 s에서 노드d까지의 홉 디스턴스를  $hopDiff(s, d)$  (s에서 d까지 거쳐야 하는 노드들의 수)라 할 때, 노드 s에서 목적지 d까지 전송비용은  $Cost_{sending}(s, d)$ 는 수학적 식 4와 같이 정의된다.

### 수학적 식 4

$$Cost_{sending}(s, d) = hopDiff(s, d) \cdot T(s)$$

[0037]

[0038] 위의 수학적식을 기반으로 다양한 조인 연산 처리 방법들의 비용식을 생성할 수 있다. 본 발명의 내용을 상술하기 위하여 몇 가지 기본 개념을 정의한다.

[0039] 정의 1. 센서 s의 센싱영역  $SR(s)$ 는 s와 s의 자손들의 위치를 포함하는 최소 경계 영역이다.

[0040] 정의 2. 질의 영역 Q에 대하여  $cov_{min}(Q)$ 는 Q를 포함하는 최소 크기의  $SR()$ 을 지니는 노드를 나타낸다.

[0041] 도 2는 위의 정의 1과 정의 2의 예를 나타내는 것으로 질의 영역 Q가 실선 사각형으로 표시될 때, 센서  $s_2$ 의 센

싱영역  $SR(s_2)$ 가  $Q$ 를 포함하는 최소 크기의 센싱영역임으로  $cov_{min}(Q) = s_2$ 가 된다.

- [0042] 도 3은 조인 연산을 포함한 질의 처리를 위한 기본적인 3가지 조인 처리 방안을 나타내는 것으로 이 방법들은 일반적으로 센서 네트워크에서 조인 연산을 처리하는 방법들이다.
- [0043] 도 3의 (a)는 가장 단순한 방법인 베이스 조인(baseJoin)으로 질의 영역에 속한 센서들의 측정값을 모두 베이스 스테이션을 전송하고 베이스 스테이션에서 조인 연산을 수행하는 것이다. 여기서 조인 연산은  $\triangleright \triangleleft$ 으로 표시된다.
- [0044] 도 3의 (b)는 커버 조인(coverJoin)으로 두 질의 영역  $Q_L$ 과  $Q_R$ 의 모든 측정값을 모으는 노드에서 조인 연산을 수행하는 방법이다. 질의 영역  $Q_L$ 과  $Q_R$ 의 모든 측정값이 모이는 노드는 정의 2에 의하여  $cov_{min}(Q_L \cup Q_R)$ 로 표현할 수 있다. 이때, 조인 연산을 수행하는 노드  $cov_{min}(Q_L \cup Q_R)$ 에서는 일반적인 조인 연산과 달리 데이터의 집합에서 조인 결과에 포함되는 데이터만 추출하는 세미조인(semi-join)을 수행하여 조인 연산에 참여하지 않는 측정값들을 베이스 스테이션으로 전송하지 않도록 한다. 세미조인은  $\triangleright <$ 로 표현된다. 이에 따라서, 측정값들의 집합  $L$ 에서 조인 결과에 포함되는 측정값을 추출하는  $L \triangleright < R$ 과 측정값들의 집합  $R$ 에서 조인 결과에 포함되는 측정값만을 추출하는  $R \triangleright L$ 이 수행된다.
- [0045] 도 3의 (c)는 사이드 조인(sideJoin)을 나타낸다. 사이드 조인에서는  $Q_L$  상의 모든 측정값이  $cov_{min}(Q_L)$ 에서 모이고 모아진 측정값들인  $L$ 을  $cov_{min}(Q_L \cup Q_R)$ 로 보낸다. 세미조인  $R \triangleright < L$ 에 있어서 데이터 집합  $R$ 중 조인 결과에 참여하는 데이터만을 추출하는데 있어서 조인 대상인  $L$ 이 모두 필요하지 않고  $L$ 중 조인 속성값들 부분만 필요하므로 이를 이용하여 세미조인을 할 수 있다. 따라서,  $cov_{min}(Q_L \cup Q_R)$ 에서는  $L$ 중에서 조인 연산의 조건이 되는 조인 속성만을 추출하고 이를  $cov_{min}(Q_R)$ 로 전송한다.  $cov_{min}(Q_R)$ 에서는 모아진  $Q_R$ 상의 모든 측정값들의 집합  $R$ 과 전송받은 조인 속성값들만을 이용하여  $R \triangleright L$ 을 수행한다. 이 수행 결과를 다시  $cov_{min}(Q_L \cup Q_R)$ 로 보내어  $L \triangleright < R$ 을 수행하고 두 세미조인의 결과를 베이스 스테이션으로 전송한다.  $R$ 의 조인 속성을  $cov_{min}(Q_L)$ 로 보내어  $L \triangleright < R$ 를 먼저 수행할 수 있다. 이는 위의 사이드 조인 설명과 동일한 원리임으로 생략한다.
- [0046] 베이스 조인 방식의 비용  $Cost_{base}(Q_L, Q_R)$ 은 수학식 5와 같이 나타낼 수 있다.

### 수학식 5

$$Cost_{base}(Q_L, Q_R) = Cost_{gathering}(cov_{min}(Q_L)) + Cost_{gathering}(cov_{min}(Q_R)) + D_L \cdot T(cov_{min}(Q_L)) + D_R \cdot T(cov_{min}(Q_R)) + D_C \cdot (T(cov_{min}(Q_L)) + T(cov_{min}(Q_R)))$$

- [0047]
- [0048] 수학식 5에서  $D_L$ 을  $cov_{min}(Q_L)$ 에서  $cov_{min}(Q_L \cup Q_R)$ 까지의 홉 거리를 나타내고  $D_R$ 은  $cov_{min}(Q_R)$ 에서  $cov_{min}(Q_L \cup Q_R)$ 까지의 홉 거리,  $D_C$ 는  $cov_{min}(Q_L \cup Q_R)$ 에서 베이스스테이션까지의 홉 거리를 나타낸다. 수학식 5에 나타나 있는 것과 같이  $cov_{min}(Q_L)$ 과  $cov_{min}(Q_R)$ 에서  $Q_L$ 과  $Q_R$ 상의 측정값을 수집하는 비용 ( $= Cost_{gathering}(cov_{min}(Q_L))$ 과  $Cost_{gathering}(cov_{min}(Q_R))$ ) 그리고 수집된 데이터를  $cov_{min}(Q_L \cup Q_R)$ 로 전송하는 비용 ( $= D_L \cdot T(cov_{min}(Q_L)) + D_R \cdot T(cov_{min}(Q_R))$ ) 마지막으로  $cov_{min}(Q_L \cup Q_R)$ 에서 베이스스테이션으로 전송하는 비용 ( $= D_C \cdot (T(cov_{min}(Q_L)) + T(cov_{min}(Q_R)))$ )으로 구성된다.
- [0049] 커버 조인에서는  $cov_{min}(Q_L \cup Q_R)$ 에서 세미조인이 수행됨으로  $L \triangleright < R$ 과  $R \triangleright < L$ 의 결과 크기를 계산해야 한다. 세미조인 선택률(semijoin selectivity)은  $L$  중  $L \triangleright < R$ 의 결과의 크기 비율을 나타낸다. 이를 위하여 Hevner [참고문헌: A. R. Hevner and S. B. Yao. Query processing in distributed database systems. *IEEE Trans. Software Eng.*, 5(3):177-187, 1979.]등이 세미조인 선택률 계산식으로 수학식 6을 제안하였다.

## 수학식 6

$$SJS(R \ltimes_A L) = \frac{size(\Pi_A(L))}{size(dom(A))}$$

[0050]

[0051] 수학식 6에서 세미조인 조건을 사용되는 속성을 A라 할 때, dom(A)는 A가 가질 수 있는 값의 범위(domain)이라고 하고  $\Pi_A(L)$ 은 L중에서 속성 A부분만을 뽑아낸 결과를 나타낸다.

[0052] 이를 반영한 커버 조인 방식의 비용  $Cost_{cover}(Q_L, Q_R)$ 의 비용은 수학식 7과 같다.

## 수학식 7

$$\begin{aligned} Cost_{cover}(Q_L, Q_R) = & Cost_{gathering}(cov_{min}(Q_L)) + \\ & Cost_{gathering}(cov_{min}(Q_R)) + \\ & D_L \cdot T(cov_{min}(Q_L)) + D_R \cdot T(cov_{min}(Q_R)) + \\ & D_C \cdot (T(cov_{min}(Q_L)) \cdot SJS_{R.A} + \\ & T(cov_{min}(Q_R)) \cdot SJS_{L.A}) \end{aligned}$$

[0053]

[0054] 수학식 6에서  $SJS(R \triangleright \triangleleft_A L)$ 은 L의 속성 A에만 영향을 받음으로  $SJS_{L.A}$ 로 표현한다. 동일하게  $SJS(L \triangleright \triangleleft_A R)$ 은  $SJS_{R.A}$ 로 표기한다.

[0055] 사이드 조인의 비용  $Cost_{side}(Q_L, Q_R)$ 은 수학식 8과 같이 계산할 수 있다.

## 수학식 8

$$\begin{aligned} Cost_{side}(Q_L, Q_R) = & Cost_{gathering}(cov_{min}(Q_L)) + \\ & Cost_{gathering}(cov_{min}(Q_R)) + \\ & D_L \cdot T(cov_{min}(Q_L)) + D_R \cdot T(cov_{min}(Q_L)) \cdot \frac{j}{r} + \\ & D_R \cdot T(cov_{min}(Q_R)) \cdot SJS_{L.A} + \\ & D_C \cdot (T(cov_{min}(Q_L)) \cdot SJS_{R.A} + \\ & T(cov_{min}(Q_R)) \cdot SJS_{L.A}) \end{aligned}$$

[0056]

[0057] 사이드 조인에서는  $Q_L$ 상의 측정값을  $cov_{min}(Q_L \cup Q_R)$ 에 전송하고  $cov_{min}(Q_L \cup Q_R)$ 에서 조인 속성부분만을  $cov_{min}(Q_R)$ 로 전송한다. 따라서 조인 속성의 크기를 j라 하고 측정값의 크기를 r이라 할 때,  $T(cov_{min}(Q_L)) \cdot j/r$  크기의 데이터가  $D_R$  거리만큼 전송하게 된다. 그리고  $cov_{min}(Q_R)$ 에서는  $R \triangleright \triangleleft L$ 의 결과를  $cov_{min}(Q_L \cup Q_R)$ 로 전송함으로써 이 비용은  $D_R \cdot T(cov_{min}(Q_R)) \cdot SJS_{L.A}$ 가 된다. 이후  $cov_{min}(Q_L \cup Q_R)$ 에서  $L \triangleright \triangleleft R$ 을 수행하고 커버 조인에서 마찬가지로 두 세미 조인의 결과를 베이스 스테이션으로 전송한다.

[0058] 본 발명에서는 위에서 설명한 일반적인 세 가지 조인 처리 방법들 외에 추가로 분할 조인 (partition join) 방법을 고안하였다.

[0059] 분할 조인 방법의 기본 개념은  $U_c R_c = R$  일 경우,  $U_c R_c \triangleright \triangleleft L = R \triangleright \triangleleft L$  과 같다는 것이다. 즉, R을 여러 개의 부분집합들로 나누고 이 부분집합들에 대하여 세미조인을 수행한 결과들의 합은 원 세미조인의 결과와 같다는 것이다. 이러한 개념을 네트워크 내 조인 처리 방법에 적용하여 정의 3과 같이 질의 영역 Q을 여러 개의 부분질의영역들로 나눌 수 있다.

[0060] 정의 3. 질의 영역 Q의 부분질의영역들의 집합  $PQ = \{Q_1, Q_2, \dots, Q_n\}$ 의 원소인 부분질의영역  $Q_i$ 는,  $cov_{min}(Q)$ 의 자식 노드 i에 대하여,  $SR(i)$ 와 Q와 겹치는 영역이다.

[0061] 분할 조인 방법은 이와 같이 분할된 영역에서 세미조인을 수행하는 방법으로 도 4는 분할 조인 방법의 예를 나타낸다.

[0062] 도 4에서 보듯이 질의 영역  $Q_R$ 에 대한  $cov_{min}(Q_R)$ 은 정의 3에 의한 부분질의영역(도 4에서는  $Q_{R1}, Q_{R2}, Q_{R3}$ 로

표현) 상에 측정값들을 수집함으로  $\text{cov}_{\min}(Q_R)$  또는  $\text{cov}_{\min}(Q_R)$ 의 자식 노드인  $c$ 에서  $R_c \triangleright < L$  를 수행할 수 있다.

[0063] 따라서, 분할 조인 방법에서는  $R_c \triangleright < L$ 를 자식 노드  $c$ 에서 할지  $\text{cov}_{\min}(Q_R)$ 에서 할지를 결정하는 문제가 있다. 또한, 이 문제는 자식 노드  $c$ 에서 재귀적으로  $c$ 의 자식 노드로 적용된다. 예를 들어 도 4에서 노드  $s_1$ 과  $s_2$ 에서는  $R_1 \triangleright < L$ 과  $R_2 \triangleright < L$ 가 수행되며  $s_3$ 에서는  $R_3$ 를  $\text{cov}_{\min}(Q_R)$ 로 전송하여  $R_3 \triangleright < L$ 을  $\text{cov}_{\min}(Q_R)$ 에서 수행하였다. 여기서 주목할 점은  $s_1$ 과  $s_2$ 에서 분할 조인이 재귀적으로 적용될 수 있다는 것이다.

[0064] 수학식9는 분할 조인 방법의 비용  $\text{Cost}_{\text{partition}}(Q_L, Q_R)$ 을 나타낸다.

### 수학식 9

$$\begin{aligned} \text{Cost}_{\text{partition}}(Q_L, Q_R) = & \text{Cost}_{\text{gathering}}(\text{cov}_{\min}(Q_L)) + \\ & \text{Cost}_{\text{partial}}(\text{cov}_{\min}(Q_L), \text{cov}_{\min}(Q_R)) + \\ & D_L \cdot T(\text{cov}_{\min}(Q_L)) + D_R \cdot T(\text{cov}_{\min}(Q_L)) \cdot \frac{j}{r} + \\ & D_R \cdot T(\text{cov}_{\min}(Q_R)) \cdot SJS_{L.A} + \\ & D_C \cdot (T(\text{cov}_{\min}(Q_L)) \cdot SJS_{R.A} + \\ & T(\text{cov}_{\min}(Q_R)) \cdot SJS_{L.A}) \end{aligned}$$

[0065]

[0066] 기본적인 비용은 수학식 8의 커버조인 비용과 유사하나 다른 점은 수학식 8의  $\text{Cost}_{\text{gathering}}(\text{cov}_{\min}(Q_R))$  항 대신  $\text{Cost}_{\text{partial}}()$ 항이 쓰였다는 것이다.

[0067]  $\text{Cost}_{\text{partial}}()$ 항은 수학식 10으로 정의된다.

### 수학식 10

$$\begin{aligned} \text{Cost}_{\text{partial}}(\text{cov}_{\min}(Q_L), \text{cov}_{\min}(Q_R)) = & \min_{C \subseteq \text{qchild}(\text{cov}_{\min}(Q_R))} \{ T_s(\text{cov}_{\min}(Q_L)) \cdot \frac{j}{r} + \\ & \sum_{c \in C} T_r(\text{cov}_{\min}(Q_L)) \cdot \frac{j}{r} + \\ & \sum_{c \in C} T(c) \cdot SJS_{L.A} + \\ & \sum_{c \in C} (\text{Cost}_{\text{partial}}(\text{cov}_{\min}(Q_L), c) + \\ & \sum_{c' \in \text{qchild}(\text{cov}_{\min}(Q_R)) - C} (\text{Cost}_{\text{gathering}}(c') + T(c')) \} \\ & , \text{ where } \text{qchild}(\text{cov}_{\min}(Q_R)) \text{ is a set of children whose } SR() \\ & \text{ overlap with } Q_R \\ & \text{ and } T_s(\text{cov}_{\min}(Q_L)) = 0 \text{ if } C = \emptyset \end{aligned}$$

[0068]

[0069] 수학식 10에서  $\text{cov}_{\min}(Q_R)$ 의 자식 중  $Q_R$ 과 중첩되는  $SR()$ 을 지니는 자식 노드들을  $\text{qchild}(\text{cov}_{\min}(Q_R))$ 이라고 할 때,  $\text{qchild}(\text{cov}_{\min}(Q_R))$ 의 부분집합  $C$ 에서 세미조인을 수행하기 위해서는  $\Pi_A(L)$ 을  $C$ 로 보내야 한다. 데이터 전송 비용  $T()$ 는 데이터 발신비용  $T_s()$ 와 데이터 수신 비용  $T_r()$  나누어 볼 수 있다. 센서 네트워크에서는 부모 노드가 여러 자식노드에게 데이터를 방송(broadcast)할 수 있으므로 데이터 전송 비용은  $T_s(\text{cov}_{\min}(Q_L)) \cdot j/r$ 이라고 할 수 있다. 이에 따라  $C$ 에 속한 노드들은 데이터를 전송받아야하므로 이 비용은  $\sum_c T_r(\text{cov}_{\min}(Q_L)) \cdot j/r$ 이다. 또한,  $C$ 에 속한 노드들은 분할 조인을 재귀적으로 적용한다. 또한,  $\text{qchild}(\text{cov}_{\min}(Q_R))$  중  $C$ 에 속하지 아니한 노드들은 데이터를 수집하고 이를  $\text{cov}_{\min}(Q_R)$ 로 전송한다.

[0070] 위의 수학식 10을 이용하여 최소 비용의 분할 조인이 이루어지는  $C$ 를 구하는 것은 동적 계획법(dynamic

programming)을 이용해도,  $n$ 이  $\text{cov}_{\min}(Q_R)$ 의 자손 노드들의 수라고 할 때, 자손 노드들 중 세미조인이 수행되는 모든 가능한 위치를 찾는 문제로 시간 복잡도가  $O(2^n)$ 으로 많은 시간이 필요하다. 따라서, 본 발명에서는 넓이 우선 탐색 (breadth first traversal)을 기반한 욕심쟁이 접근 (greedy approach) 방법을 이용한 최소 비용 계산 모듈을 고안하였다.

[0071] 본 발명에서 고안한 분할 조인 욕심쟁이 모듈의 의사코드(pseudo code)를

```

Procedure compute_partial($\text{cov}_{\min}(Q_R), \text{cov}_{\min}(Q_R)$)
begin
1. Queue Que
2. Que.enqueue($\text{cov}_{\min}(Q_R)$)
3. $\text{Cost}_{\text{patial}} := 0$;
4. while(Que is not empty) {
5. aNode := Que.dequeue()
6. $C := \emptyset, \text{cost}_C := 0$
7. for each $c \in \text{qchild}(\text{aNode})$ do {
8. if($T(c) > T_r(\text{cov}_{\min}(Q_L)) \cdot \frac{i}{r} + T(c) \cdot \text{SJS}_{L,A}$) {
9. $C := C \cup c$
10. $\text{cost}_C += T_r(\text{cov}_{\min}(Q_L)) \cdot \frac{i}{r} + T(c) \cdot \text{SJS}_{L,A}$
11. } else {
12. $\text{Cost}_{\text{patial}} += (\text{Cost}_{\text{gathering}}(c) + T(c))$
13. }
14. }
15. if($T_s(\text{cov}_{\min}(Q_L)) \cdot \frac{i}{r} < \sum_{c \in C} T(c) - \text{cost}_C$) {
16. Que.enqueue(C)
17. $\text{Cost}_{\text{patial}} += (T_s(\text{cov}_{\min}(Q_L)) \cdot \frac{i}{r} + \text{cost}_C)$
18. } else {
19. for each $c \in C$ do
20. $\text{Cost}_{\text{patial}} += (\text{Cost}_{\text{gathering}}(c) + T(c))$
21. }
22. }
end

```

[0072]

[0073] 와 같이 나타낸다. 기본적으로 큐(Queue)를 이용한 넓이 우선 탐색을 수행하는데 있어서 qchild()에 속한 자식 노드  $c$ 가 세미 조인을 수행한다면,  $c$ 는  $\Pi_A(L)$ 를 받고  $R_c \triangleright L$ 의 결과를 반환한다. 따라서,  $c$ 에서 세미조인을 수행하지 않고 바로 데이터를 전송하는 전송 비용  $T(c)$ 는  $\Pi_A(L)$ 의 수신비용 과 세미조인 결과 전송비용  $T(c) \cdot \text{SJS}_{L,A}$ 의 합보다 크다면 (Line 8)  $c$ 는 세미조인이 수행될 수 있는 후보자식이 된다. 이러한 후보 자식들을  $C$ 로 모은다 (Line 9). 또한 이 비용을  $\text{cost}_C$ 로 합한다. 이러한 후보 자식 노드들이 세미조인을 수행하려면 부모노드 (aNode)가  $\Pi_A(L)$ 를 전송해야 한다. 따라서,  $\Pi_A(L)$ 를 발신하는 비용( =  $T_s(\text{cov}_{\min}(Q_L))$ )보다 세미조인을 하는데 따른 이득( =  $\sum_c T(c) - \text{cost}_C$ )보다 작다면 (Line 15-17)  $C$ 에 속한 노드들에서 세미조인을 수행한다. 또한, 이에 따른 비용을 합산한다(Line 17). 또한,  $C$ 를 큐에 넣어 (Line 16)  $C$ 의 자식들에 대하여 추가로 고려한다. 그러지 않다면 (Line 18-21)  $C$ 에서 세미조인을 수행하는 것이 좋지 않다는 것을 나타낸다. 따라서  $C$ 에서 데이터를 모아서 부모 노드로 보내는 비용을 합산한다(Line 20, 21).

[0074] 위에서 설명한 사이드 조인 및 분할 조인에서는  $\text{cov}_{\min}(Q_L \cup Q_R)$ 에서  $\Pi_A(L)$ 를  $\text{cov}_{\min}(Q_R)$ 로 보낸다. 따라서,  $\Pi_A(L)$  대신 요약정보인 시놉시스(synopsis)를 활용한다면 이 전송비용을 줄일 수 있을 것이다. 따라서, 본 발명에서는 시놉시스를 사용하는 조인 방법인 시놉시스 조인을 고안하였다. 본 발명에서는 시놉시스로 블룸필터(bloom filter) [참고문헌: B. H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7), 1970.]를 사용하였다. 블룸필터는 크기가  $m$ 인 비트열과  $k$ 개의 해시함수들로 구성된다. 각 해시함수는 임의 값을  $[1, m]$  사이의 정수값으로 변환한다. 따라서, 하나의 값은 비트열에서  $k$ 개 해시함수에 의한  $k$ 개의 위치들의 비트를 1로 하여 표현한다. 따라서, 특정 값이 블룸필터상 존재하는지 아닌지는 특정값의 비트열의 해시 함수의  $k$ 개 위치의 비트들이 모두 1인지를 확인하면 된다. 이러한 특성을 이용하여 조인 연산에 활용할 수 있다.

[0075] 시놉시스 조인은 분할조인과 유사하나  $\Pi_A(L)$ 대신  $\Pi_A(L)$ 의 블룸필터를  $\text{cov}_{\min}(Q_L \cup Q_R)$ 에서  $\text{cov}_{\min}(Q_R)$ 로 보내는 것

이 다르다.

[0076] 이때,  $m$  크기의  $\Pi_A(L)$ 를 위한 블룸필터의 전송 비용  $T_{sysL}$ 은 수학적식 11과 같이 표현할 수 있다.

### 수학적식 11

[0077]  $T_{sysL} = m \cdot P(\text{cov}_{\min}(Q_L))$

[0078] 여기서  $P(\text{cov}_{\min}(Q_L))$ 은  $\text{cov}_{\min}(Q_L)$  에서 데이터를 전송할 확률이다.

[0079] 어떤 노드  $s$ 에서 데이터를 전송할 확률  $P(s)$ 는  $1 - \text{Prob}(s \text{가 데이터를 전송하지 않음})$ 가 같다. 트리 구조의 네트워크에서 노드  $s$ 가 데이터를 전송하지 않으려면 노드  $s$ 의 측정값이 선택조건  $P_0$ 를 만족하지 않고  $s$ 의 자식들로  $s$ 로 데이터를 전송하지 않는 것이다. 따라서  $P(s)$ 는 수학적식 12와 같이 표현할 수 있다.

### 수학적식 12

[0080]  $P(s) = 1 - (1 - \text{sel}(PQ)) \cdot \prod_{c \in \text{child}(s)} (1 - P(c))$

[0081] 시놉시스 조인은 분할 조인과 유사함으로 수학적식 11을 기반으로 시놉시스 조인의 비용을 수학적식13과 같이 표현할 수 있다.

### 수학적식 13

$$\begin{aligned} Cost_{\text{partial,synopsis}}(\text{cov}_{\min}(Q_L), \text{cov}_{\min}(Q_R)) = & \min_{C \subseteq \text{qchild}(\text{cov}_{\min}(Q_R))} \{ \\ & T_{sysL,s} + \sum_{c \in C} T_{sysL,r} + \\ & \sum_{c \in C} T(c) \cdot SJS_{L,A} + \\ & \sum_{c \in C} Cost_{\text{partial,synopsis}}(\text{cov}_{\min}(Q_L), c) + \\ & \sum_{c' \in \text{qchild}(\text{cov}_{\min}(Q_R)) - C} (Cost_{\text{gathering}}(c') + T(c')) \} \\ & , \text{where } T_{sysL,s} = 0 \text{ if } C = \emptyset \end{aligned}$$

[0082]

[0083] 수학적식 13에서  $T_{sysL,s}$ 와  $T_{sysL,r}$ 은  $\Pi_A(L)$ 의 블룸필터의 발신 및 수신 비용을 나타낸다. 또한, 시놉시스 조인 방식을 위한 욕심쟁이 모듈은 도 6의 Line 8,10,15,17의  $T_r$ 과  $T_s$ 를  $T_{sysL,s}$ 와  $T_{sysL,r}$ 로 대치함으로써 구할 수 있다.

[0084] 시놉시스 조인에서는  $Q_L$  영역의 측정값들인  $L$ 이  $\text{cov}_{\min}(Q_L \cup Q_R)$ 로 보내진다. 이 경우  $L$ 에는 조인 결과에 참여하지 않은 것들이 있으므로 에너지를 소비하게 된다. 또한, 본 발명에서는 시놉시스 조인을 확장하여 완전시놉시스 조인 (full synopsis join) 방법을 고안하였다. 완전 시놉시스 조인에서는  $L$  대신에  $\Pi_A(L)$ 의 블룸필터를  $\text{cov}_{\min}(Q_L)$ 에서  $\text{cov}_{\min}(Q_L \cup Q_R)$ 을 전송하고  $\text{cov}_{\min}(Q_L \cup Q_R)$ 을  $\text{cov}_{\min}(Q_R)$ 로 전송한다. 완전 시놉시스 조인은 다음의 4 단계로 구성된다.

[0085]

1)  $Q_L$ 에서  $\Pi_A(L)$ 의 블룸필터를 생성

[0086]

2) 생성된 블룸필터를  $\text{cov}_{\min}(Q_R)$ 로 전송하여  $Q_R$ 에서 시놉시스 조인을 수행

[0087]

3) 2)의 조인 결과를  $\text{cov}_{\min}(Q_L \cup Q_R)$ 로 보내고  $\text{cov}_{\min}(Q_L \cup Q_R)$ 에서 이를 이용하여 블룸필터를 생성하고 이를  $\text{cov}_{\min}(Q_L)$ 로 전송하고  $Q_L$ 에서 시놉시스 조인을 수행

[0088]

4) 3)의 결과를  $\text{cov}_{\min}(Q_L \cup Q_R)$ 로 보내고  $\text{cov}_{\min}(Q_L \cup Q_R)$ 에서는 두개의 시놉시스 조인 결과를 베이스 스테이션으로 전송

[0089]

첫번째 단계에서 서로 다른 두 데이터 집합  $A1$ 과  $A2$ 의 합집합  $A$ 의 블룸필터는  $A1$ 의 블룸필터와  $A2$ 의 블룸필터의

OR 연산 결과와 동일함으로  $Q_L$  상의 한 노드는 자식 노드들로 받은 블룸필터들을 OR연산으로 합치고 자신의 측정값을 블룸필터에 넣음으로써 블룸필터를 생성할 수 있다. 따라서 이 첫단계의 비용은 수학적 식 14와 같이 나타낼 수 있다.

#### 수학적 식 14

$$Cost_{con}(cov_{min}(Q_L)) = \sum_{c \in child(cov_{min}(Q_L))} (Cost_{con}(c) + P(c) \cdot m)$$

두번째 단계에서는 블룸필터가  $cov_{min}(Q_R)$ 로 전송되고 시놉시스 조인이 수행됨으로 이 비용은 수학적 식 15과 같다.

#### 수학적 식 15

$$D_L \cdot T_{sysL} + D_R \cdot T_{sysL} + Cost_{partial,synopsis}(cov_{min}(Q_L), cov_{min}(Q_R))$$

세번째 단계에서는 두번째 단계의 결과인  $R \supset L$ 를  $cov_{min}(Q_L \cup Q_R)$ 로 전송하고  $cov_{min}(Q_L \cup Q_R)$ 에서 이 결과를 이용하여  $\Pi_A(R \supset L)$ 의 블룸필터를 생성하고 이를  $cov_{min}(Q_L)$ 로 전송하고  $L \supset R$ 을 위한 시놉시스 조인을 수행한다. 이 비용은 수학적 식 16과 같다.

#### 수학적 식 16

$$D_R \cdot T(cov_{min}(Q_R)) \cdot SJS_{L.A} + D_L \cdot T_{sysR} + Cost_{partial,synopsis}(cov_{min}(Q_R), cov_{min}(Q_L))$$

마지막으로 세번째 단계의 결과인  $L \supset R$ 을  $cov_{min}(Q_L)$ 에서  $cov_{min}(Q_L \cup Q_R)$ 로 전송하고  $cov_{min}(Q_L \cup Q_R)$ 에서 두번째 단계의 결과인  $R \supset L$ 과 세번째 단계의 결과  $L \supset R$ 을 베이스 스테이션으로 전송한다. 이 비용은 수학적 식 17과 같다.

#### 수학적 식 17

$$D_L \cdot T(cov_{min}(Q_L)) \cdot SJS_{R.A} + D_C \cdot (T(cov_{min}(Q_L)) \cdot SJS_{R.A} + T(cov_{min}(Q_R)) \cdot SJS_{L.A})$$

본 발명에서는 위에서 상기한 베이스조인, 커버조인, 사이드조인, 분할 조인, 시놉시스조인, 완전 시놉시스 조인 중 상황에 따라 가장 효율적인 조인 방법을 선택하기 위하여 도 5에 도시한 바와 같은 모듈로 구성된 시스템을 구성하였다.

도 5를 참조하면, 질의문에 대한 구문분석 모듈이 수행되고 이를 기반으로 센서 네트워크상에서 중요한  $cov_{min}(Q_L)$ ,  $cov_{min}(Q_R)$ ,  $cov_{min}(Q_L \cup Q_R)$ 을 결정하는 결정기 모듈이 수행되고 이 위치를 기반으로 각 비용계산기들은 각 조인방법의 비용을 계산하고 계산 값을 최소 비용 선택모듈로 전송한다. 최소비용 선택 모듈에서는 이중 가장 좋은 방법을 선택하고 이를 질의 실행 모듈에 통보하여 조인 연산을 수행하게 함으로써 조인 결과를 얻을 수 있다.

본 발명의 효율성을 측정하기 위하여 조인 선택률  $SJS_{R.A}$ 의 값을 변화시키며 각 조인 방법들의 예측비용 및 실제 수행 비용을 비교하였다. 실험에 사용된 파라미터는 다음과 같다. 또한, 실험에서 각 센서 측정값의 크기는 44byte, 조인 속성값의 크기는 8byte, 시놉시스 조인에 사용되는 블룸필터의 크기는 30byte로 하였다. 실험에서는 1000 단위 시간 동안의 예측비용 및 실제 비용을 누적하였다. 또한, 실험에서는  $Q_R$ 쪽에서 세미조인이 먼저 수행되는 방법에 접두사 R을 붙였으면 이와 반대이면 접두사 L을 붙였다.

표 1

| Parameter  | Default value |
|------------|---------------|
| $sel(P_Q)$ | 0.5           |
| $SJS$      | 0.5           |
| $D_C$      | 1             |
| $D_L$      | 2             |
| $D_R$      | 2             |
| $ Q_L $    | 10            |
| $ Q_R $    | 10            |

[0100]

[0101]

도 6은 실험결과에 따른 예측 비용 및 실제 비용을 나타낸다. 결과에서 보듯이 다양한  $SJS_{R,A}$ 에 대하여 가장 최소 비용이라고 예측한 방법이 실제 수행에서도 가장 적은 실행 비용을 나타냄을 볼 수 있다.

[0102]

전술한 바와 같은 본 발명의 상세한 설명에서는 구체적인 실시예에 관해 설명하였다. 그러나 본 발명의 범주에서 벗어나지 않는 한도 내에서는 여러 가지 변형이 가능하다. 본 발명의 기술적 사상은 본 발명의 전술한 실시예에 국한되어 정해져서는 안 되며, 특허청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

### 도면의 간단한 설명

[0103]

도 1은 특정 질의 영역에 속한 센서들이 무선 통신을 통하여 임의의 노드 d 까지 측정값을 전송하는 예를 나타내는 개념도.

[0104]

도 2는 정의 1과 정의 2의 예를 나타내는 개념도.

[0105]

도 3은 조인 연산을 포함한 질의 처리를 위한 기본적인 3가지 조인 처리 방안을 나타내는 개념도.

[0106]

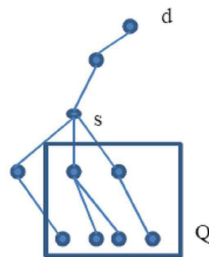
도 4는 분할 조인 방법의 예를 나타내는 개념도.

[0107]

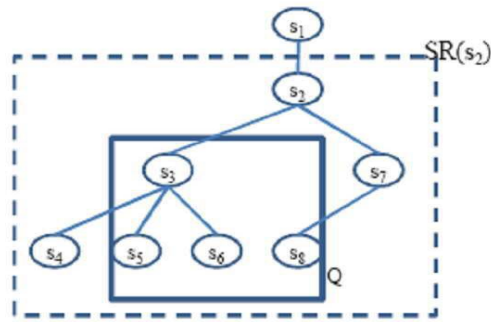
도 5는 본 발명에 따른 비용 기반 최적 조인 처리 시스템을 나타내는 블록도.

### 도면

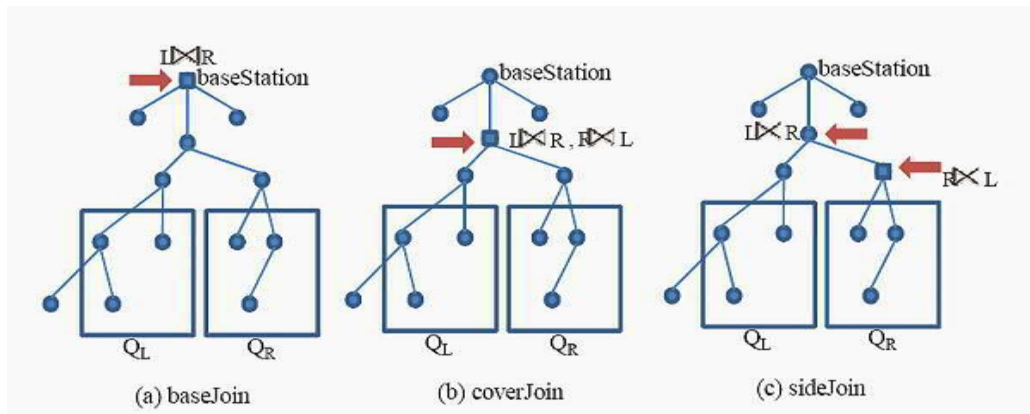
#### 도면1



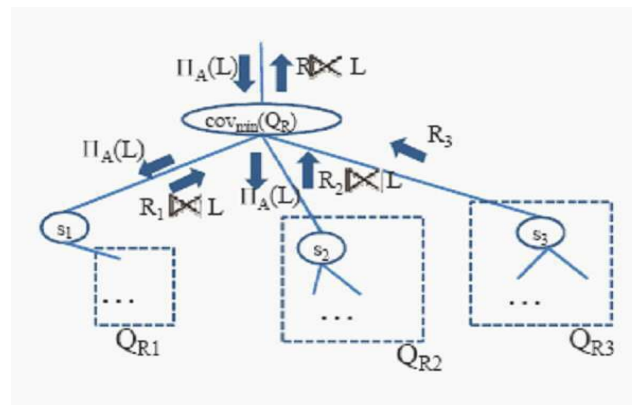
도면2



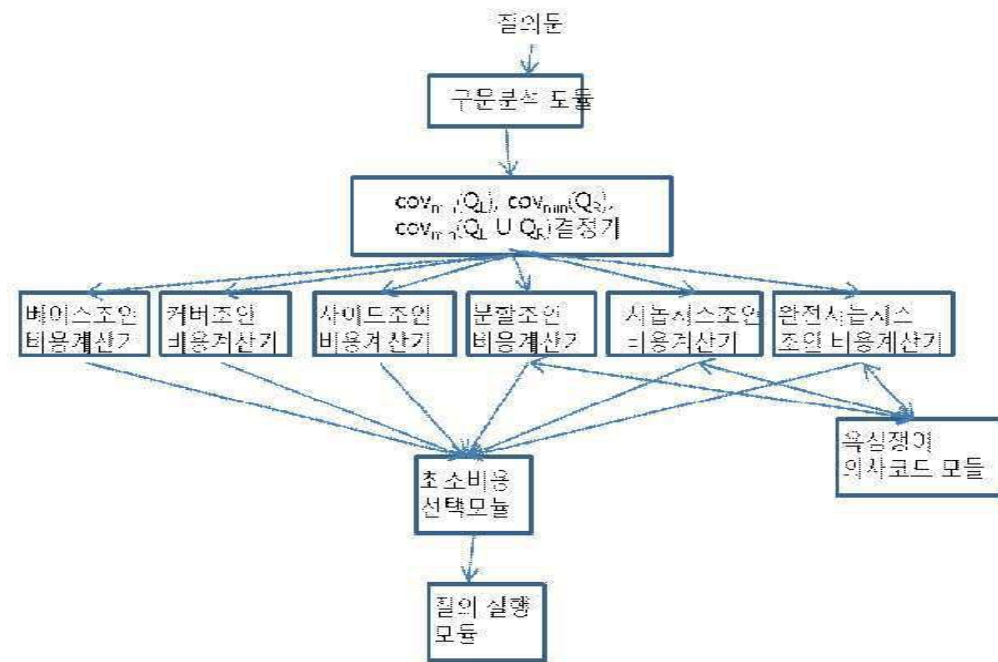
도면3



도면4



도면5



도면6

