

논문 2005-42TE-1-6

# 분산 이동 객체 데이터베이스의 효율적인 범위 질의 처리

## (An Efficient Range Query Processing of Distributed Moving Object)

전 세 길\*, 우 찬 일\*\*

(Se Gil Jeon and Chan Il Woo)

### 요 약

최근 들어 이동 통신에서 이동하는 고객의 위치정보와 관련된 서비스가 중요하게 부각되고 있다. 이동 객체의 경우 갱신 연산이 많고 부하가 특정 지역에 집중되는 특징이 있다. 이러한 LBS 응용에서 시공간 영역질의는 중요한 부분이고, 질의 처리 속도의 개선을 위해서 디스크 입출력 시간을 최대한 줄일 필요가 있다. 본 논문에서는 갱신연산을 최소화하기 위해 고안되어진 시공간 인덱스 구조로 비 균등 2단계 격자 인덱스 구조를 적용한다. 또한 질의처리를 향상을 위해서 공간 관련성과 시간 관련성을 이용한 스케줄링 기법과 time zone 개념을 사용하여 시공간 개념이 결합된 질의 처리 기법을 제안한다. 다양한 질의 범위를 사용하여 실험한 결과 제안된 방법이 기존의 방법보다 우수함을 알 수 있었다.

### Abstract

Recently, the location based service for moving customers is becoming one of the most important service in mobile communication area and for moving object applications, there are lots of update operations and such update loads are concentrated on some particular area unevenly. The primary processing of LBS application is spatio-temporal range queries and to improve the throughput of spatio-temporal range queries, the time of disk I/O in query processing should be reduced. In this paper, we adopt non-uniform two-level grid index structure, which are designed to minimize update operations. We propose query scheduling technique using spatial relationship and time relationship and a combined spatio-temporal query processing method using time zone concepts to improve the throughput of query processing. Some experimental results are shown for range queries with different query range to show the performance tradeoffs of the proposed methods.

**Keywords :** Location Based Service, Query Scheduling, Location Based Query Processing.

### I. 서 론

이동통신 사용자의 급증으로 휴대폰 가입자의 위치 관련 서비스가 더불어 증가하고 있으며, 이러한 서비스를 위치 기반 서비스(LBS : Location Based Service)라 한다. 위치 기반 서비스는 위치 결정, 위치 응용 플랫폼, 위치 응용 서비스의 3가지 중요한 기술로 구성되며, 대부분의 위치 기반 서비스는 시간 및 공간에 관련하여 각 시간대 별로 위치 정보를 저장하게 된다. 그러나 이

동 객체의 경우 위치 정보를 처리하기 위하여 위치 정보가 수시로 변한다는 것과 공간 질의가 유사한 지역에 빈번하게 발생해 중복 연산이 많은 문제점이 있다<sup>1,2)</sup>.

시공간 데이터베이스에 대해서 발생하는 질의의 유형은 특정 시간 내에서 특정 객체가 움직인 궤적을 질의 하는 경우와 특정 시간 내에서 특정 지점에 머문 객체들을 구하는 질의가 있다. 또한 특정 시간 범위에 특정 영역 내에 있던 객체들을 구하는 질의가 있을 수 있으며, 이를 범위 질의라 한다. 범위 질의는 사각형이나 원형 형태로 구성되며 각 범위 질의 간에는 유사한 겹침 영역이 발생할 수 있다. 그러나 동일한 영역에 대한 질의 간에는 질의 결과를 공유할 수 있으며 같은 데이터에 대한 디스크 재 검색 시간을 줄임으로써 질의 처리를 향상시킬 수 있다. 질의 처리 큐에 입력된 질의

\* 정회원, 건국대학교 생명 분자 정보학 센터  
(Bio Molecular Informatics Center, Konkuk University)

\*\* 종신회원, 서일대학 정보통신전공  
(Dept. of Information and Communication Engineering, Seoil College)

접수일자: 2004년10월27일, 수정완료일: 2004년11월22일

에 대해서 유사한 영역을 함께 처리하기 위해서는 질의 우선순위를 재조정 할 수 있는 질의 스케줄링 기법이 필요하다.

본 논문에서는 빈번한 갱신연산을 줄이고 대량의 데이터를 효율적으로 처리하기 위해 비 균등 2단계 격자 인덱스 구조와 time zone 인덱스를 사용한 질의 스케줄링 기법을 제안한다. 비 균등 2단계 격자 인덱스 구조에서 첫 번째 단계는 크기를 유동적으로 구성하고 두 번째 단계에서는 각 셀 간의 순서를 Z-ordering 기법을 사용한다. 또한 시간적 인덱스 구조로 time zone 기법을 사용하여 시간대에 따라 4개의 zone에 나누어 필터링하여 저장한다<sup>[3-6]</sup>. 본 논문에서 제안한 스케줄링 기법에서는 질의 간에 공간적으로 겹치는 셀의 개수에 따른 공간 관련성과 각 질의에 대해서 시간 범위가 겹치는 정도를 나타내는 시간 관련성을 적용한다. 또한 다중 노드를 위하여 질의와 각 노드간의 겹침 관계를 적용한 노드 관련성도 함께 적용하여 기존의 이동 사용자 패턴을 적용한 데이터를 이용하여 스케줄링 시 적용되는 각각의 가중치 영향에 대해서 실험 평가 한다.

## II. 관련 연구

### 1. 이동 객체

이동 객체는 시간의 변화에 따라 객체의 공간 위치정보가 연속적으로 변경되는 객체로 크게 이동점과 이동영역으로 나눌 수 있다<sup>[7]</sup>.

그림 1은 시간에 따라 변화하는 이동 점 객체와 이동 영역 객체의 움직임을 3차원 공간에 표현하였다. 사람, 자동차, 모바일 폰 등과 같은 이동 점 객체는 시간의 변화에 따라 자신의 공간 위치만 변화하지만 이동 영역 객체는 시간에 따라 자신의 공간 위치뿐만 아니라 모양까지 변화하는 것으로 한 국가의 행정 구역, 몸속에 퍼진 암세포의 상태 등을 들 수 있다. 이동 객체와 관련한 질의들은 객체의 이동 경로나 방향, 거리, 시간 등에 대한

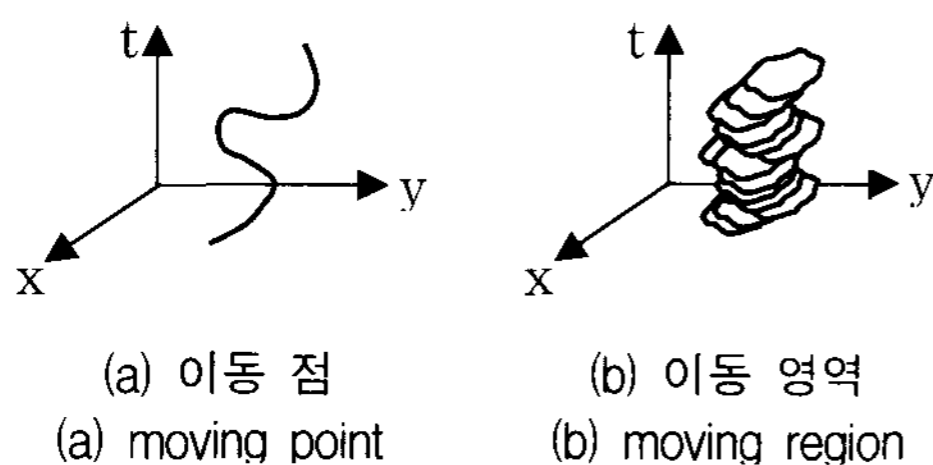


그림 1. 이동 객체  
Fig. 1. Moving object.

질의들로 표현할 수 있으며 이러한 질의는 이동하는 특정객체에 대한 위치를 확인 가능하게 한다. 본 논문에서의 이동 객체는 이동 점 객체 중 휴대폰 가입자로 한정한다.

### 2. 기존의 시공간 인덱스 구조

시공간 객체의 위치정보를 인덱싱 하기 위해 제안된 기존의 구조는 RT-트리, 3D R-트리, TB-트리, TPR-트리 등이 있다<sup>[8-10]</sup>. 이러한 구조들은 대량의 이동객체의 움직임에 따라 많은 노드의 분할 및 합병이 발생하여 삽입, 삭제 시 디스크에 많은 부하를 초래하며 처리 시간도 기하급수적으로 증가한다.

## III. 시공간 모델링

### 1. 2단계 공간 인덱싱 기법

이동 객체의 이동정보를 저장하기 위한 고정 그리드 기반 2단계 인덱싱 구조는 그림 2에 나타내었다. 1단계는 전체 지역을 담당하는 최 상위 단계의 비 균등 그리드 구조로 구성되고 2단계는 균일한 크기로 분할되는 균등 그리드 구조로 구성된다. 1단계에서는 인구 밀도나 기타 정보에 따른 지리적인 지역으로 25.6km 단위의 사각형 형태로 분할된다. 이러한 사각형을 매크로 셀(macro-cell)이라 하며 지역에 따라 크기가 다르게 나타나게 된다.

### 2. Time Zone 기법을 이용한 시간 인덱싱

이동 객체의 위치정보는 계속적으로 변화하므로 샘플링 간격을 길게 한다고 해도 대량의 정보로 인하여

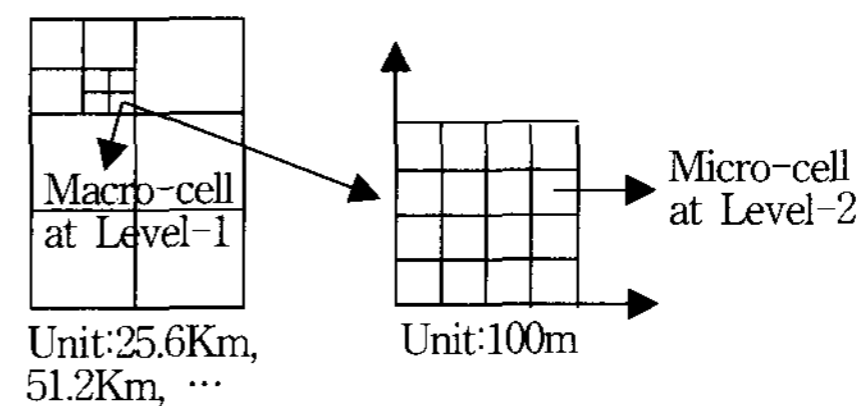


그림 2. 비 균등 2단계 그리드 구조  
Fig. 2. A non-uniform 2 level grid structure.

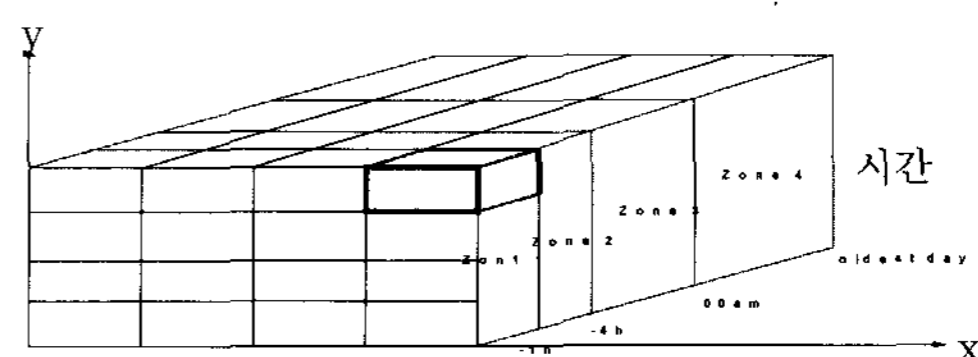


그림 3. 4개의 Time Zone이 적용된 시공간  
Fig. 3. Spatio-temporal model adopted 4 time zone.

데이터 검색 시간이 기하급수적으로 증가하게 된다. 따라서 본 논문에서는 위치 정보 저장 테이블을 zone 이라는 시간의 범주에 따라 시간대 별로 서로 다른 테이블에 저장한다.

그림 3은 4개의 time zone 시공간 데이터 모델을 나타내고 X, Y축으로 이루어진 각 면은 Z-ordering을 이용한 고정 그리드이며 Z축은 시간에 따른 변화를 나타낸다. Z축의 zone1은 100m 이상 움직였으며 현재 시간으로부터 1시간 전의 위치정보를 저장하고 있으며, zone2는 400m 이상을 움직였으며 1시간에서 4시간사이의 위치 정보를 저장한다. zone3는 3.2km 이상을 움직이고 4시간 이후부터 그날의 시작까지의 위치 정보를 저장하며, zone4는 이전날짜의 데이터로써 백업 역할을 담당한다.

#### IV. 스케줄링 기법을 이용한 시공간 질의 처리

##### 1. 시공간 질의 유형

그림 4는 다중 노드에서 원형 질의나 사각형 질의 시에 발생 할 수 있는 질의 형태를 나타낸다. 여기서 사각형 질의 일부가 해당 노드에 포함되는 경우에는 부분적으로 포함된 사각형 질의와 노드 좌표계 간의 겹치는 좌측 하단 좌표와 우측 상단 좌표를 구하는 작업이 필요하며 원형 질의가 해당 노드에 일부 포함되는 상태는 모든 객체와 질의 중심점 간의 거리 계산이 필요하다.

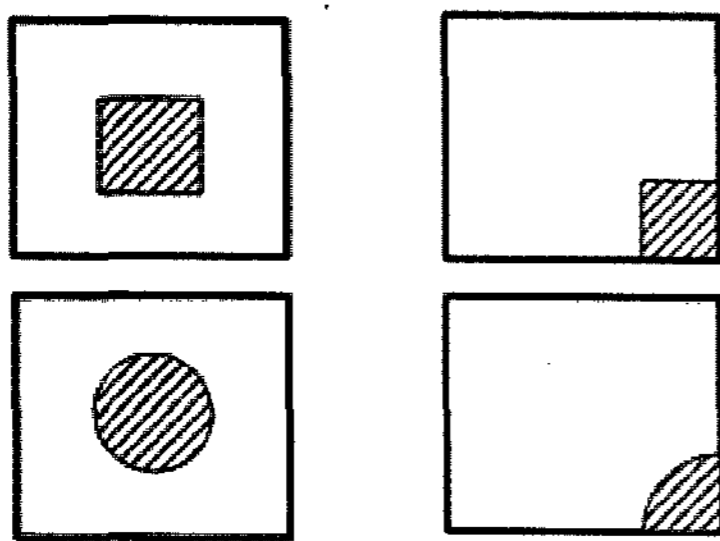


그림 4. 다중 노드에서 범위 질의 유형  
Fig. 4. Range query type for multiple node.

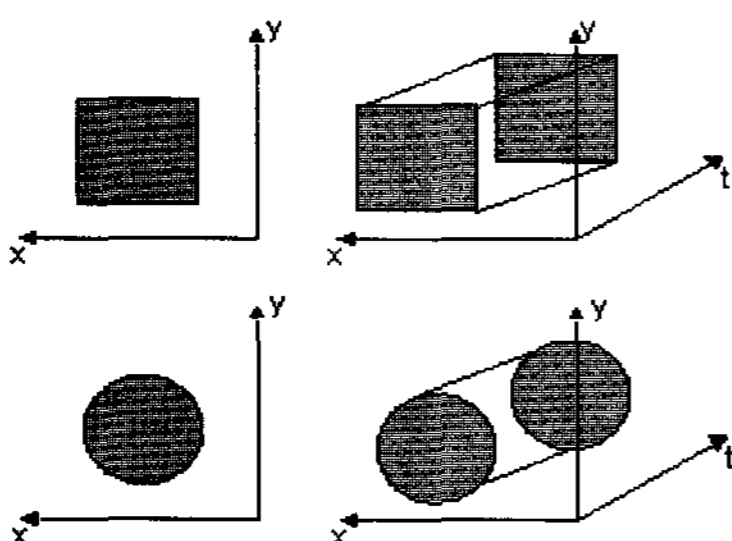


그림 5. 시공간 질의 유형  
Fig. 5. Typical spatio-temporal query shapes.

그림 5는 시공간 질의 유형을 나타낸 것으로 2차원 그래프는 범위 내에 있는 객체를 검색하는 질의어를 나타내고, 3차원 그래프는 특정 시간 범주를 가지는 범주 질의를 나타낸다.

##### 2. 시공간 질의 처리 흐름

LLDS(Long-term Location Data Server)는 하나의 마스터 LDP(Location Data Processor)와 여러 개의 슬레이브 LDP들로 구성되고, 위치 정보의 과거 궤적 정보를 저장하는 역할을 한다. 질의는 마스터 LDP의 질의 큐에 계속적으로 쌓이고 마스터 LDP는 질의가 현재 영역이면 현재 노드에서 처리하고 아니면 네트워크를 통해 다른 노드로 질의를 전송한다. 마스터 노드에 포함되는 질의의 경우 질의 형태에 따라 해당되는 질의 처리 알고리즘을 호출한다. 각 질의 중에서 현재 마스터 노드에 부분적으로 포함되어 있는 질의는 네트워크를 통해 다른 노드로 전송한다.

##### 3. 시공간 질의 스케줄링

시공간 질의 스케줄링은 질의 큐에 입력된 일정 질의를 시공간적으로 다른 처리기와의 관련성을 고려하여

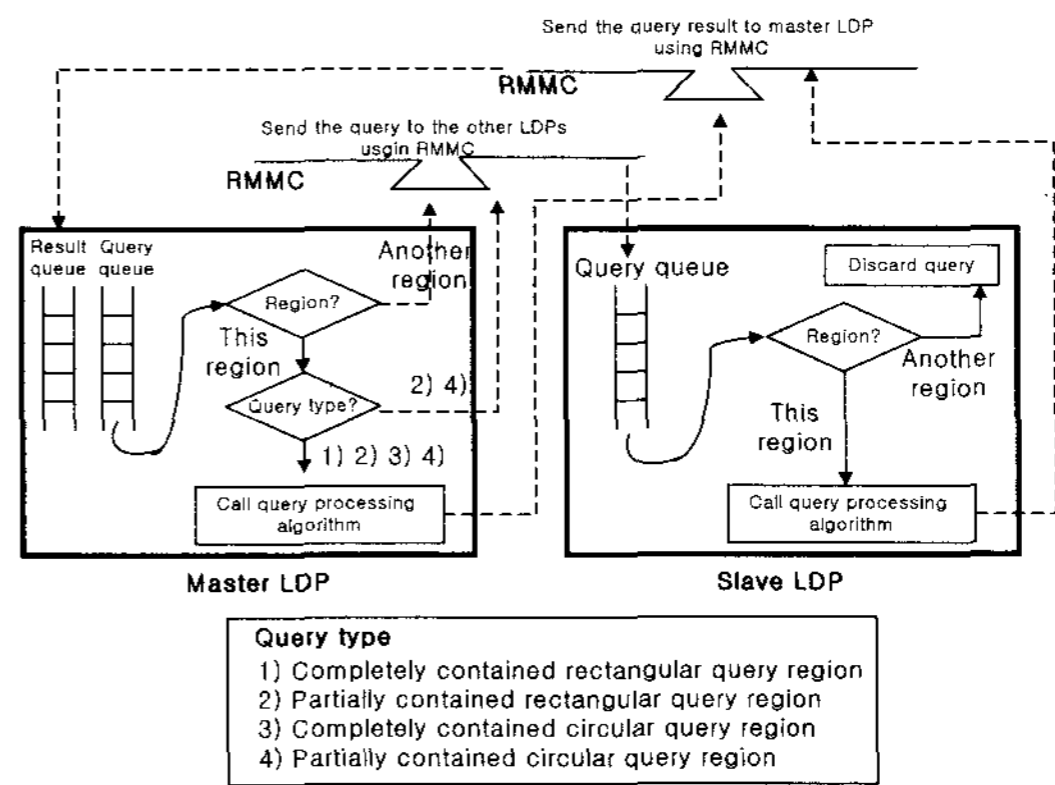


그림 6. LLDS 질의 처리 흐름  
Fig. 6. LLDS query processing flow.

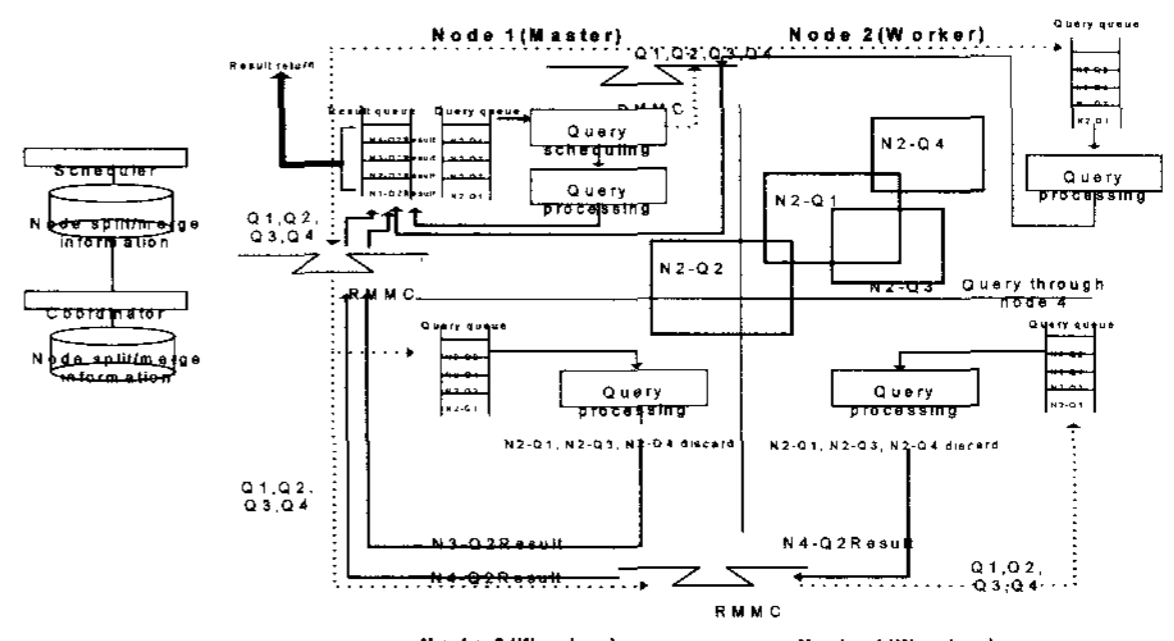


그림 7. Master 노드가 스케줄링을 담당하는 구조  
Fig. 7. Scheduling structure controlled master node.

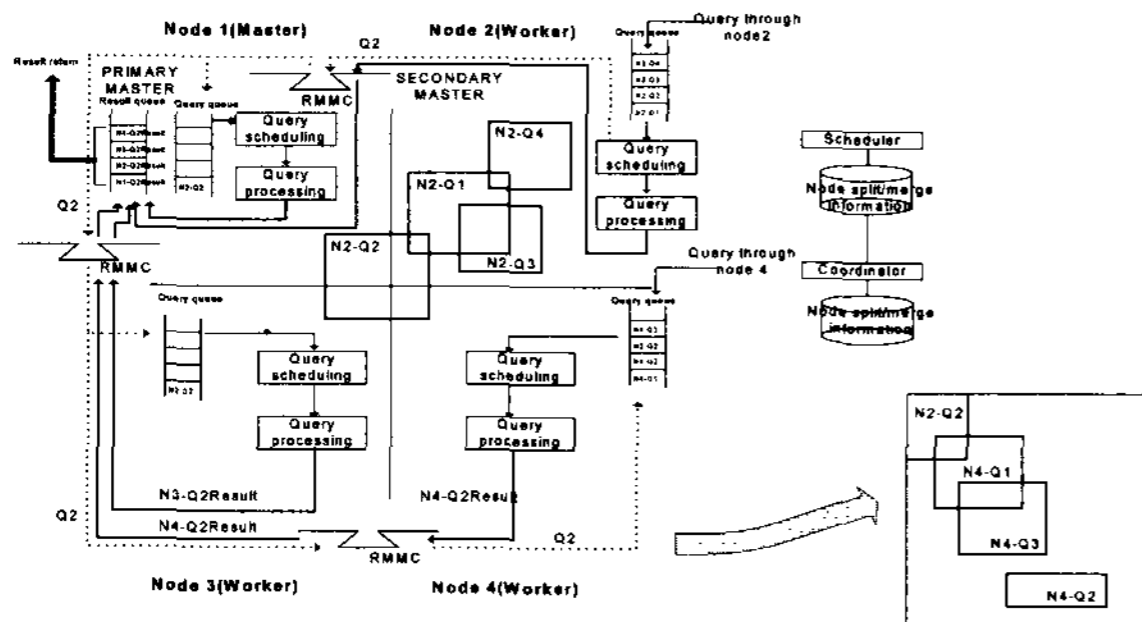


그림 8. 각 노드가 스케줄링을 담당하는 구조  
Fig. 8. Scheduling structure controlled each node.

처리 순서를 재조정함으로써 질의 처리률을 높이는 방법이다. 본 논문에서는 질의 처리 순서 결정을 위해 위치 관련성, 시간 관련성, 노드 관련성을 고려한다.

그림 7은 LDP Master 노드가 스케줄링을 담당하는 구조이다. 이 구조에서는 Master 노드로 입력된 질의를 Master 노드가 여러 관련성을 적용하여 스케줄링 하여 각 노드로 전송한다.

그림 8은 스케줄링 기능을 가지고 있는 각 노드에서 독립적으로 질의를 수용하는 것을 나타내고 있다. 일반적으로 질의 집중 지역에 대해 수행된 질의들은 공간적으로 서로 겹쳐 있는 경우가 많고 겹쳐진 지역 내에 존재하는 셀을 사용할 확률이 크며 질의 간의 위치 관련성은 영역 질의 간의 공간 관련성을 나타내는 기준이 된다.

가. 셀 번호를 이용한 위치 관련성

정의 1) 질의 간 위치 관련성

$$LR = \text{COUNT}(\text{Overlap\_Cell\_No}(Q_{\text{previous}}, Q_{\text{current}}))$$

LR(Location Relationship) = 질의 Q의 위치 관련성

Overlap\_Cell\_No(Q1, Q2) = 질의 Q1, Q2가 겹치는 셀 번호

Qprevious = 위치 관련성 비교 대상 기준 질의

Qcurrent = 현재 대기 중인 질의

질의 간의 위치 관련성은 질의들이 겹치는 셀의 수에 비례하고 위치 관련성이 크다는 것은 해당 질의가 서로 인접하다는 것을 의미하며 같은 데이터를 공유 할 확률이 높다. 위치 관련성이 높은 질의들을 처리하게 되면 인접 지역에 집중되는 영역 질의를 같이 처리하므로 데이터 검색 시 필요한 탐색의 중복을 최소화 할 수 있다.

나. Time Zone 개념을 이용한 시간 관련성

질의 간의 시간 관련성은 주어진 질의의 시간 범위 간의 겹치는 정도를 의미한다. 공간 관련성과 마찬가지로

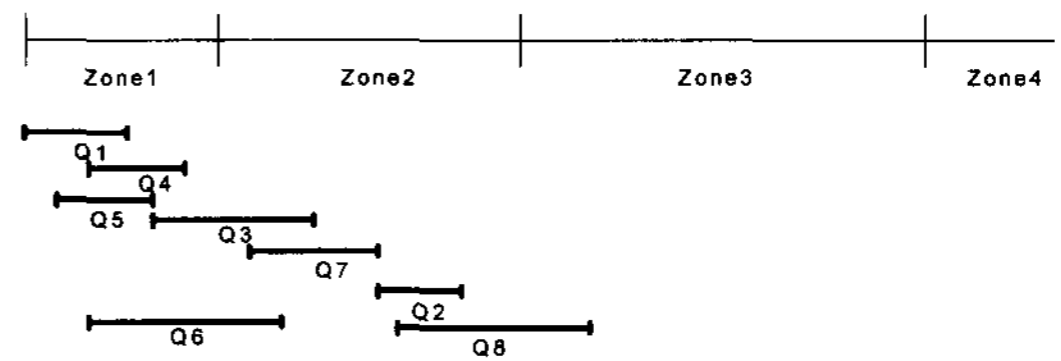


그림 9. Time zone에서 질의 간 시간 관련성  
Fig. 9. Time relationship inter query using time zone.

로 시간 범위가 주어진 궤적 질의의 경우 서로 시간적으로 겹쳐 있는 경우가 많이 발생할 수 있으며 데이터를 공유할 확률이 높다. 그림 9에서 각 zone은 하나의 테이블을 의미하므로 같은 zone내에 있는 질의들을 인접하여 처리하는 것이 좋고 기준 질의에 대해서 시간적으로 겹치는 정도가 높은 질의를 우선 처리하는 것이 중복 탐색을 최소화 할 수 있다.

정의 2) 질의 간 시간 관련성

$$TR = \text{Overlap\_Time}(Q_{\text{previous}}, Q_{\text{current}})$$

TR(Time Relationship) = 질의 Q의 위치 관련성

Overlap\_Time(Q1, Q2) = 질의 Q1과 Q2가 겹치는 시간

Qprevious = 시간 관련성 비교 대상 기준 질의

Qcurrent = 현재 대기 중인 질의

다. 노드 관련성과 가중치 조절

질의 스케줄링을 위해서는 질의 처리 큐에 대기하고 있는 질의들의 우선순위를 결정해야 하며 본 논문에서는 여러 노드에 대해 겹쳐서 수행되는 범위 질의를 사용하므로 노드간의 관련성도 고려하여야 한다.

정의 3) 질의 간 노드 관련성

$$NR = \text{COUNT}(\text{Overlap\_Node\_No}(Q_{\text{current}}))$$

NR(Node Relationship) = 질의 Q의 노드 관련성

Overlap\_Node\_No(Q1) = 질의 Q1와 겹치는 노드 번호

Qcurrent = 현재 대기 중인 질의

정의 4) 각 관련성간의 가중치 조절후의 질의 관련성

$$QR = LR * W_L + TR * W_T + NR * W_N$$

QR(Query Relationship) = 질의 Q의 관련성

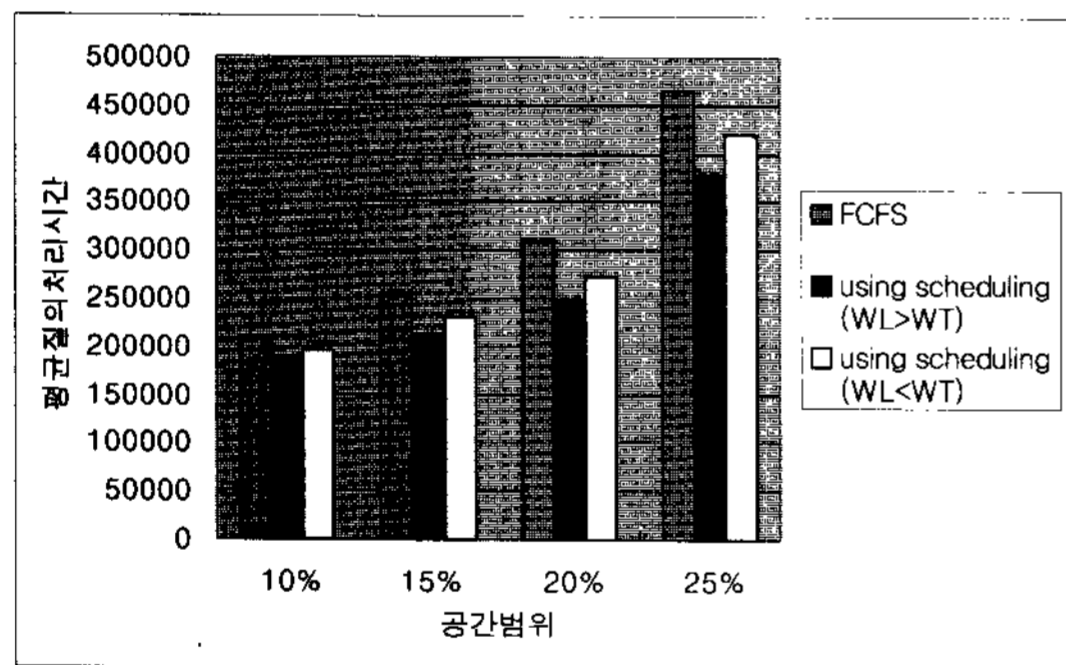
여기서,  $W_L$ 은 위치 관련성에 대한 가중치,  $W_T$ 는 시간 관련성에 대한 가중치 그리고  $W_N$ 은 노드 관련성에 대한 가중치를 나타낸다.

### V. 실험 결과

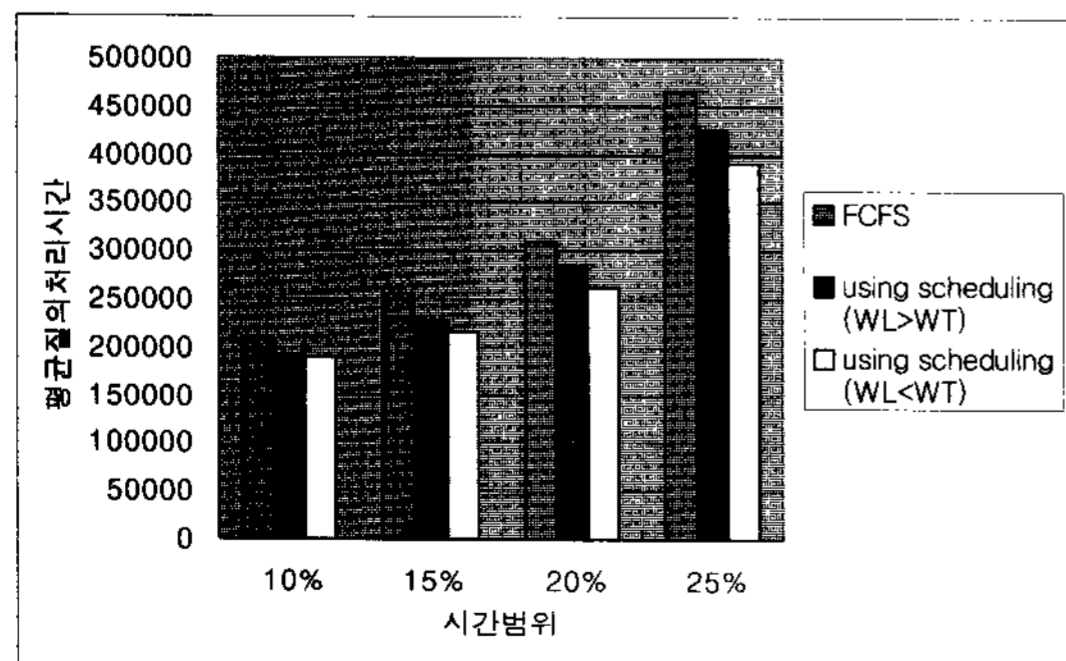
스케줄링 기법을 이용한 제안 방법에서는 전체 영역을 25.6km×25.6km로 구성하여 2000개의 이동 객체와 200개의 질의 수를 시간 및 공간 범위의 크기를 변화시켜 질의를 생성하였다. 이동객체의 위치정보를 발생시키는 데이터로는 휴대폰 사용자 패턴을 적용한 시뮬레이터의 데이터를 사용하였다.

그림 10의 (a)는 공간적인 질의 크기를 변화 시켜 스케줄링 한 실험한 결과로 선입선출 방법으로 처리한 경우 보다 질의 처리 시간이 줄어드는 것을 알 수 있으며, 스케줄링의 경우 위치 관련성의 가중치를 크게 하는 것이 효과적임을 알 수 있다. 그림 (b)는 시간적인 질의 크기를 변화 시켜가며 실험한 결과로 시간 관련성의 가중치를 크게 한 것이 평균 질의 처리 시간이 적게 나타나는 것을 알 수 있다.

실험 결과 스케줄링 시 적용되는 가중치는 공간적으로 변화 폭이 큰 질의들이 많은 경우 위치 관련성 가중치를 크게 하고 시간적으로 변화폭이 큰 경우 시간 관련성 가중치를 크게 하는 것이 우수함을 알 수 있다.



(a) 공간 범위를 변화 시킨 경우  
(a) A variety of spatial range



(b) 시간 범위를 변화 시킨 경우  
(b) A variety of temporal range

그림 10. 가중치 크기에 따른 질의 처리시간  
Fig. 10. Range query processing time.

### VI. 결 론

위치 기반 서비스에서 저장 및 인덱싱의 대상이 되는 이동 객체는 시간이 지남에 따라 위치가 계속적으로 변화하므로 시간에 따른 위치 정보는 기하급수적으로 증가하게 된다. 또한, 이동 객체가 휴대폰 가입자일 경우 이동 경로는 불규칙적이며 특정 노드에 집중되는 현상이 발생할 수 있다. 이러한 경우 부하를 분산시킬 수 있는 방법과 누적되는 데이터양을 줄일 수 있는 방법이 필요하며 이동 객체의 위치 변화에 따라 전체 인덱스 구조의 변화를 최소화 할 수 있는 방안이 제공 되어야 한다. 기존에 제안된 R-트리 계열의 인덱싱 방법은 이동 객체의 움직임에 따라 갱신 연산이 이루어지므로 대량의 위치 정보를 인덱싱 하기에는 부적합 하다고 볼 수 있다.

본 논문에서는 이동 통신 서비스에서 지속적으로 이동하는 고객의 위치정보를 저장 및 검색하기 위해 각 지역의 객체 수에 따라 동적으로 변하는 단계 그리고 특정 지역을 균등 분할하는 비 균등 2단계 고정 그리드 인덱싱 방법을 사용하였고, 질의 간의 관련성을 공간, 시간 노드로 정의 하고 이들 간의 관련성을 이용해 동일 데이터의 중복 탐색을 최소화 할 수 있는 스케줄링 방법을 제안 하였다. 실험 결과 질의 영역이 큰 경우 질의 스케줄링을 이용하면 효율적이라는 것을 알 수 있었으며, 공간 적으로 변화 폭이 큰 경우 공간 관련성의 가중치를 높이고 시간 범위의 변화 폭이 큰 경우 시간 관련성의 가중치를 높이는 것이 효율적임을 알 수 있다. 향후 연구 과제로는 다양한 형태의 시공간 질의에 대해서 스케줄링을 수행할 수 있는 방법에 대한 연구가 이루어져야 할 것이다.

### 참 고 문 헌

- [1] 박원순, 전세길, 나연목, "대용량 이동 객체의 위치 정보 인덱싱", 정보과학회 추계학술발표논문집, Vol. 29, No. 2, pp.49-51, 2002년 10월.
- [2] 전세길, 나연목, "고정그리드 인덱싱에서 VP 필터링을 이용한 범위 질의 처리", 정보처리학회 춘계 학술발표논문집, 제10권 1호, pp.1531-1534, 2003년 5월.
- [3] Y.Nah, M.H.Kim, T.Wang, K.H.Kim, Y.K.Yang, "TMO-structured Cluster-based Real-Time Management of Location Data on Massive Volume of Moving Items," STFES 2003, Hakodate, Japan.

- [4] M.H.Kim, T.Wang, K.H.Kim, Y.Nah, Y.K.Yang, "Distributed Adaptive Architecture for Managing Very Large Volume of Moving Items," IDPT 2003, Beijing, China.
- [5] 나연목, K. H. Kim, 왕태형, 김문희, 이종훈, 양영규, "GALIS : LBS 시스템의 클러스터 기반 신축성소유 아키텍처", 데이터베이스연구, 제18권 4호, pp.33-47, 2002년 12월.
- [6] Segil Jeon, Yunmook Nah, "Range Query Processing for Distributed Real-time Moving Object Databases", APIS 2004, Istanbul, Turkey.
- [7] Saltenis, S., et al, Jensen, E., Leutenegger, S., Lopez, M., "Indexing the Positions of Continuously Moving Objects" Proc. ACM SIGMOD, 2000.
- [8] Theodoridis, Y., Vazigannis, M., and Sellis, T., "Spatio-Temporal Indexing for Large Multimedia Applications", Proc. 3rd IEEE Conf. on Multimedia Computing and Systems(ICMCS), 1996.
- [9] J. Nievergelt, H. Hinterberger, and K.C.Sevcik. The grid file : an adaptable, symmetric multikey file structure. ACM TODS, 9(1) : 38-71, March 1984.
- [10] Xu, X., Han, J., and Lu, W., "RT-tree: An Improved R-tree Index Structure for Spatiotemporal Databases", Proc. 4th Int'l Symp. on Spatial Data Handling(SDH), 1990.

---

 저 자 소 개
 

---



전 세 길(정회원)

1998년 단국대학교 컴퓨터공학과  
공학사.

2000년 단국대학교 컴퓨터공학과  
공학석사.

2004년 단국대학교 전자컴퓨터공  
학과 공학박사.

2005~현재 건국대학교 생명 분자 정보학 센터  
연구원

<주관심분야: 시공간 데이터베이스, 메타 데이터,  
데이터 모델링, 데이터 그리드>



우 찬 일(종신회원)

1993년 단국대학교 전자공학과  
공학사.

1995년 단국대학교 전자공학과  
공학석사.

2003년 단국대학교 전자공학과  
공학박사.

2004년~현재 서일대학 정보통신전공 교수.

<주관심분야: 디지털 워터마킹, 정보보호 시스템,  
스마트카드 보안, 데이터베이스 보안>