

물류 데이터의 연속 질의 처리를 위한 다차원 색인 기법

Multi dimensional index technique for continuous query of logistics data*

추병조*, 홍봉희, 김기홍

Byung-jo Chu*, Bonghee Hong, Gi-hong Kim

부산대학교 컴퓨터공학과

{cbj1004*, bhhong, buglist}@pusan.ac.kr

요 약

EPCglobal은 기업 간의 물류 활동 촉진과 글로벌 유통물류 시스템 구축을 위하여 EPCglobal Architecture Framework을 제시 하였다. EPCglobal Architecture Framework의 한 구성요소인 EPCIS(Electronic Product Code Information Services)는 EPC 기반 물류 관련 정보에 대한 저장 및 검색 서비스를 제공한다. EPCIS는 단발성 질의(poll)와 연속 질의(subscribe) 검색 서비스를 제공한다. EPCIS의 연속 질의는 시스템 자동화 및 재고 관리, 공급망 관리를 위해 다양한 응용에서 활용이 가능하다. 일반적으로 연속 질의의 성능을 향상시키기 위해서는 질의 색인을 사용한다. 그러나 EPCIS는 13차원의 도메인과 모든 데이터 필드가 필수 조건이 아니라는 것으로 인해, 차원의 저주 및 무한 영역 질의 문제를 발생 시킨다.

본 논문에서는 EPCIS의 물류 데이터의 연속 질의 처리를 위한 다차원 색인 기법을 제안한다. 13차원의 도메인을 여러 개의 질의 색인으로 구성하고, 등록된 질의 및 입력되는 데이터에 의해 변경되는 동적 질의 실행 계획을 제한함으로써, 차원의 저주와 무한 영역 질의의 문제를 해결하고, EPCIS에서 다수의 연속 질의 등록 시 효율적으로 처리가 가능하도록 한다.

1. 서론

RFID(Radio Frequency Identification)는 각종 물품에 소형 칩을 부착해 사물의 정보와 주변 환경정보를 무선주파수로 전송·처리하는 비접촉식 자동 인식시스템이다. 이러한 RFID는 빠른 인식 속도, 인식률, 데이터 저장능력으로 인해 여러 응용분야에서 주목 받고 있는 기술이다. RFID의 응용 분야는 재고관리, 물류의 공급망 관리, 공장 자동화 등으로 다양하게 확산되고 있다 [1][2][3]. RFID 기술은 생산자에게는 상품의 판매 상황을 실시간으로 정확한 파악이 가능하게 하며 소비자에게는 상품의 유통 경로나 상품정보에 대한 검색을 가능하게 한다. 또한 현재 상품의 재고를 파악하여 부족한 물품을 자동 재 입고하여 기업의 이익을 확대시킨다.

EPC(Electronic Product Code)는 기존의 바코드와 마찬가지로 상품을 식별하는 코드이다. 바코드는 상품 단위 식별이 가능했

지만, EPC는 상품의 단품 레벨까지 식별이 가능하다. EPCglobal은 RFID 관련 기술을 보급 및 표준화를 담당하는 비영리 국제기구이다. EPCglobal은 개체에 EPC 고유 식별자를 넣은 RFID 태그를 붙여 관리하고 인터넷을 통해 이에 대한 정보획득 절차를 표준화하여 글로벌한 유통물류 시스템 구축을 목표로 하고 있다. 이를 위해 EPCglobal은 EPCglobal Architecture Framework를 제안하였다. EPCglobal Architecture Framework의 한 구성요소인 EPCIS(EPC Information Services)는 EPC 관련 정보를 저장 및 검색 서비스를 제공한다. EPCIS는 원하는 검색 조건을 기술하고, 그 결과를 즉시 받는 단발성 질의(poll)와 검색 결과를 주기적으로 보고 받는 연속 질의(subscribe) 검색 서비스를 제공한다. 일반적으로 연속 질의의 성능을 향상시키기 위해서는 질의 색인을 사용한다. 그러나 EPCIS는 13차원의 도메인과 모든 데이터 필드가 필수 조건이 아니라는

* “이 논문은 2008년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임”(지방연구중심대학육성사업/차세대IT기술연구사업단)

것으로 인해, 차원의 저주 및 무한 영역 질의 문제를 발생 시킨다. 본 논문에서는 EPCIS의 물류 데이터의 연속 질의 처리를 위한 다차원 색인 기법을 제안하였다. 이를 통해 차원의 저주와 무한 영역 질의 문제를 해결하고, EPCIS에서 다수의 연속 질의 등록 시 효율적으로 처리가 가능하도록 하였다.

본 논문의 구성은 다음과 같다. 2장은 관련 연구로써 EPCglobal Architecture Framework와 연속 질의를 위한 질의 색인을 소개 한다. 3장에서는 이 논문의 대상 환경과 문제를 정의 한다. 4장에서는 문제를 해결하기 위한 방법으로써, 여러 개의 색인 구성 방법과 동적 질의 실행 계획을 제안한다. 마지막으로 5장에서는 결론 및 향후 연구 방향에 대해 기술한다.

2. 관련 연구

2.1 EPCglobal Network

EPCglobal은 개체에 EPC 고유 식별자를 넣은 RFID 태그를 붙여 관리하고, 이를 사용하여 공급, 유통망을 강화를 통한 물류 활동 촉진시키기 위해 그림 1과 같은 EPCglobal Architecture Framework를 제안하였다.[4]

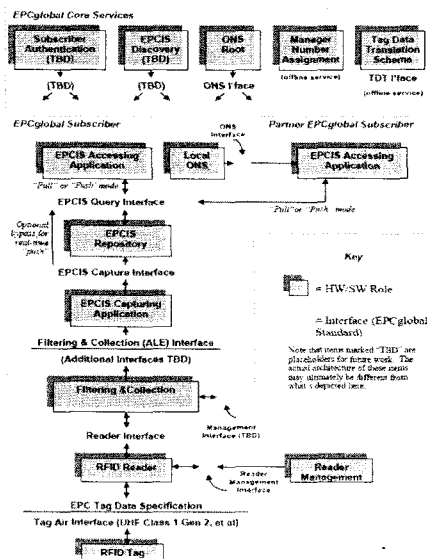


그림 1 EPCglobal Network Architecture

EPCglobal Network Architecture의 주요 구성 요소의 역할은 다음과 같다.

- RFID Tag: 상품을 유일하게 구분할 수 있는 EPC를 보유
- RFID Reader: 하나 이상의 안테나 반경 내에 있는 RFID 태그의 EPC를 인식
- Filtering & Collection (ALE): 하나 이상의 RFID Reader로부터 수신된 데이터를 중복 제거 및 그룹화
- EPCIS Capturing Application: EPC 관련 비즈니스 이벤트 생성 및 EPCIS로 이벤트 전송
- EPCIS: EPC 관련 정보를 저장하고, 검색 서비스를 제공
- ONS: 특정 EPC에 관련 자세한 정보를 제공해주는 EPCIS 주소를 검색

위의 주요구성 요소 중에서 우리는 정보 저장소의 역할을 하는 EPCIS의 검색 서비스에 대해서 연구해보고자 한다.

2.2 EPCIS(EPC Information Services)

EPCIS의 특징은 ALE, Reader와 같은 하위 시스템과 달리 과거 데이터를 저장하고, 비즈니스 정보를 다룬다. EPCIS 데이터는 크게 4개의 기본 이벤트로 구성이 된다. 이를 통해 비즈니스 상에 일어나는 사건을 표현한다. 그림 2는 EPCIS의 기본 이벤트 타입을 보여준다.[5]

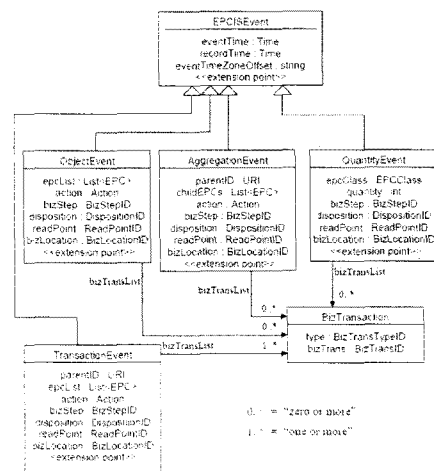


그림 2 EPCIS 기본 이벤트

4가지 이벤트는 다음과 같은 의미를 가진다.

- Object Event: 어떤 시간과 장소에서 어떤 비즈니스 단계에서 관측된 EPC의 리스트를 표현한다. Object Event는 단순 관측 정보를 가진다.
- Aggregation Event: 어떤 시간과 장소에서 어떤 비즈니스 단계에서 부모 단계의 EPC에 물리적으로 결합 또는 해체 되는 EPC의 리스트를 가진다.
- Quantity Event: 어떤 시간과 장소에서 관측된 상품단위의 EPC의 수량을 가진다.
- Transaction Event: 비즈니스와 연결된 정보를 저장하는데 사용하는 이벤트이다.

EPCIS는 위와 같은 4가지 이벤트를 다양한 응용에서 검색이 하기 위한 그림 3과 같은 EPCIS Query Control Interface를 제공한다.

```

<<interface>>
EPCISQueryControlInterface
---
subscribe(queryName : String, params : QueryParams, dest :
URI, controls : SubscriptionControls, subscriptionID :
String)
unsubscribe(subscriptionID : String)
poll(queryName : String, params : QueryParams) :
QueryResults
getQueryNames() : List // of names
getSubscriptionIDs(queryName : String) : List // of Strings
getStandardVersion() : String
getVendorVersion() : String
<<extension point>>
    
```

그림 3 EPCIS Query Control Interface

이중 가장 중요한 함수는 poll과 subscribe이다. poll은 원하는 검색 조건을 Query Params로 기술하고, 그 결과를 즉시 받는 단발성 질의이며, subscribe는 원하는 검색 조건을 기술하고, 검색 결과를 주기적으로 보고 받는 연속 질의이다. 본 논문에서는 EPCIS에서 연속 질의인 subscribe를 효율적으로 처리 하는 기법에 대해서 연구 하고자 한다.

2.3 질의 색인

연속 질의 성능을 개선하기 위해 사용하는 방법 중 대표적인 것이 질의 색인이다. 질의 색인의 요구사항은 실시간 연속 질의 수행을 목적으로 하기 때문에 빠른 검색 속

도를 위해 기본적으로 메인 메모리 기반의 색인 구조를 가지며, 메인 메모리 환경에서 메모리 사용 공간을 최소화하고 삽입되는 질의 개수에 독립적인 빠른 검색 성능을 보장하는 것이다.

대표적인 다차원 질의 색인 구조로써, VCR (Virtual Construct Rectangle) Index[6][7]와 CQI(Cell-based Query) Index[8]가 있다.

3. 대상 환경 및 문제 정의

3.1 대상 환경

EPCIS는 EPC 관련 정보를 저장하는 정보 저장소 역할을 한다. 그리고 저장된 데이터를 단발성 질의(poll)와 연속 질의(subscribe) 형태로 검색 서비스를 제공한다. EPCIS의 연속 질의는 시스템 자동화 및 재고 관리, 공급망 관리를 위해 다양한 응용에서 활용이 가능하다. 그림 4는 EPCIS 연속 질의를 이용한 재고 관리 시스템 응용을 나타낸다.

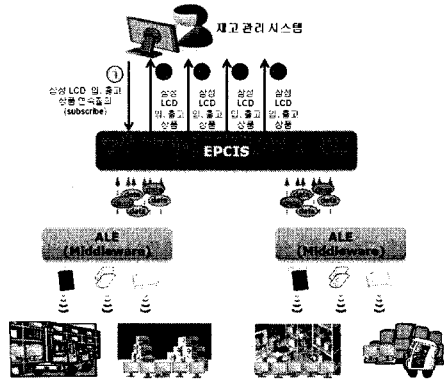


그림 4 EPCIS 연속 질의를 이용한 재고 관리 시스템

재고 관리 시스템 관리자가 작업 진행 상황을 모니터링 하기 위해 매 10분마다 삼성 LCD 입, 출고 상품에 대한 정보가 필요하다고 하면, EPCIS의 연속 질의를 이용하여 이 요구사항을 처리 할 수 있다. 그림 4와 같이 첫째, EPCIS에 삼성 LCD 입, 출고 상품에 대해 10분의 보고 주기를 가지는 연속 질의를 등록 한다. 그 후 관리자는 매 10분마다 연속 질의의 결과인 삼성 LCD 입, 출고 상품 정보 획득이 가능하다. 그림 4의

예제와 같이 EPCIS 연속 질의는 대량의 데이터에서 사용자가 원하는 데이터만을 효율적으로 검색하여, 제공해 주는 역할을 한다. 또한 시스템의 편리함을 제공하고 자동화를 보조해주는 역할을 한다.

3.2 일반적인 질의 색인 처리 기법

일반적으로 연속 질의의 성능을 향상시키기 위해서는 질의 색인을 사용한다. 동작 방법은 그림 5와 같다. 첫째 질의가 등록이 되고, EPCIS 데이터가 입력이 되면 질의 색인을 검색하여(스태빙), 매치되는 연속 질의가 있을 경우 해당 연속 질의의 결과로 추가한다. 그리고 연속 질의 처리기(Continuous Query Processor)에서는 해당 질의 결과 반환 사점이 되면 검색 결과를 사용자에서 전송하게 된다. 질의 결과는 메인 메모리에 유지 된다.

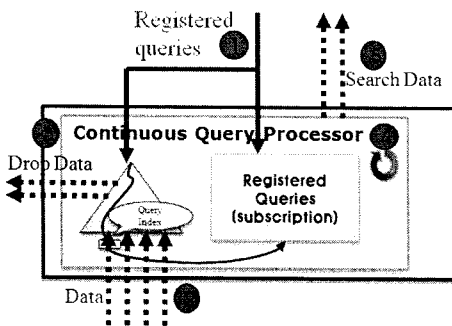


그림 5 질의 색인을 이용한 연속 질의 처리

3.3 문제정의

EPCIS의 경우 그림 6과 같이 13가지가 넘는 고차원의 도메인 조건과 입력 데이터 필드가 필수가 아니라는 2가지 문제점으로 인해 일반적인 질의 색인의 사용이 유용하지 않다.

첫 번째 문제는 EPCIS 표준에 정의된 기본적인 도메인만 해도 13가지로써, 도메인을 색인화 할 경우 차원의 저주로 검색 성능이 현격히 떨어진다. 차원의 저주라고 함은 다차원 색인의 차원이 높아질 경우, 중복 되는 비율이 높아져서, 검색 성능이 현격히 떨어지는 것을 말한다.

	Domain Size	ObjectEvent	Aggregation Event	Quantity Event	Transaction Event
eventType	2bit				
eventTime	64bit				
recordTime	64bit				
epc	96bit				
parentID	96bit				
epcClass	64bit				
quantity	32bit				
action	2bit				
bizStep	8bit이하				
disposition	8bit이하				
readPoint	64bit 이하				
bizLocation	64bit 이하				
bizTransaction	96bit 이상				

Entity Domain Size	Small Domain Size	Mandatory Domain	Optional Domain	Orated Domain
		Always Point Query	Sometimes Infrity Query	Always Infrity Query

그림 6 EPCIS 데이터 도메인 분석

그림 7은 실제 데이터 입력 시 차원이 증가함에 따라 현격히 증가하는 데이터 중복 비율을 보여준다.[9]

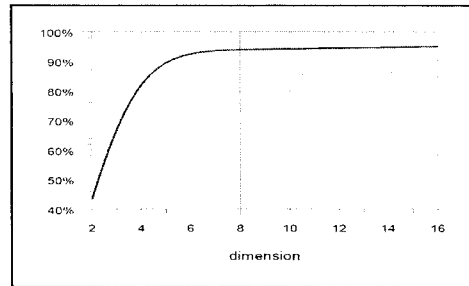


그림 7 차원의 증가에 따른 중복 비율

두 번째 문제는 입력되는 모든 데이터 필드가 필수 아니라는 것이다. 일반적인 질의 색인에 입력되는 데이터는 모든 필드가 정의된 점으로 표현이 된다. 그러나 EPCIS 표준 스펙에 의하면, EPCIS 데이터(이벤트)의 경우 거의 모든 필드가 선택 조건이라는 것이다. 이것은 아주 큰 문제를 야기 시킨다. 데이터의 하나의 필드가 빠진다는 것은 그 차원(필드)의 검색 조건이 무한대가 되는 것을 이야기 한다. 검색 조건이 빠진 필드의 경우 모든 영역이 검색 영역에 해당함으로써, 무한대 질의를 해야 하는 것이다. 2가지의 심각한 문제점 때문에 EPCIS에 기존 질의 색인의 사용은 효율적이지 못하다.

4. 문제해결 방법

4.1 Basic Idea

본 논문에서는 13차원의 차원의 저주 및 무한 영역 질의를 해결하기위해 여러 개의

색인으로 EPCIS 질의 색인을 구성한다. 색인은 각 도메인의 크기, 도메인 필수 유무를 고려하여, 도메인을 그룹화 하여 구성한다. 실제적인 동작과정은 그림8과 같다.

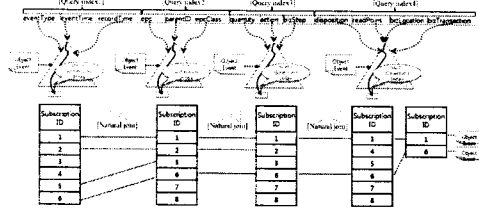


그림 8 여러 개의 색인으로 구성된 질의 색인 동작과정

여러 개의 색인을 구성하고, 각각의 색인을 처리한 결과를 자연 조인하여, 모든 색인 결과에 해당하는 연속 질의가 있을 경우에는 연속 질의의 검색 결과로 데이터(이벤트)를 추가 시키는 것이다.

또한 성능을 향상시키기 위해 다양한 방법으로 차원을 축소시킬 수 있다. 그림 9와 같이 epcClass 도메인과 epc 도메인의 연관성을 활용하여, epc 도메인 상에서 선분으로 표현함으로써, epcClass와 epc 도메인은 하나의 차원으로 구성이 가능하다.

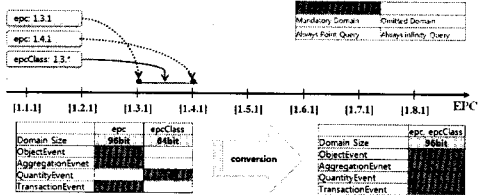


그림 9 epcClass, epc 도메인 병합

BizLocation 도메인은 readPoint를 기반으로 발생하는 비즈니스적인 위치라는 것을 고려하여, 그림 10과 같이 병합을 함으로써, BizLocation과 readPoint 도메인을 하나의 도메인으로 구성이 가능하다.

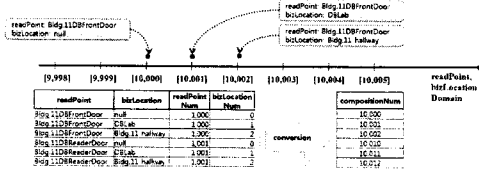


그림 10 BizLocation, readPoint를 도메인 병합

4.2 동적 질의 실행 계획

여러 개의 색인으로 질의 색인을 구성할 경우 그림 11과 같이 색인을 처리하는 순서에 따라 성능의 차이가 발생한다.

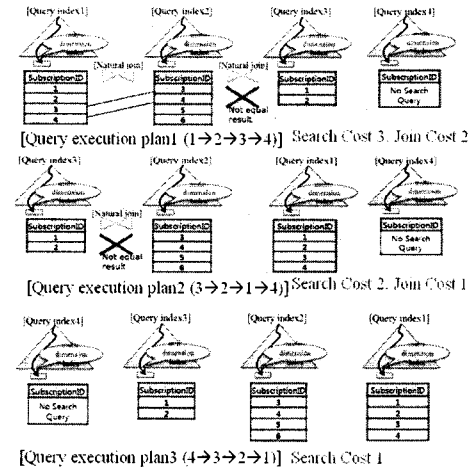


그림 11 질의 실행 계획에 따른 성능차이

이는 각각의 색인을 결과를 자연 조인 후 결과가 있을 경우에 자연 조인의 결과만을 가지고 다음 색인을 처리하기 때문이다. 그림 11의 질의 실행 계획3과 같이 검색 결과를 적게 만드는 색인을 먼저 처리 하는 것이 좋다. 검색 결과를 만드는 조건으로 도메인 크기가 크거나, 해당 도메인이 필수 조건이었을 경우이다(그림 6 참고). 도메인의 크기가 클 경우 EPCIS 데이터가 질의에 포함될 확률이 줄어든다. 해당 도메인 필수 조건이 아닐 경우 데이터 필드가 빠짐으로써, 질의 색인에서 무한 영역 질의를 발생시킬 수 있다. 검색 조건을 적게 만드는 조건 중 가장 중요한 것은 각 도메인에 등록된 질의이다. 사용자가 특정 도메인에 질의 조건을 하나도 등록하지 않았다면 그 도메인의 이벤트 질의 결과는 등록된 모든 질의가 될 것이다. 현재 시스템에 등록된 질의들의 도메인 조건이 중요한 것이다. 본 논문에서는 도메인 크기, 도메인 필수 유무, 등록된 유저의 질의를 고려하여, 입력된 각각의 이벤트에 대해서 최적화된 동적 질의 실행 계획을 제안한다. 특히 유저의 질의 조건은 가변적이고, 실시간으로 변경이 되므로, 연속 질의 등록 및 삭제 시 동적으로 정보를 갱신하여, 항상 최적화된 질의 실행

계획이 가능하도록 한다.

4.3 연속 질의 처리기 구조

그림 12은 동적 질의 실행 계획을 적용하기 위한 연속 질의 처리기 구조도를 나타내고 있다.

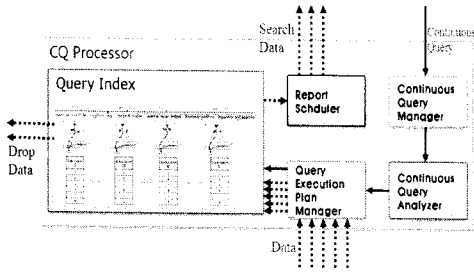


그림 12 연속 질의 처리기 구조도

제안한 연속 질의 처리기는 5개의 모듈로 구성되어 있다. 각 모듈의 역할 및 특징 다음과 같다.

- Query Index: 여러 개의 색인으로 구성되며, 입력된 데이터를 각각의 색인을 검색하여 조건에 맞는 질의가 있을 경우에 연속 질의 검색 결과로 추가 시키고, 없을 경우에는 삭제된다.
- Report Scheduler: 등록된 모든 연속 질의 보고 주기를 관리한다.
- Continuous Query Manager: 등록된 모든 연속 질의를 관리하고, 연속 질의 결과로 검색된 모든 데이터를 결과 전송 시점까지 관리한다.
- Query Execution Plan Manager: 분석된 연속 질의 정보를 기반으로 입력된 데이터에 대해 최적화된 질의 수행 계획을 생성한다.
- Continuous Query Analyzer: 등록된 연속 질의의 검색 조건의 도메인 크기, 질의 크기 등을 분석하여, Query Execution Plan Manager에게 연속 질의 분석 정보를 제공한다. 연속 질의 등록, 삭제 시점에 질의를 분석한다.

5. 결론 및 향후 연구

EPCIS는 EPCglobal Network Architecture에서 EPC 관련 정보 저장 및 검색을 제공하는 시스템이다. EPCIS에서 제공하는 연속

질의(subscribe)는 시스템 자동화 및 재고 모니터링, 공급망 관리와 같은 다양한 응용에서 사용 될 수 있다. 연속 질의의 성능을 향상시키기 위해 사용되는 일반적인 질의 색인은 EPCIS의 13차원의 도메인과 모든 데이터 필드가 필수 조건이 아니라는 것으로 인해, 차원의 저주 및 무한 영역 질의 문제를 발생 하였다.

본 논문에서는 13차원의 도메인을 여러 개의 질의 색인으로 구성하고, 등록된 질의 및 입력되는 데이터에 의해 변경되는 동적 질의 실행 계획을 제한함으로써, 차원의 저주와 무한 영역 질의의 문제를 해결하고, EPCIS에서 다수의 연속 질의 등록 시 효율적으로 처리가 가능하도록 하였다.

향후 연구로는 본 논문에서 설계된 연속 질의 처리기에서 사용되는 여러 개의 질의 색인을 RFID 환경의 특성을 고려하여 구성하여 보다 향상된 연속 질의 처리 성능을 제공하는 것이 필요하다.

6. 참고문헌

- [1] K. Romer, T. Schoch, F. Mattern and T. Dubendorfer, "Smart Identification Frameworks for Ubiquitous Computing Applications" Proc. of Pervasive Computing and Communications, pp. 256-262, 2003
- [2] S. E. Sarma, S. A. Weis, and D. W. Engels. "RFID Systems and Security and Privacy Implications". Springer-Verlag, pp. 454-469, 2002
- [3] K. Romer, T. Schoch. "Infrastructure Concepts for Tag-Based Ubiquitous Computing Applications". Proc. of Ubicomp, pp. 253-262, 2002
- [4] EPCGlobal, "The EPCglobal Architecture Framework 1.2", 2007
- [5] EPCGlobal, "EPC Information Services (EPCIS) Version 1.0 Specification", 2007
- [6] Kun-Lung Wu, Shyh-Kwei Chen, Philip S. Yu, "VCR indexing for fast event matching for highly-overlapping range predicates." . SAC 2004, pp740-747, 2004.
- [7] Kun-Lung Wu, Shyh-Kwei Chen, Philip S. Yu. "Processing Continual Range

Queries over Moving Objects Using VCR-
Based Query Indexes.” , MobiQuitous
2004: 226-235

- [8] D.V.Kalashnikov, S.Prabhakar, W.G.Aref,
and S.E.Hambrusch. “Efficient evaluation
of continuous range queries on moving
objects.” , In Proc. of 13th DESA,
2002.
- [9] Stefan Berchtold, Daniel A. Keim,
Hans-Peter Kriegel. “The X-tree: An Index
Structure for High-Dimensional Data
Stefan Berchtold” Proc. 22nd VLDB
Conf., 28-39 1996