

실시간 센서 데이터 배포를 위한 효율적 매칭

An efficient matching mechanism for real-time sensor data dissemination

석 보 현* 이 필 우** 허 의 남***
Bo-Hyun Seek Pill-Woo Lee Eui-Nam Huh

요 약

실시간적인 데이터의 수집과 더불어 수집한 데이터의 실시간적인 전송을 기반으로 사용자가 센서데이터를 보다 폭넓게 활용할 수 있는 환경을 제공하기 위해 시스템에서 자동적으로 정보를 배포해주는 정보배포 시스템의 필요성이 증대되고 있다. 이러한 요구에 맞추어 본 논문에서는 그리드 환경을 기반으로 센서네트워크에서 유입되는 방대한 양의 데이터를 처리 및 공유하기 위한 정보배포 시스템과 보다 효율적으로 데이터와 사용자의 요구를 매칭하는 방법을 제공하는 CGIM 알고리즘을 제안하였다.

Abstract

In the ubiquitous environment, sensor network technologies have advanced for collecting information of the environment. With the rapid growth of sensor network technology, it is necessary and important to share the collected sensor data with a large base of diverse users. In order to provide dissemination of sensor data, we design an information dissemination system using an independent disseminator between provider and consumer.

This paper describes how we designed the information dissemination system using one of the possible dissemination patterns for sensor networks, and an efficient matching algorithm called CGIM (Classed Grouping Index Matching) which employs a dynamic re-grouping scheme.

☞ keyword : Information Dissemination, Matching mechanism, Grid, Sensor data

1. 서 론

비쿼터스 사회에서 정보의 수집은 사용자들에게 보다 쾌적하고 편리한 환경을 제공하기 위한 가장 기본적인 요구라고 할 수 있다. 이러한 요구에 따라 많은 정보를 실시간적으로 수집하기 위해 센서 네트워크에 대한 연구가 활발하게 진행 중이며 이를 이용하여 자연 재해나 각종 위험들을 방지하기 위해 자연 환경 정보를 수집하거나, 노인이나 환자의 지속적인 관리와 위급상황이 발

생하는 경우 인명구조, 실내에서의 자동 조명이나 환기 조절 등 보다 쾌적한 환경을 위해서 사용될 수 있다.

위와 같은 목적을 실현하기 위해서는 실 시간적인 데이터의 수집과 더불어 수집한 데이터의 실 시간적인 전송이 기반이 되어야 하며 사용자가 이러한 센서데이터를 보다 폭넓게 활용하기 위해, 사용자가 정보를 찾아가는 것이 아니라 시스템이 사용자의 요구정보를 이용하여 원하는 정보를 자동적으로 배포해주는 정보배포 시스템의 필요성이 대두되고 있다. 따라서 본 논문에서는 센서 데이터 실시간 처리와 공유를 위한 시스템으로 방대한 양의 데이터를 처리, 관리할 수 있는 그리드 환경에서 사용자들이 필요로 하는 정보를 자동적으로 찾아 배포해주는 정보배포 시스템을 제안한다.

* 준 회 원 : 경희대학교 컴퓨터공학과 석사과정
bohyun@icns.khu.ac.kr

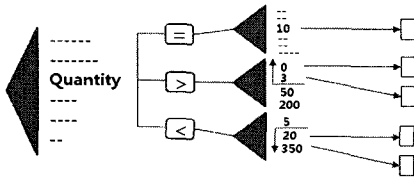
** 정 회 원 : 한국과학기술정보연구원 책임연구원
pwlee@kisti.re.kr

*** 종신회원 : 경희대학교 컴퓨터공학과 교수
johnhuh@khu.ac.kr(교신저자)

[2007/09/07 투고 - 2007/09/11 심사 - 2007/09/17 심사완료]

subscription id를 상호 연계하여 해당되는 subscription들을 찾을 수 있도록 하였다.

[4-5]에서 제안하는 forwarding algorithm은 attribute별 인덱싱 및 그룹핑을 통해 다양한 attribute를 가지는 predicate를 찾아낸다. predicate들은 여러 attribute를 가지는데, 각 predicate별로 각각 다른 종류와 수의 attribute를 가질 수 있다. 같은 attribute와 value값을 가지는 경우를 제외한 모든 attribute들을 indexing하여, 데이터와 일치하는 attribute에 해당하는 predicate를 찾는다. 그림(2)는 Indexing한 attribute별로 predicate의 value들을 크기 순으로 정렬하고 정렬된 value값을 이용하여 데이터와 일치하는 값부터 value값의 끝까지 순차적으로 매칭한다. Value값이 일치하면 해당 attribute를 가지는 predicate를 찾는다.



(그림 2) Forwarding Algorithm

앞에서 설명한 연구들에서 제시하는 매칭 알고리즘들은 단일 operation에 의한 predicate들의 매칭을 제공하는데, 이러한 알고리즘은 두 개의 범위를 조합한 predicate 매칭에는 매우 취약하다는 단점이 있다. 복합 predicate를 이용한 subscription명세가 필요한 환경에서 이러한 알고리즘들을 사용할 경우, 각각의 operator에 대한 단일 predicate 매칭 결과 값을 조합하는 과정이 필요한데, 이는 불필요한 자원과 시간의 낭비를 초래한다. 또한, 부분적인 교집합을 가지는 predicate들과 포함 관계를 가지는 predicate들 또한 고려해야 한다는 어려움이 있다. 따라서 이러한 문제점들을 고려하여 정보배포 시스템에서의 효율적인 매칭을 수행하고자 본 논문에서는 u-Health 환경에 맞추어 범위 조건을 매칭시키는

CGIM 알고리즘을 제안하였다. CGIM알고리즘을 제안하는데 앞서 정보배포 시스템에 대해 설명하겠다.

3. 정보배포 시스템

정보배포 시스템은 외부로부터 수집한 각종 정보를 필요로 하는 사용자를 선별하여 배포해주는 기능을 제공한다. 본 논문에서는 정보배포 시스템을 OGF의 INFO-D 표준의 Single Disseminator Pattern을 기반으로 설계하였다.

3.1. OGSA-DAI (Open Grid Service Architecture Data access and Integration)

정보배포 시스템의 구조와 동작과정을 설명하기에 앞서 보다 효율적인 데이터 관리와 배포를 위해 사용한 OGSA-DAI[8]의 장점에 대해 간략하게 설명하고자 한다. OGSA-DAI는 데이터 그리드 환경에서 다양한 종류의 데이터베이스를 보다 효과적으로 사용하기 위해 통합된 인터페이스를 제공하는 미들웨어로서 데이터의 저장, 관리뿐만 아니라 변환, 전송 기능까지 제공해주는 장점이 있다.

OGSA-DAI의 장점으로 부각되는 기능들을 살펴보면 다음과 같다.

첫째, XML DB, Relational DB, 파일 시스템 등 다양한 데이터 스토리지를 위한 통합된 인터페이스를 제공한다. 어떤 형태의 데이터든지, 통합된 인터페이스를 통해 각각의 DB에 저장, 관리 될 수 있다.

둘째, 데이터의 저장 및 관리뿐만 아니라 형식을 변환시키거나 전송하는 기능을 제공한다. 단순히 데이터를 저장, 관리하는 데이터베이스의 개념뿐만 아니라 일반 데이터를 XML형식으로 변환시키거나 Ftp, E-mail 또는 웹서비스를 이용한 전송 기능까지 제공해주므로, 정보배포 시스템에서 각 기능을 일일이 구현하지 않아도 된다는 장점이

있다.

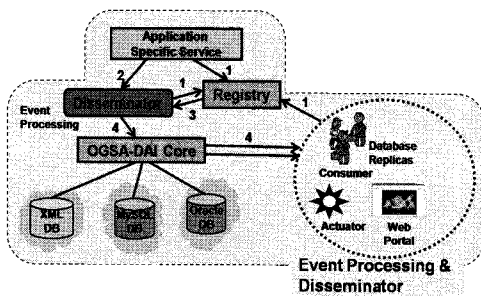
셋째, OGSA-DAI에서 기본적으로 제공되는 activity뿐만 아니라 API를 사용하여 원하는 activity를 생성하여 실행할 수 있다. 개발자가 직접 개발하는 어플리케이션에 맞는 activity를 생성하고 실행 함으로서 OGSA-DAI를 어플리케이션에 보다 적절하게 사용할 수 있다.

넷째, Activity생성뿐만 아니라 실행 순서를 설정할 수 있는데 이는perform문서의 명세를 통해 가능하다. 따라서 유동적인 OGSA-DAI의 사용이 가능하다.

본 논문에서는 위에서 언급한 OGSA-DAI의 장점들을 정보배포 시스템에 적용하여 설계하였고, 이러한 정보배포 시스템의 동작 과정은 그림(3)에서 나타내고 있다.

3.2. 정보배포 시스템 구조 및 동작과정

정보배포 시스템의 동작과정은 다음과 같다.



(그림 3) 정보배포 시스템의 구조

1. 각 컴포넌트들은 (Application Specific Service, Disseminator, Consumer) Registry에 자신의 정보를 등록한다.

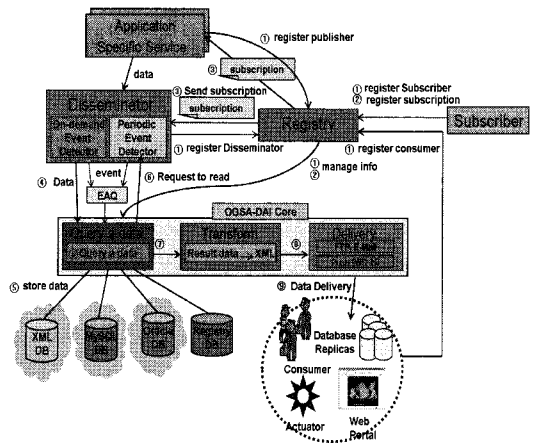
2. Application Specific Service는 "Stream Processing & Knowledge Processing"에서 필터링 및 가공 처리된 센서 데이터를 Disseminator에게 전달한다.

3. Disseminator는 Registry로부터 Consumer의 요구조건이 명세된 subscription을 전달받고 이를 통

해 어떤 사용자에게 어떤 주기로 정보를 전달할지를 스케줄링한다.

4. Disseminator는 원하는 activity수행 정보를 perform문서로 작성하여 OGSA-DAI를 통해 각각의 DB에 데이터를 저장하고 스케줄링한 정보배포 순서에 따라 전송하는 등의 activity실행을 요청한다. OGSA-DAI는 Disseminator에서 요청한 activity들을 수행함으로써 쿼리를 실행하여 데이터를 저장, 검색하거나 consumer들에게 전송하는 역할을 수행한다.

그림(4)는 Information Disseminator부분을 보다 자세하게 나타내고 있다.



(그림 4) OGSA-DAI Core Architecture

① Application Specific Service는 생산자로서, 병원이나 의사와 같은 유저들은 구독자로서, 또한 Disseminator와 Subscriber를 Registry에 자신의 정보를 등록한다.

② 구독자는 자신의 정보와 어떤 정보를 어떤 주기로 배포받기를 원하는지를 명시한 정보를 Registry에 등록된 Subscriber에게 전달하고 Subscriber는 이를 이용하여 구독정보를 작성하여 Registry에 등록한다. Registry는 ①과②에서 등록받은 정보들을 OGSA-DAI의 registryDB에 저장한다.

③ 레지스트리는 등록된 구독정보들을 주기적으로 배포자와 생산자에게 전달한다.

④ 하나 또는 여러 Application Specific Service 들을 통해 가공된 데이터들이 들어오면, 배포자는 먼저 On-demand Event Detector에서 구독정보에 따라 이벤트인지 일반 데이터인지를 분류한다. 분류한 데이터는 구독정보에 따라 단일 이벤트만 충족되어야 할 경우와 여러 이벤트가 동시에 충족되어야 하는 경우로 나뉘는데, 이벤트 데이터는 두 경우에 따라 각기 다르게 동작한다. 센서 데이터가 일단 이벤트 데이터로 구분이 되면 구독정보에 단일 이벤트인지 복합 이벤트인지를 구분한다. 단일 이벤트로 분류되면 ⑤를 거쳐 바로 ⑦과정으로 넘어가고, 여러 이벤트가 동시에 충족시켜야 하는 경우에는 이벤트 데이터를 구독정보에 표기되어 있는 주기에 따라 EAQ(Event Aggregation Queue)에 저장한다. "만일 정해진 주기 내에 모든 이벤트가 발생하지 않거나, 이벤트로 판명되지 않는 경우 각 subscription에 해당되는 EAQ는 초기화된다."

⑤ 배포자에서 분류된 데이터들은 OGSA-DAI를 통해 DB에 저장된다. 이벤트 데이터일 경우에는 저장된 후 바로 ⑦과정으로 넘어간다.

⑥ 배포자의 Periodic Event Detector에서는 단일 이벤트일 경우, 구독정보에 따라 구독자가 원하는 주기별로 데이터 전송을 스케줄링하여 해당 주기에 맞춰 OGSA-DAI에 센서 데이터를 검색하는 request를 전달한다. 여러 이벤트가 동시에 충족시켜져야 하는 경우에는, 구독 정보에 따라 충족시켜야 하는 모든 이벤트 데이터가 EAQ에 저장된 후, ⑤, ⑦ 과정을 통해 데이터를 저장, 배포한다.

⑦ 검색된 데이터는 OGSA-DAI의 Transform Activity에 전달되며 XML형식의 데이터가 아닐 경우 XML 형식으로 변환된다.

⑧ XML형식으로 변환된 이벤트 데이터는 구독자의 정보에 따라 FTP, E-mail 등의 수단으로 배포된다.

4. CGIM(Classed Group Index Matching)

알고리즘

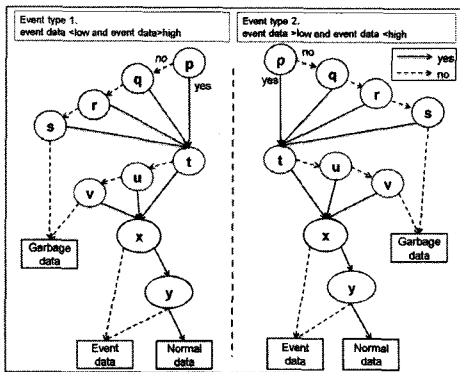
우리가 제안하는 정보배포 시스템에서 사용자는 원하는 종류/범위의 센서데이터를 실시간적으로 받을 수 있어야 하는데, 이를 위해 사용자의 요구를 명세하는 것이 subscription이다. Subscription의 "조건(condition)"과 센서 데이터를 비교, 매칭하기 위해 몇 가지 고려해야 할 사항들이 있다. 첫째, 특정 이벤트에 대한 실시간적인 전송뿐만 아니라 사용자가 지정한 주기별 전송 또한 제공되어야 한다. 사용자가 이벤트 데이터로서 실시간적으로 전송 받고자 하는 내용 이외에 따로 저장되는 데이터는 subscription에 명세한 주기 별로 사용자에게 전송되어야 한다. 둘째, u-Health 시스템에서 제공하는 data의 종류가 센서의 종류에 따라 여러 가지가 될 수 있다. 셋째, 센서 데이터를 어플리케이션이나 각종 서비스에서 사용할 경우, 범위를 이용하여 요구사항을 나타내는 것이 더욱 더 효율적이다. 따라서, 사용자의 요구를 충분히 반영하기 위해서는 범위를 적용한 요구사항 명세가 필요하다. 넷째, 센서 데이터가 들어왔을 때, disseminator의 모든 subscription과 일일이 매칭하게 되면 이벤트 데이터를 찾는데 많은 시간이 소모될 것으로 예측된다. 다수의 사용자에게 본 논문에서 제안하는 배포 시스템 서비스를 제공하기 위해서는 방대한 양의 subscription과의 매칭이 반드시 필요하게 되는데, 이를 위해 보다 효율적인 매칭 알고리즘이 필요하다.

4.1 순차적 매칭 알고리즘

본 논문에서 제안하는 매칭 알고리즘인 CGIM 매칭 알고리즘을 설명하기에 앞서 순차적인 방식으로 센서 데이터의 이벤트 여부를 찾아내는 순차적 매칭 알고리즘을 설명한다. 정보배포 시스템에서 사용되는 센서 데이터는 실제 감지한 데이터 값뿐만 아니라 센서를 부착한 환자의 정보에

따라 병과(part)와 유형(event_type) 정보를 포함한다. Part는 환자가 가지고 있는 병력 정보를 나타내는데, 이는 순환계, 소화계, 호흡계, 신장으로 나뉘고 event_type은 센서의 종류에 따라 혈압, 체온, 맥박 등으로 나뉜다. 그림(5)은 이러한 순차적인 방법으로 매칭하는 매칭 알고리즘을 BDD(binary Decision Diagram)를 통해 모델링 한 것이다.

매칭 순서는 그림(5)에서 보이는 바와 같이 동작한다. p(순환계), q(소화계), r(호흡계), s(신장)는 part를 나타내는 것으로 센서 데이터의 part가 p,q,r,s 중 하나와 일치하면 다음 단계로 data_type은 t(data_type a), u(data_type b), v(data_type c)와 매칭한다. 센서 데이터가 attribute인 part와 event_type에 모두 해당이 되면, 사용자의 요구 범위 중 최소값을 나타내는 x와 최대값을 나타내는 y에 비교하여 조건에 맞는지 여부를 확인한다. 최종적으로 센서 데이터가 x보다 크고 y보다 작을 경우 매칭한다는 결과값을 갖게 된다. 그러나, 위와 같은 방법으로 순차적인 모델을 통해 매칭하는 경우, 센서 데이터가 들어올 때마다 데이터를 모든 subscription과 일일이 매칭해야 하므로 자원과 시간을 낭비하는 문제점을 초래하게 된다. 이러한 문제점을 해결하기 위해 본 논문에서는 subscription의 조건에 해당하는 predicates의 그룹핑을 이용하여 보다 효율적인 매칭을 제공하는 매칭 알고리즘을 제안한다.

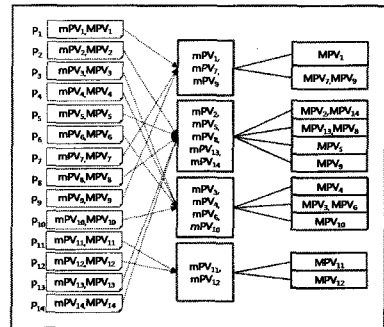


(그림 5) 순차적 매칭 알고리즘의 BDD모델

4.2. 효과적인 매칭을 위한 predicate 그룹핑

Subscription은 사용자의 정보 외에 사용자의 조건을 나타내는 predicate로 이루어진다. Predicate는 $P=(event_type, N, op, mPV, MPV)$ 와 같은 형태로 구성되는데, 이때, event_type은 데이터의 종류를 명세하는 부분이고, N은 범위를 선택하는 방법에 대해 나타낸다. op는 <<또는 >>와 같은 범위를 나타내는 부등호 operation을 나타낸다. 또한, mPV는 범위의 최소 값을, MPV는 범위의 최대값을 나타낸다. 이러한 범위는 $mPV < data < MPV$ 로 나타낼 수 있다.

아래 그림(6)은 매칭을 위한 그룹핑을 여러 predicate들을 그룹핑한 사례로서 나타낸 것이다. mS는 각각의 predicates들이 포함하는 mPV들의 모임 ($mS = \{mPV_1, mPV_2, \dots, mPV_i\}$) 이고, MS는 MPV의 모임 ($MS = \{MPV_1, MPV_2, \dots, MPV_i\}$) 이다. 본 논문에서는 mPV의 값에 따라 그룹 mS로 그룹핑하고 이러한 mS 내에서 다시 MS로 하위 그룹을 그룹핑하였다.

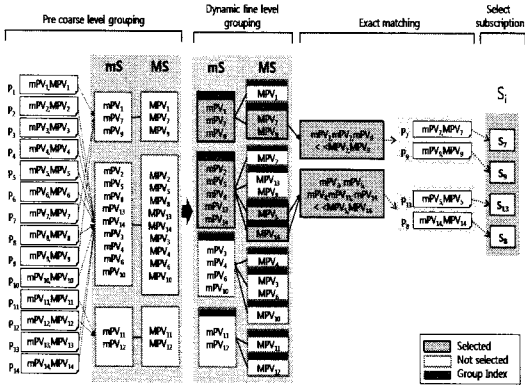


(그림 6) predicate 그룹핑 사례

4.3. CGIM 매칭 알고리즘

이번 절에서는 본 논문에서 제안하는 매칭 알고리즘의 전체적인 매칭 과정을 그림(7)을 통해 설명하고자 한다.

4.4. 그룹 분할



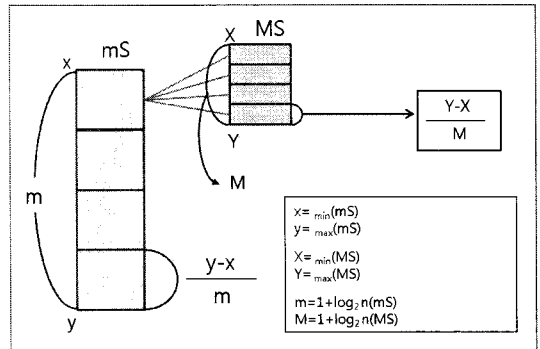
(그림 7) CGIM 알고리즘

Pre coarse level grouping은 이미 5.2절에서 설명한 방법으로 predicates들이 mS그룹과 하위 MS 그룹으로 그룹핑 되는 과정을 말한다. Pre coarse level grouping은 subscription이 등록되면 센서 데이터가 들어오기 전에 미리 시스템에서 그림(6)과 같이 트리 형태로 그룹핑한다. Dynamic fine level grouping에서는 실질적인 센서 데이터 매칭이 시작되는데 일차적으로 해당되는 mS그룹을 매칭하고 이렇게 선별된 mS 그룹 내에서 다시 해당되는 MS그룹을 선별한다. Pre coarse level grouping과 Dynamic fine level grouping을 통해 생성된 2계층의 subscription 그룹은 각 그룹마다 대표값이 정해지고, 이러한 대표값들이 인덱싱되어 센서데이터가 들어올 경우, 인덱싱 된 대표값과 적은 횟수의 비교만으로도 조건에 해당되지 않는 그룹을 제외할 수 있다. 최종적으로 Exactly matching에서는 선택된 그룹 내의 predicate들 중에서 정확하게 매칭하는 predicate들을 찾는 과정을 나타낸다. Exactly matching 과정을 거쳐 정확하게 매칭되는 predicate들의 id를 이용하여 해당 subscription을 발체한다. 발체한 subscription은 데이터를 전송받을 사용자와 전송할 방법을 선택하는데 이용된다.

그룹을 생성하는 데 있어서 고려해야 할 점은 그룹을 얼마나 분할하는 가에 있다. 본 논문에서는 두 가지 방법을 이용하여 그룹 수를 나누고 1차적으로 실험을 하였다. 첫 번째는 통계적인 방법으로 그림(8)에 나타나 있듯이, subscription의 개수(n)을 이용하여 식 ①과 같이 그룹수(m)을 결정한다. 도출된 m을 이용하여 식 ②를 계산하면 간격(p)를 결정할 수 있다. 이러한 방법을 이용하여 그룹핑하고 사례를 실험하였다.

$$m = 1 + \log_2 n \text{ ----- ①}$$

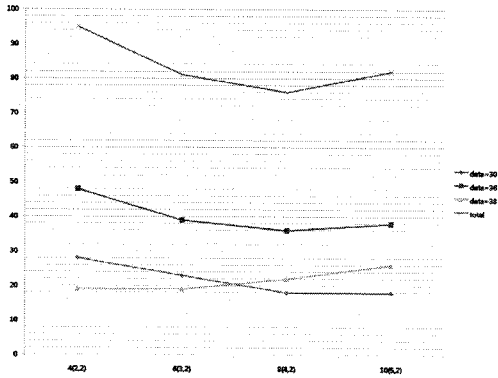
$$p = (y - x) / m \text{ ----- ②}$$



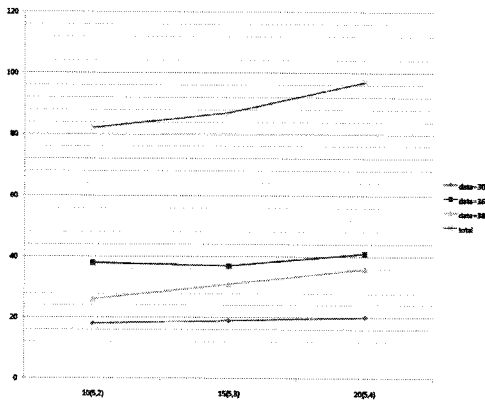
(그림 8) 통계적 계급 수 지정에 따른 그룹핑

두 번째 방법으로, 계급 수를 다양하게 설정하여 실험하였다. 상위 그룹과 하위 그룹의 그룹 수를 나타내는 m과 M을 다양하게 설정하여 식 ②를 이용하여 간격 값을 도출하고 간격 값에 따라 predicate들을 그룹핑 하였다.

50개의 subscription을 그룹핑한 후, 매칭 알고리즘을 사용하여 5가지 종류의 데이터를 이벤트 매칭한 결과는 그림(9), (10)의 그래프를 통해 나타난다.



(그림 9) mS 그룹수 변경에 따른 그룹핑



(그림 10) MS 그룹수 변경에 따른 그룹핑

위의 실험들을 통해 아래와 같은 결과를 도출해낼 수 있다.

- 유입되는 data의 값이 조건 그룹 후반부에 해당되면 매칭 시간은 그룹 수에 영향을 많이 받는다. 그룹 수가 증가할수록 성능이 떨어진다.
- 유입되는 data의 값이 조건 트리의 전반부에 해당되면 매칭 시간은 전체적인 그룹 수 증가에 따른 그래프와 유사한 형태의 성능을 보인다.
- sub group의 수가 많아질수록 성능이 떨어진다.
- Pre coarse level grouping 에서의 group 수가 증가되면 특정 값 이상의 그룹 수부터는 성능이 떨어진다.

5. 실험

본 논문에서 제안하는 CGIM 알고리즘의 성능을 분석하기 위해 CGIM 알고리즘과 기존의 연구에서 제안한 단일 predicate를 사용하는 Forwarding 알고리즘[4-5]과 최악의 경우로 모든 subscription을 매칭하는 sequential matching 을 비교하였다. 각각의 알고리즘을 설명하면 다음과 같다.

1. Worst case(Sequential matching): 데이터가 들어오면 모든 subscription을 처음부터 순차적으로 매칭한다.

2. Forwarding 알고리즘: 단일 predicate를 사용하는 Forwarding 알고리즘에서는 먼저 operator별 그룹을 생성하고 각각의 그룹 별로 모든 value를 정렬한다. 정렬한 모든 value를 인덱싱하고 데이터가 들어오면 operator에 따라 데이터와 일치하는 값에서 시작하여 해당하는 모든 value를 순차적으로 찾는다. 실험과정은 두 가지의 단일 operator의 매칭 결과값을 도출하고 그 결과들을 조합하여 원하는 매칭 결과를 산출한다.

3. CGIM 알고리즘: Predicate의 operator와 value 값을 그룹핑 하고 predicate그룹을 인덱싱하여 매칭을 제공한다.

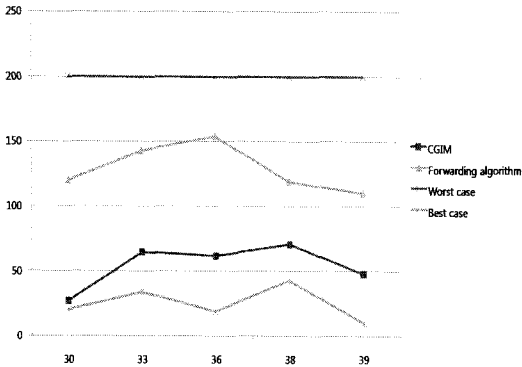
5.1. 매칭 성능 비교

시스템이 100개의 subscription정보를 가지고 있다고 했을 때, 5개의 data를 입력 받아 해당되는 subscription을 찾는 과정에서 수행되는 매칭 횟수를 실험하였다. 입력한 데이터는 30, 33, 36, 38, 39 이고 각각 다른 개수의 그룹을 통해 실험하였다.

본 실험에서는 Predicate Case1에 해당하는 (mPV<data<MPV) 형태의 조건 매칭 실험하였다.

그림(11)은 표1 중에서도 각 알고리즘 별로 데이터 5개를 모두 매칭 한 후의 매칭횟수를 그래

프로 나타낸 것이다. forwarding algorithm을 사용하여 매칭하였을 때, 5개의 데이터를 모두 매칭한 횟수는 646번, Worst case는 1000번, CGIM 알고리즘을 사용하면 그룹 수에 따라 차이가 있지만 평균적으로 262번의 매칭 이 발생했다.



(그림 11) Predicate case1 에 해당하는 predicate를 이용한 매칭 실험결과

Worst case의 sequential matching을 수행하는 경우에는 874번, Forwarding 알고리즘을 사용하여 매칭하는 경우에는 520번의 매칭과정이 낭비된다.

낭비되는 매칭 수 :

Best case 매칭 시간 - 해당 알고리즘 매칭 시간

본 논문에서 제안하는 CGIM 알고리즘을 사용하면 낭비되는 매칭과정은 147번으로 Worst case에 비해 83%의 매칭시간이, Forwarding 알고리즘에 비해 71%의 매칭 시간을 줄일 수 있다는 결과를 가질 수 있다.

5.2. 효과적인 그룹핑 방법

위의 간단한 실험 결과들을 살펴보면, 그룹을 어떻게 생성하느냐에 따라 매칭의 효율성이 달라지며, 효율성을 높이기 위해서는 보다 효과적인 그룹 지정 방식이 필요하다는 것을 알 수 있다.

이에 본 문서에서는 다양한 침도로 분류되는 subscription 패턴 예제들을 사용하여 침도 값에 따라 효과적인 그룹핑 방식을 제안한다.

첫째, 일정간격 그룹핑은 전체 구간 $minimum < maximum$ 범위를 앞에서 설명한 통계적인 방법으로 그룹 수를 설정하고 해당 그룹수(m)에 따라 동일한 간격으로 자른다.

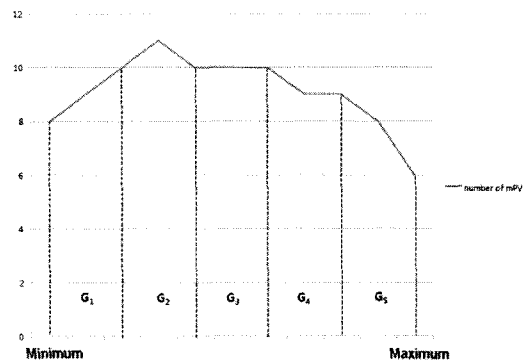
$$m=1+\log_2 n(mS)$$

둘째, 동적 재분할 그룹핑은 먼저, 전체 범위를 세 그룹으로 나누는데, 첫번째 그룹은 $G_1 = Minimum < a$ 최소값부터 a 사이의 범위를 가지고 두번째 그룹은 $a < \beta$ 범위를, 마지막 그룹은 $G_5 = \beta < Maximum$ β 값부터 최대값 사이의 범위를 가진다. 다음 단계로 $a < \beta$ 범위는 통계적인 방법을 통해 그룹 수를 설정하고 그에 해당하는 수의 subgroup을 나눈다.

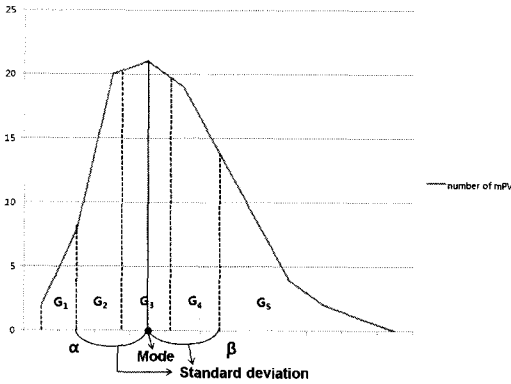
$$a = Mode - Standard\ deviation$$

$$\beta = Mode + Standard\ deviation$$

그림(12), 그림(13)은 각각 일정간격 그룹핑과 동적 재분할 그룹핑 방법을 나타낸 그림이다.



(그림 12) 일정간격 그룹핑 방법 그림



(그림 13) 동적 재분할 그룹핑 방법

일정간격 그룹핑은 침도가 낮은, 즉, subscription의 분포가 고른 경우에 효과적으로 사용할 수 있는 방법이고, 동적 재분할 그룹핑은 침도가 높은, 즉, 특정 간격에 분포가 밀집된 분포에 효과적으로 사용할 수 있는 그룹핑 방법이다.

5.2.1. 그룹핑 방법에 따른비교 실험

다음 실험에서는 정규 분포를 나타내는 침도 값 3을 나타내는 subscription예제를 이용하여 같은 수의 그룹 수를 가질 때, 일정간격 그룹핑과 동적 재분할 그룹핑을 비교하였다. 실험 결과는 다음 표(1), 표(2)에 나타나있다.

(표 1) 일정간격 그룹핑 사용 시, 그룹핑 실험 결과

일정간격그룹핑	29	31	32	33	35	Total
4	26	26	80	80	100	312
5	10	56	56	93	103	318
6	10	56	56	93	101	316

(표 2) 동적 재분할 그룹핑 사용 시, 그룹핑 실험 결과

동적 재분할그룹핑	29	31	32	33	35	total
4	10	26	81	81	104	302
5	10	26	57	82	105	280
6	10	27	58	83	106	284

각각 4개, 5개, 6개의 그룹을 나누었을 때, 평균적으로 일정간격 그룹핑에 따른 그룹핑은 315회, 동적 재그룹핑 방법에 따른 그룹핑은 288회이다. 또한 낭비되는 매칭 횟수는 일정간격 그룹핑이 63회의 결과를 나타낼 때, 동적 재분할 그룹핑은 36회로 43%의 불필요한 매칭 횟수를 줄일 수 있다는 결과를 나타낸다. 따라서, subscription의 분포의 침도 높은 특정 부분의 그룹수를 증가시키는 방법이 더 효율적인 것으로 나타났다.

5.2.2. 침도에 따른 다양한 subscription 패턴 별 그룹핑 방법 결정 실험

위의 실험 결과에 따라 다음 실험에서는 각기 다른 침도를 가지는 subscription 예제를 이용하여 subscription분포 패턴에 따라 그룹핑 방법을 결정하기 위한 방법에 대해 연구하였다.

다양한 값의 침도에 따른 다양한 subscription 패턴 예제는 다음 표(3)과 같다.

(표 3) 다양한 침도에 따른 다양한 subscription 패턴

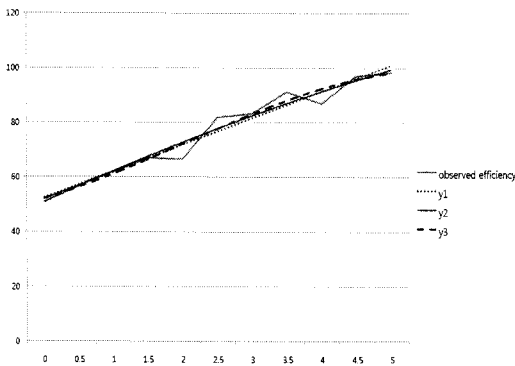
Subscription Type	Type 1	Type 2	Type 3	Type 4	Type 5
침도	2	2.5	3	3.5	4
왜도	0.232812	0.406678	0.1255	0.642479	0.490915
표준편차	2.989865	2.557046	1.489221	1.5	1.623191
분산	8.939293	6.538485	2.217778	2.25	2.634747
평균	32.51	31.87	31.38	31.25	31.54
최빈도	31	31	31	31	31

낭비되는 매칭횟수의 절약을 패턴 파악하여 어떤 그룹핑 방법을 선택할 지 여부를 결정할 기준을제안하기 위해 침도에 따라 동적 재분할 그룹핑 사용 시, 일정간격 그룹핑에 비해 침도에 따라 낭비되는 매칭 횟수의 절약을 실험하였다. A가 동적 재분할 그룹핑 사용 시 낭비되는 매칭 횟수이고 B가 일정간격 그룹핑 사용 시 낭비되는 매칭 횟수 일 때, 절약을 s는 $s = 100 - \{(A/B)*100\}$ 수식에 따라 아래의 표(4)와 같은 결과를 나타낸다.

(표 4) 동적 재분할 그룹핑과 일정간격 그룹핑 비교를 통한 절약률 패턴 분석

Subscription Type	Type 1	Type 2	Type 3	Type 4	Type 5
동적 재분할 그룹핑	16	4	4	5	4
일정간격그룹핑	48	22	24	58	31
절약률	66.70%	81.90%	83.40%	91.40%	87.10%

그림(14)는 위의 표(4)를 통해 선형회기 방식을 사용하여 절약률 패턴을 나타낸 것이다. 이러한 그림(14)를 살펴보면 침도가 높아질수록 낭비되는 매칭 횟수의 절약률이 보다 효율적이라는 결과를 알 수 있는데, 이러한 그래프를 기반으로 추세선(Trend line)을 이용하여 패턴에 대한 수식을 도출해낼 수 있다.



(그림 14) 추세선을 이용한 수식 도출

첫번째, 1차 방정식을 이용한 수식은 $y_1 = 4.860x + 47.57$ 이고 그래프와의 일치율은 $R^2 = 0.960$ 이다. 두번째, 2차 방정식을 이용한 수식은 $y_2 = -0.103x^2 + 6.098x + 44.88$ 이고 그래프와의 일치율은 $R^2 = 0.963$ 이다. 셋째, 2차 방정식을 이용한 수식은 $y_3 = -0.037x^3 + 0.568x^2 + 2.734x + 48.95$ 이고 그래프와의 일치율은 $R^2 = 0.967$ 이다. 위에서 언급한 추세선을 이용한 다항식을 사용하여, 침도 값에 따라 그룹핑 방식을 선택할 수 있는 기준을 제시하였고, 이를 통해 subscription의

분포가 특정 간격에 치우치는 패턴을 보일 경우 전체적으로 동일한 간격으로 그룹핑을 하는 방법에서 더 나아가 특정 간격에서의 그룹 수를 증가함으로써 보다 효율적인 그룹핑 방법을 제안하였다.

6. 결론

환경이나 각종 정보를 수집하는 역할을 하는 센서네트워크로부터 수집되는 센서 데이터를 효율적으로 처리, 공유하여 보다 폭넓게 활용할 수 있도록 하는 시스템의 필요가 증대되면서 정보배포 시스템의 중요성은 점차 커지고 있다. 이러한 요구에 맞추어 본 논문에서는 그리드 환경을 기반으로 센서네트워크에서 유입되는 방대한 양의 데이터를 처리 및 공유하기 위한 정보배포 시스템을 제안하였다. 또한, CGIM알고리즘을 제안함으로써 센서 데이터와 subscription과의 매칭을 보다 효율적으로 제공하도록 하여 정보배포 시스템을 보다 빠르고 정확하게 동작하도록 그 기능을 향상시켰다. CGIM알고리즘은 한가지 데이터가 들어올 때뿐만 아니라 혈압데이터와 같은 두 가지 데이터를 매칭하는 방법을 제공하고 알고리즘의 큰 변화 없이 세가지 패턴의 predicates 매칭이 가능하다는 장점이 있다. 또한 매칭을 위한 수식 ($<$, $>$ or $<$, $>$ and $<$)의 변화는 있으나, 그룹핑 방법이나 전체적인 매칭 흐름은 동일하게 적용할 수 있으며 매칭 시간을 효율적으로 줄일 수 있다는 장점을 가지고 있다. 또한 subscription의 패턴, 즉 침도에 따른 두 가지 그룹핑 방법을 제공함으로써, 앞에서 제안한 CGIM 알고리즘에 효율성을 더욱 증대할 수 있도록 하였다.

본 연구는 센서 데이터를 이용하는 실시간적인 인체정보 탐색을 통한 응급환자 정보 전달 서비스, 환자건강 모니터링서비스나 환경정보 수집을 통한 스마트 홈의 구현에 사용될 수 있을 것으로 예상된다.

Acknowledgement.

본 연구는 부분적으로 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터지원사업의 연구결과로 수행되었음. (IITA-2008(C1090-0802-0002))

참고 문헌

[1] SDI(Selective Dissemination of Information)서비스 개선방안 연구, 한국정보과학회 2004

[2] Marcos K. Aguilera et al, Matching Events in a Content based Subscription System, Proceedings of PODC, May 1999

[3] Ghazaleh Ashayer et al, Predicate Matching and Subscription Matching in Publish/Subscribe Systems, Proceedings of the 22 nd International

Conference on Distributed Computing Systems Workshops (ICDCSW'02)

[4] Antonio Carzaniga Alexander L.Wolf, "Fast Forwarding for Content-Based Networking", Technical Report CU-CS-922-01, Department of Computer Science, University of Colorado, November, 2001. Revised, September 2002.

[5] Antonio Carzaniga, Alexander L. Wolf, "Forwarding in a Content-Based Network", Proceedings of ACM SIGCOMM 2003. p. 163-174. Karlsruhe, Germany. August, 2003.

[6] Open Grid forum, <http://www.ogf.org>

[7] Vijay Dialani, Ronny Fehling et al. "INFOD Base Specification", www.ogf.org

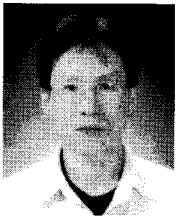
[8] "www.ogsadai.org", OGSA-DAI

◎ 저자 소개 ◎



석 보 현(Bo-Hyun Seok)

2006년 서울여자대학교 컴퓨터공학 학과 졸업(학사)
 2008년 경희대학교 일반대학원 컴퓨터공학 학과 졸업(석사)
 관심분야 : 그리드 컴퓨팅, 센서 네트워크, 정보 배포, 매칭 알고리즘 etc.
 E-mail : witcheye@paran.com



이 필 우(Pill-Woo Lee)

1991년 Univ. of Tsukuba 컴퓨터공학 학과 졸업(석사)
 1997년 Univ. of Tsukuba 컴퓨터공학 학과 졸업(박사)
 2006~현재 한국 과학기술 정보 연구원 CI 미들웨어연구팀 팀장
 관심분야 : 분산컴퓨팅, 그리드 컴퓨팅, etc.
 E-mail : pwlee@kisti.re.kr



허 의 남(Eui-Nam Huh)

1990년 부산국립대학교 컴퓨터공학 학과 졸업(학사)
 1995년 The University of Texas at Arlington 컴퓨터공학 학과 졸업 (석사)
 2002년 The Ohio University 컴퓨터공학 학과 졸업 (박사)
 2005년 9월~현재 경희대학교 컴퓨터공학과 부교수
 관심분야 : 분산컴퓨팅, 센서 네트워크, 보안. etc
 E-mail : johnhuh@khu.ac.kr