

u-Health care 를 위한 publish/subscribe 시스템에서의 효율적인 매칭 메커니즘

석보현*, 이필우**, 허의남*

*경희대학교 컴퓨터공학과

**한국과학기술정보연구원

e-mail : bohyun@icns.khu.ac.kr*, johnhuh@khu.ac.kr* , pwlee@kisti.re.kr**

An Efficient Matching Mechanism in Publish/Subscribe System for U-Health care

Bo-Hyun Seok*, Pill-Woo Lee**, Eui-Nam Huh*

*Dept. of Computer Science, Kyung-Hee University

**KISTI

요 약

실 시간적인 데이터의 수집과 더불어 수집한 데이터의 실 시간적인 전송을 기반으로 정보를 보다 폭넓게 활용할 수 있는 환경을 제공하기 위해 시스템에서 자동적으로 정보를 배포해주는 Publish/Subscribe 시스템에 대한 요구가 증대되고 있다. 이러한 pub/sub 시스템은 사용자의 요구사항을 미리 저장, 이를 이용하여 수집되는 정보와 사용자의 요구와 사용자를 찾아 배포해주는 방법을 사용하는데, 이때, 일치 여부를 확인하는 매칭 과정에서의 많은 자원과 시간의 소모가 문제점으로 대두되고 있다. 따라서, 논문에서는 보다 효율적으로 데이터와 범위를 이용하여 나타내는 사용자의 요구를 매칭하는 방법을 제공하는 CGIM 알고리즘을 제안하였다.

1. 서론

유비쿼터스 환경에서 정보의 수집은 사용자들에게 보다 쾌적하고 편리한 환경을 제공하기 위한 가장 기본적인 요구사항이라고 할 수 있다. 이러한 요구에 따라 많은 정보를 실 시간적으로 배포하기 위해 센서 네트워크에 대한 연구가 활발하게 진행 중이며 센서들을 이용하여 자연 재해나 각종 위험들을 방지하기 위해 자연 환경 정보를 수집하거나, 노인이나 환자의 지속적인 관리와 위급상황이 발생하는 경우 인명구조, 실내에서의 자동 조명이나 환기 조절 등 보다 쾌적한 환경을 위해서 사용될 수 있다.

이렇듯 센서네트워크의 발전에 따라 수집되는 정보의 종류와 양이 증가하고 기하급수적으로 증가되는 정보의 양에 비례하여 정보의 수요 또한 증가하는 이러한 환경에서 정보로의 실시간적인 접근은 정보를 이용하는 이용자들에게 매우 중요한 기능이라고 할 수 있다. 그러나 현재 정보를 얻기 위해 가장 많이 사용되는 방법은 직접 홈페이지나 정보검색시스템에 접근하여 정보를 검색하는 것으로 실시간적으로 필요로 하는 정보를 제공받는데 한계가 있다. 이러한 문제를 해결하고자 자동적으로 사용자들에게 정보를 배포해주는 정보배포 시스템, 즉, publish/subscribe 시스템에 대한 연구가 현재까지 활발하게 진행되고 있다 [1]. Pub/sub 시스템은 미리 사용자의 요구사항을 시스템에 등록 받고 이러한 요구정보를 이용하여 수집한

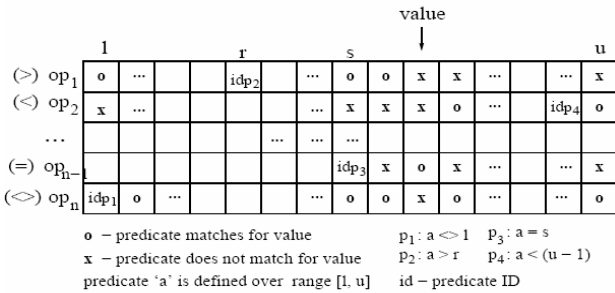
정보와 요구사항이 일치하는 사용자에게 배포해주는 기능을 한다. 이 때, 수집하는 정보와 많은 사용자의 요구사항들을 일일이 매칭시키는 과정에서 많은 자원과 시간을 소모한다는 문제점을 가지고 있다. 이러한 문제점들을 해결하고자 효율적으로 매칭 과정을 줄이고자 하는 연구가 진행되고 있으나, 본 논문의 기본 환경이 되는 u-Health care 시스템에서는 사용자의 요구를 범위를 이용하여 명세하는데, 이러한 범위를 사용하는 요구사항을 매칭하는 연구는 매우 미미한 실정이다.

따라서 본 논문에서는 pub/sub 시스템에서 정보가 범위를 이용한 사용자의 요구사항과 일치하는지 매칭하고 이벤트 여부를 결정하기 위해 효율적인 매칭을 제공하는 CGIM(Classed Group Index Matching) 매칭 알고리즘을 제안한다.

2. 관련연구

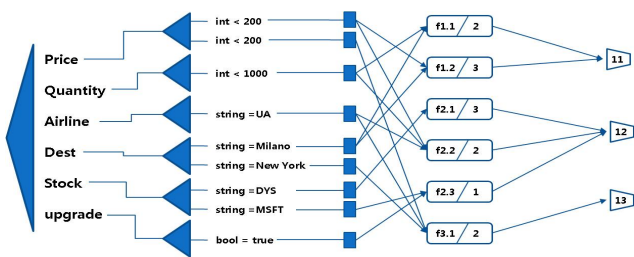
Pub/sub 시스템에서 효율적으로 매칭과정을 수행하기 위한 방안들에 대한 연구들이 진행되고 있다. 현재까지 진행된 매칭 관련 연구로는 첫째, predicate 그룹핑을 이용한 매칭 알고리즘[2], 둘째, predicate 값의 인덱스를 이용한 매칭 알고리즘[3-4] 등이 있다. 이번 섹션에서는 각 알고리즘들에 대해 알아보하고자 한다. [2]에서는 동일한 predicate 를 가지는 subscription 으로 불필요하게 낭비되는 predicate 매칭을 효율적

으로 줄여보고자, 같은 predicate 들을 그룹핑하는 방법을 제안하고 있다. 그림(1)과 같은 형태로 predicates 들을 operation 별로 인덱싱하고 이를 predicate id 와 subscription id 를 상호 연계하여 predicate 그룹에 해당하는 subscription 들을 찾을 수 있도록 하였다.

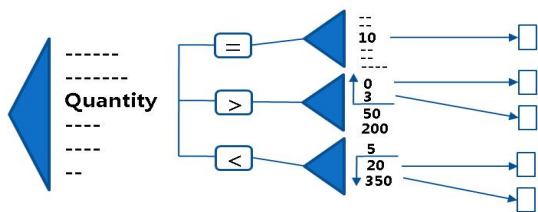


(그림 1) Index of predicates

[3-4]에서 제안하는 forwarding algorithm 은 그림(2)에서 보이는 바와 같이 attribute 별 인덱싱 및 그룹핑을 통해 다양한 attribute 를 가지는 predicate 를 찾아낸다. predicate 들은 여러 attribute 를 가지는데, 각 predicate 별로 각각 다른 종류와 수의 attribute 를 가질 수 있다. 같은 attribute 와 value 값을 가지는 경우를 제외한 모든 attribute 들을 indexing 하여, 데이터와 일치하는 attribute 에 해당하는 predicate 를 찾는다. 그림(3)은 Indexing 한 attribute 별로 predicate 의 value 들을 크기 순으로 정렬하고 정렬된 value 값을 이용하여 데이터와 일치하는 값부터 value 값의 끝까지 순차적으로 매칭한다. Value 값이 일치하면 해당 attribute 를 가지는 predicate 를 찾는다.



(그림 2) Forwarding Algorithm attribute tree



(그림 3) Forwarding Algorithm predicates groups

앞에서 설명한 연구들에서 제시하는 매칭 알고리즘들은 단일 operation 에 의한 predicate 들의 매칭을 제공하는데, 이러한 알고리즘은 두 개의 범위를 조합한 predicate 매칭에는 매우 취약하다는 단점이 있다. 복합 predicate 를 이용한 subscription 명세가 필요한 환

경에서는 단일 predicate 를 이용한 위의 알고리즘을 사용할 경우, 각각의 operator 에 대한 결과값을 조합하는 과정이 필요한데, 이는 불필요한 자원과 시간의 낭비를 초래한다.

u-Health care 시스템에서와 같이 바이오 센서 데이터를 필요로 하는 실제 요구 환경에서는 보다 효율적으로 사용자의 요구를 정의하기 위해 복합적인 범위로 정의된 predicate 를 요구 하기 때문에 기존의 연구에서 진행된 매칭 알고리즘을 적용시키는데 한계가 있다. 따라서 본 논문에서는 u-Health 환경에 맞추어 범위 조건을 매칭시키는 효율적인 매칭 알고리즘을 제안하고자 한다.

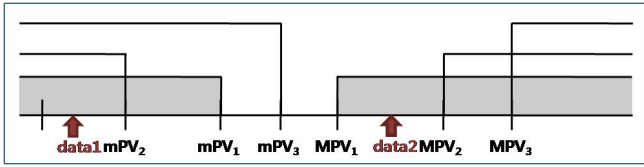
3. CGIM(Classed Group Index Matching) 알고리즘

본 논문의 기본 환경이 되는 u-Health care 정보매포시스템[5]에서 사용자는 원하는 종류/범위의 센서 데이터를 실시간적으로 받을 수 있어야 하는데, 이를 위해 사용자는 자신의 요구를 명세하는 subscription 을 등록한다. Subscription 의 조건(condition) 과 센서 데이터를 비교, 매칭하기 고려사항은 다음과 같다. 첫째, u-Health 시스템에서 제공하는 data 의 종류는 센서의 종류에 따라 여러 가지가 될 수 있다. 둘째, 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. 셋째, 센서 데이터가 들어왔을 때, 시스템의 모든 subscription 과 일일이 매칭하게 되면 특히 범위를 이용한 요구사항을 사용하는 subscription 매칭을 수행할 때, 이벤트 데이터를 찾는데 많은 시간과 자원이 소모될 것으로 예측된다. 이에 따라, 방대한 양의 subscription 과의 매칭을 위해 보다 효율적인 매칭 알고리즘이 필요하다.

3.1. 효과적인 매칭을 위한 predicate 그룹핑

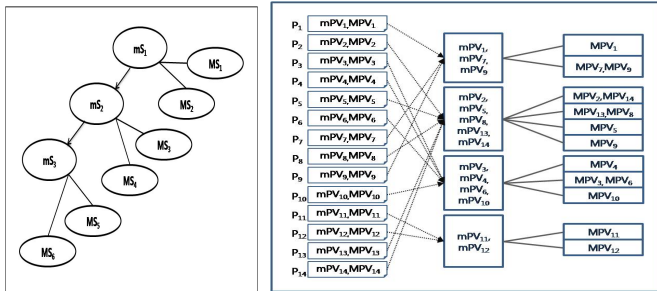
Subscription 은 사용자의 요구조건을 나타내는 predicate 로 이루어지는데 predicate 는 5 개의 속성으로 구성된다. 첫째, event_type 은 데이터의 종류를 명세하는 부분이고, 둘째, N 은 범위를 선택하는 방법에 대해 나타낸다. 셋째, op 는 <<또는 <>와 같은 범위를 나타내는 부등호 operation 을 나타내고, 넷째, mPV 는 범위의 최소 값을, 다섯째, MPV 는 범위의 최대값을 나타낸다. 이러한 범위는 mPV<data<MPV 로 나타낼 수 있다.

본 논문에서는 u-Health care 시스템에서 아래 그림(4)에서와 같이 두 종류의 데이터를 수집하고 동시에 각기 다른 범위의 조건에 복합적으로 일치하는 (mPV>data1 || MPV<data2)로 나타나는 범위를 이용하여 알고리즘을 설계하였다. 아래 그림(4)를 살펴보면, 데이터가 들어왔을 때, data1 은 mPV₁, mPV₂, mPV₃ 이 모두 일치하지만, data2 는 MPV₁만 해당되는걸 볼 수 있다. 따라서 아래 그림(4)와 같은 경우 결과는 data1 과 data2 가 모두 일치하는 (mPV₁, MPV₁)만 일치한다.



(그림 4) predicate 사례

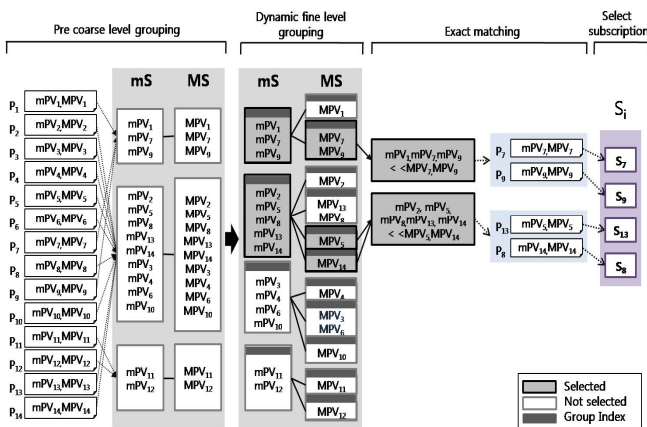
CGIM 알고리즘에서는 실제 매칭 수행 단계에 들어가기 전에 predicate 를 그룹핑하고 인덱싱하여 일치하는 predicate 의 범위를 줄여나가면서 찾는 방법을 사용한다. 아래 그림(5)는 CGIM 에서의 매칭을 위한 그룹핑을 개념적으로 나타낸 것이다. mS 는 각각의 predicates 들이 포함하는 mPV 들의 모임 ($mS = \{mPV_1, mPV_2, \dots, mPV_i\}$) 이고, MS 는 MPV 의 모임 ($MS = \{MPV_1, MPV_2, \dots, MPV_i\}$) 이다. 본 논문에서는 mPV 의 값에 따라 그룹 mS_i 로 그룹핑하고 이러한 mS_i 내에서 다시 MS_i 로 하위 그룹을 그룹핑 하였다.



(그림 5) predicate 트리

3.2. CGIM 알고리즘 동작과정

CGIM 알고리즘의 동작과정은 다음 그림(6)과 같다.



(그림 6) CGIM 동작과정

▶ Pre coarse level grouping: predicate 들의 mS(mPV 그룹)를 그룹핑하고 거기에 종속된 MS(MPV 그룹)를 그룹핑한다. 생성된 그룹들은 인덱싱한다.

▶ Dynamic fine level grouping: Pre coarse level grouping 에서 생성된 predicate 들의 그룹들을 이용하여 일차적으로 해당되는 mS 그룹을 매칭하고 이렇게 선별된 mS

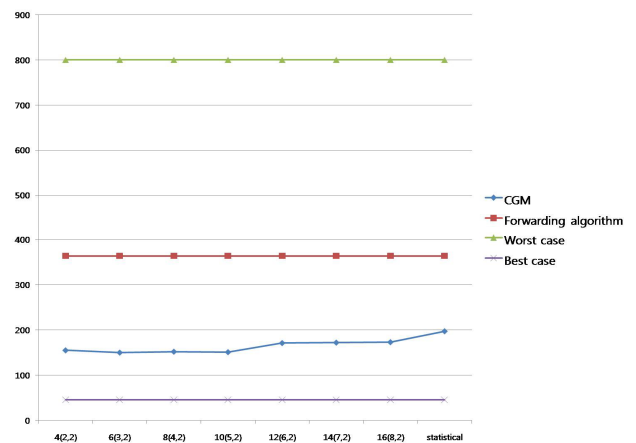
그룹 내에 종속되는 MS 그룹을 선별한다.

▶ Exactly matching: 최종적으로 선별 된 그룹 내에서 실질적으로 정확하게 매칭하는지 순차적인 매칭을 수행한다.

▶ Select Subscription: 정확하게 매칭되는 predicate 들의 id 를 이용하여 해당 subscription 을 발체한다. 발체한 subscription 은 데이터를 전송받을 사용자와 전송할 방법을 선택하는데 이용된다.

4. 실험

본 논문에서 제안하는 알고리즘의 성능을 평가하기 위해 100 개의 subscription 과 5 쌍의 센서 데이터를 매칭시키는 사례로 각각 ($mPV=29, MPV=38$), ($mPV=29, MPV=33$), ($mPV=33, MPV=37$), ($mPV=32, MPV=35$)를 이용한 5 개의 예제를 실행하였다. 아래 그림(7)에서 나타나 있는 그래프는 같은 subscription 과 데이터 예제를 사용한 Forwarding 알고리즘과 CGIM 알고리즘의 실험 결과를 비교하여 나타낸 것이다.



(그림 7) 실험 결과

그림(7)에서 나타내는 바와 같이 Forwarding 알고리즘을 이용하여 매칭할 때 364 번의 매칭 과정이 소요되는데, CGIM 알고리즘을 사용하면 평균 165 번의 매칭으로 약 2 배가량 매칭 시간을 줄일 수 있게 된다. 불필요하게 낭비되는 매칭 수를 검사해보면 Worst case 에서는 754 번, Forwarding 알고리즘은 318 번의 매칭과정이 낭비되지만, 본 논문에서 제시하는 알고리즘을 사용하면 평균 119 번으로 Worst case 보다 85%, Forwarding 알고리즘보다 63%의 낭비과정을 줄일 수 있다는 결과를 나타낸다.

5. 결론

유비쿼터스 시대에 센서 네트워크를 비롯한 각종 데이터 수집환경에서 사용자들에게 자동적으로 필요로 하는 정보를 배포해주는 pub/sub 시스템의 효율성을 높이고자 CGIM 알고리즘을 제안함으로써 데이터와 subscription 과의 매칭을 보다 효율적으로 제공하도록 하였다. 본 논문에서 제안하는 CGIM 알고리즘은 혈압

데이터와 같이 두 가지 데이터를 매칭할 때, 즉, 두 개의 단일 조건이 복합적으로 일치해야 할 경우 매칭 시간을 효율적으로 줄일 수 있도록 그 기능을 제공하며 정보 배포 시스템을 보다 빠르고 정확하게 동작하도록 그 기능을 향상시켰다.

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.

참고문헌

- [1] SDI(Selective Dissemination of Information)서비스 개선방안 연구, 한국정보과학회 2004
- [2] 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)
- [3] 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.
- [4] Antonio Carzaniga, Alexander L. Wolf, "Forwarding in a Content-Based Network", Proceedings of ACM SIGCOMM 2003. p. 163-174. Karlsruhe, Germany. August, 2003.
- [5] 석보현, 이필우, 허의남, "그리드상에서의 센서 데이터를 위한 관리 및 배포 시스템 설계", 한국정보처리학회 춘계학술발표대회 제 14 권 제 1 호,2007