

## 분산 객체의 확률적 비례 검색 기반 전송률 향상 검색 알고리즘

김 분 회\*

### Search Algorithm for Advanced Transmission Rate based on Probabilistic Proportion Search of Distributed Objects

Boon-Hee Kim\*

#### 요 약

P2P 분산 시스템의 가장 큰 특징은 해당 피어들이 항상 온라인 상태일 것이라는 보장이 없다는 것이다. 즉 P2P 시스템을 이용할 때에는 해당 피어로부터 파일을 다운로드 받다가 다운로드 되지 않는 경우가 발생하게 된다. 이를 해결하기 위한 연구의 대부분은 재전송이라는 방법에 의존하고 있다. 이는 P2P 시스템의 성능 저하의 원인이 되므로 이에 대한 해결책이 필요하다. 본 연구에서는 해당 P2P 시스템을 이용하는 사용자의 평균 이용 시간대의 분석 자료를 자원 제공자 선택의 기준으로 적용하여 자원 전송 보장성을 높이고, 또한 인기도 높은 자원에 대해서 자료 전송 기회를 높여주는 역할의 기존의 분산 객체 리플리케이션 기법들과의 조합에 의한 분산 객체 전송률이 향상된 검색 알고리즘을 제안한다.

#### Abstract

A special feature of P2P distributed system isn't always the guarantee of online status for peers. In other words we want to download the file from the peer when we use the P2P system but it sometimes caused this system to fail the download. Many studies to resolve this problem depend on re-transmission method. It caused to lower performance so we have to resolve this problem. In this study, we analysis the average usage time of P2P application user and raise the resource transmission guarantee to apply the selection criteria of resource supplier. Moreover the combinations of distributed object replication techniques, the role to enhance the data transmission opportunity of high popularity resource, will cause this search algorithm to advance.

▶ Keyword : 재전송(Re-transmission), 분산 객체(Distributed Object), 확률(Probability)

---

• 제1저자 : 김분회  
• 접수일 : 2006.04.30, 심사일 : 2006.06.23, 심사완료일 : 2006.07.05  
\* 동명대학교 멀티미디어공학과 전임강사

## I. 서론

기존 P2P 환경에서의 자원 검색 알고리즘은 자원을 보유한 피어를 선택하는데 있어 다양한 연구가 진행되어왔다 [1][2][6][7][8][10]. 최소 응답시간을 보장하는 피어를 선택하거나 자원과의 거리와 관련된 표를 이용하여 거리가 가장 가까운 피어를 선택하는 알고리즘이 제안되고 있다. 그리고 응답시간의 평균값을 구하여 해당 피어의 가치를 측정하고 이러한 값을 이용하여 피어를 선택하는 알고리즘, 여러 가지 데이터를 수집하여 회귀분석하고 서로 다른 피어에 메시지를 전달하기 위한 시간을 예측하여 가장 낮은 전달 시간을 가지는 피어를 선택하는 알고리즘, 네트워크 지연 시간 등을 관찰하여 응답시간이 가장 작은 피어를 선택하는 알고리즘 등이 있다. 이런 알고리즘들이 해당 분야에서 주요 알고리즘으로 평가되고 있는 내용들이다. 그러나 이러한 노력의 대부분은 하나의 자원 요청 피어 당 하나의 자원 지원 피어를 할당하고 요구에 대한 서비스가 일어나는 동안 전송 실패가 일어날 가능성을 고려하지 않고 있다. 즉 P2P 분산 시스템의 가장 큰 특징은 해당 피어들이 항상 온라인 상태일 것이라는 보장이 없다는 것이다. 이로 인해 P2P 시스템을 이용할 때에는 해당 피어로부터 파일을 다운로드 받다가 다운로드 되지 않는 경우가 발생하게 된다. 이를 해결하기 위한 연구의 대부분은 재전송이라는 방법에 의존하고 있다. 이는 P2P 시스템의 성능 저하의 원인이 되므로 이에 대한 해결책이 필요하다.

본 논문에서는 기존의 TO[1] P2P 자원 검색 알고리즘 기반 하에 자원 전송률을 높이기 위한 검색 알고리즘을 제안한다. 본 연구에서는 해당 P2P 시스템을 이용하는 사용자의 평균 이용 시간대의 분석 자료를 자원 제공자 선택의 기준으로 적용하여 자원 전송 보장성을 높이고, 또한 인기도 높은 자원에 대해서 자료 전송 기회를 높여주는 역할의 기존의 분산 객체 리플리케이션 기법들과의 조합에 의한 분산 객체 전송률이 향상된 검색 알고리즘을 제안한다. 제안한 검색 알고리즘의 성능 평가 결과 각 피어의 검색 성공률 측면에서 의미 있는 결과를 확인할 수 있었다. 본 논문의 구성은 다음과 같다. 2장에서 관련연구, 3장에서는 본 논문의 핵심 부분인 제안한 검색 알고리즘을 4장에서 이에 대한 평가를, 마지막으로 5장에서 결론을 맺는다.

## II. 관련 연구

이 장에서는 본 논문의 자원 검색 알고리즘으로써의 기반이 되는 TO 알고리즘에 대해 설명하고자 한다. TO 알고리즘은 해당 논문에도 기술되었듯이 브로드캐스팅 기반 검색 알고리즘[3][5] 가운데 검색 성공률 측면에서 가장 우수한 APS 알고리즘과 유사한 성능을 보이며, 특정 실험 상태에서는 동등한 성능 이상의 결과를 보이며 의미 있는 결과를 나타냈다. 제안된 P2P 기반 검색 알고리즘은 K Walks 기반 P2P 시스템의 자원 검색 과정에서 발생하는 네트워크 트래픽 양을 최소화한 것을 목표로 한다. 제안된 P2P 기반 약결합 시스템에서 검색 기법의 주요 설계 목표는 순수 P2P 모델 기반 자원 검색 과정에서 발생하는 네트워크 트래픽 양을 고려한 효과적 검색 지원, 자원 검색 횟수가 증가할수록 해당 자원 보유 피어로 검색 방향성이 결정되어 자원 보유 피어의 증가된 근접성 지원, 자원 검색 횟수의 증가에 따라 학습 작용이 일어나 적중률 증가시키는 것이다. 제안한 검색 기법은 운영 환경 상에서의 몇 가지 가정 하에 설계되었다. 먼저 자원 요청을 위한 검색 진행 메시지의 발생 및 진행의 과정에서 초기 이웃 피어의 선택과정은 K Walks의 기본 원리를 따른다. 그리고 제안한 P2P 시스템을 실제로 구현하여 실험하기에는 객관적인 실험 결과 값을 얻기가 어려우며 타 시스템과의 비교 분석에 상당히 복잡한 메커니즘이 필요하므로 이 연구에서는 논외로 한다. 이는 P2P 검색 분야의 일반적인 가정이다. 제안한 알고리즘은 독자적인 검색 진행 메시지의 발생 및 진행의 과정에서 초기 이웃 피어의 선택과정은 K Walks의 기본 원리를 적용한다. [그림 1]에서 제안한 알고리즘을 표현하고 있는데, 기본적으로 검색을 진행하고자 하는 주체 피어는 찾고자 하는 자원에 대해 하나의 엔트리로 구성된 지역 인덱스 유지한다. 또한 검색의 과정을 거치게 되는 해당 이웃 피어들도 각각의 이웃 피어에 대한 지역 인덱스 값을 유지하게 된다. 이 지역 인덱스는 각 자원 당 하나로 배당되며, 인덱스에 저장되는 값은 이웃피어와의 연관성 확률 값을 반영한 값이다.

[그림 1]에서 제안한 알고리즘을 보여주는데 기본적으로 순수 P2P 모델을 기반으로 하며, P2P 오버레이 네트워크의 토폴로지는 대부분의 P2P 검색 논문에서 적용한 랜덤 그래프 모델을 바탕으로 하며, 질의와 자원의 분포는 Zipfian 이외의 실제 P2P 시스템에서 보여 지는 분포와 유사한 다양한 분포 법칙을 적용한다. 기본 입력값으로 모든

자원의 수를 나타내는 값이 요구되며, 보유 자원의 종류에 따라 인덱스로 구분하게 되고, 자원을 보유한 전체 노드의 수를 파악할 수 있다. 전체 노드 수는 오버레이 네트워크를 구성하는데 필요한 전체 노드 수로 초기 네트워크 구조를 설정하는데 필요조건이다. 각 노드의 평균 차수는 전체 오버레이 네트워크를 랜덤 그래프 모델로 설정함에 있어 노드 간의 링크 관계를 설정하는데 있어 필수 요소이다. Walker는 질의 발생 노드의 K 수만큼 질의가 발생되어 K 수에 따라 전체 메시지 발생량에 큰 영향을 줌으로 적절한 수의 K를 결정하는데 있어 주요한 역할을 한다.

```

The_TOX ) {
Initialization : pure P2P, random graph model
Definition : Qinode, Rjnode, i, K, Niindex
For (i = 0 ; i <= K; i++) {
  If (Niindex = null) {
    Original_Neighbor_Select;
    Random_Walks_Rule; }
  else {
    Max_size(all_of_Niindex);
    Neighbor_Select; }
  If (Qinode = Rjnode) {
    Backward_Forwarding(update_index);
    For (All neighbors of the hit node) {
      TO_hit_ratio_update
      constantV+recentTOHitRatio(0.1) }
    }
  When Broadcast_Flooding(Kwalker_success);
}
    
```

그림 1. TO 알고리즘  
Fig. 1. TO Algorithm

이웃노드에 관한 연관성 확률값 보유 인덱스인 Niindex의 값이 초기값 즉 null 값을 가지고 있는 경우 Random Walks의 규칙에 따라 다음에 진행할 이웃노드를 선택하게 된다. 그렇지 않을 경우는 Niindex의 값을 분석하여 연관성 확률이 가장 높은 값에 해당하는 이웃 피어로 진행된다. 여기서 동일한 연관성 확률값을 가진 인덱스가 여럿 있을 경우는 임의의 이웃을 선택한다. 이와 같은 초기화 작업 후 검색의 과정을 진행하는 Walker를 기반으로 실제적인 검색 과정이 시작되는데, 질의의 분석 과정에서 지역 인덱스의 연관성 확률 값을 바탕으로 상위에 있는 이웃피어를 선택하여 질의가 전파된다. 이때 원하는 자원을 발견하였을 시 질의 진행 방향과 역방향으로 연관성 확률 값을 갱신하면서 질의 발생 피어에 도달하게 된다. 즉 질의를 발생한 노드 Qjnode와 부합하는 자원을 지닌 노드 Rjnode을 찾은 경우 역방향으로 연관성 확률 값을 갱신하면서 질의 발생 피어에 도달하

게 되는 것이다. 여기서 constantV는 갱신단위 연관성 확률값으로 식 1)의 계산 값에 따르며, recentTOHitRatio는 이미 그 피어에 있는 연관성 확률 값을 표현한 것으로써 0과 1사이의 값을 보유하고 있다. 제안한 알고리즘의 기본 과정은 이러한 단계를 거치게 된다. 연관성 확률 값은 0과 1사이의 값으로써 검색의 결과가 적중되었을 때 해당 Walker에 의해 적중된 피어에서 검색 주체 피어의 방향으로 갱신함으로써 자원을 찾은 Walker의 검색 경로에 위치한 피어는 그 확률 값이 증가된다. 이 연관성 확률 값은 초기값은 0으로 셋팅되어 있지만 검색 적중률에 따라 비례적으로 증가된다. 검색 주체 피어가 원하는 자원에 대해 검색하고자 할 때 우선 작업이 바로 유지하고 있는 지역 인덱스의 해당 자원에 대한 연관성 확률 값을 기반으로 상대적으로 1에 더 가까운 이웃피어를 다음 경로로 선택하는데 이용하는 것이다.

갱신 단위 연관성 확률 값( $\alpha$ ) =  $1/($ 질의 발생 피어의 수 X Walker 수), ( $1 \leq$  Walker  $\leq 15$ )—식 1)

식 1)은 연관성 확률값을 갱신하는데 있어 기존의 값에 합할 단위를 나타낸다. 갱신 단위 연관성 확률값은 질의를 발생시킬 피어의 수에 하나의 피어 당 발생시킬 Walker의 수에 비례한 값을 1로 나눈 값으로써, 하나의 피어가 모든 질의에 대해 질의 파생 경로 내에 포함되고 모든 Walker의 진행 경로에 포함한다 하여도 그 결과 나타나는 최종 연관성 확률 값의 총 합이 1을 넘지 않는 범위이다. 본 논문에서는 연관성 확률 값이 1에 해당하는 경우를 OK 인수로 칭하며, 이는 검색 주체 피어에서 요청한 해당 자원이 위치해 있는 피어의 모든 이웃피어에게 OK값을 부여한다. 이는 검색 주체 피어가 해당 자원이 있는 피어로의 근접성을 높이기 위한 정책으로써 기본 메시지 발생량에 자원 보유 피어의 이웃 피어의 개수 이하의 발생량이 부가적으로 발생된다. 이러한 메시지 발생량과 근접율의 증가로 인해 P2P 시스템의 성능에 어떠한 영향을 미칠지에 대해서는 실험 및 평가를 다루는 절에서 다루겠다. 요청을 전달하기 위해 N개의 이웃 피어에서 K개를 선택하고, K의 개수는 선택에 따라 달라지겠지만, 기본적으로 K의 값보다 피어의 N 값보다 크면, 전체 이웃 피어 개수만큼의 Walker가 생성된다. 또한 새로운 이웃 발견 시 해당 자원의 인덱스는 초기값 0으로 할당하고, 피어의 연결 종료 시 관련 이웃 피어의 엔트리를 모두 삭제하는 APS의 기본 프로토콜을 따른다.

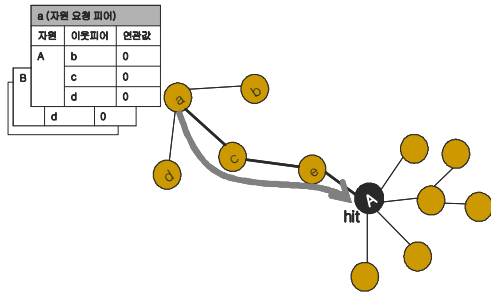


그림 2. 자원과 인덱스 구성  
Fig. 2. A formation of resource and index

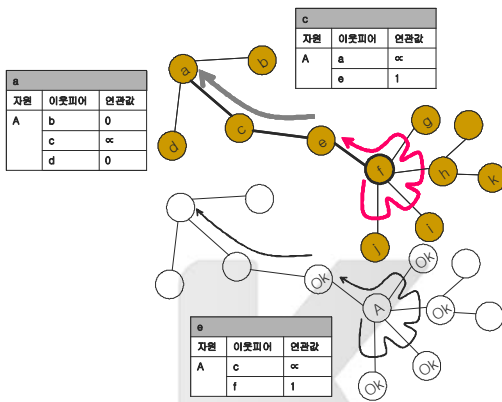


그림 3. 지역 인덱스 구조  
Fig. 3. Structure of local index

[그림 2]의 예에서처럼 A 자원을 찾는 시점에서 자원 요청 피어가 있는 a 피어 방향으로 각 피어들의 지역 인덱스 값을 갱신하여야 하는데, [그림 3] 그 예이다. 자원을 보유한 f 피어를 기점으로 이웃피어 e, g, h, i, j의 A 자원에 대한 이웃피어 f의 연관성 확률값을 1로 갱신한다. 제안한 알고리즘에 표현한 Ok 값이 연관성 확률값 1을 나타낸다.  $\infty$  값은 식 3-1)에서 표현한 갱신단위 연관성 확률값에 이전에 저장되어 있던 값을 합한 결과이다. e 피어의 경우 이웃 피어인 c에 대해서는  $\infty$  값을 f 피어에 대해서는 Ok 값을 보유하고 있고, a의 경우 이웃 피어 b, c, d 가운데 c 피어만  $\infty$  값을 가지고 있고, 나머지 b, d는 이전과 그대로인 0 값을 저장하고 있다. 여기서 표현된  $\infty$  값은 각 피어, 자원, 이웃 피어의 종류에 따라 그 결과값은 모두 자체 저장되어 있는 값에서 갱신단위 연관성 확률값을 합한 값이므로 각기 다른 값을 의미한다.

브로드캐스트 검색을 지원하는 P2P 시스템들의 메시지

발생량을 줄이기 위해 해당 피어 보유한 이웃 피어와의 차수 가운데 일부분을 선택하여 검색이 진행되는 연구가 요구되지만 어렵게 찾아진 자원 보유 피어에 대한 정보를 다수의 피어의 다양한 방향에서의 접근될 수 있도록 근접성을 넓히는 것은 다양한 검색 성능에 영향을 준다. 제한된 검색 알고리즘의 근접성을 개선하기 위한 검색 모델로 확장하고자 한다. 검색 진행 메시지의 발생 및 진행의 과정에서 초기 이웃 피어의 선택과정은 K Walks의 기본 원리를 적용하며, [그림 4]에 제안한 알고리즘을 표현하고 있다. 기본적으로 검색을 진행하고자 하는 주체 피어는 찾고자 하는 자원에 대해 하나의 엔트리로 구성된 지역 인덱스 유지하며, 검색의 과정을 거치게 되는 해당 이웃 피어들도 각각의 이웃 피어에 대한 지역 인덱스 값을 유지하게 한다. 또한 이 지역 인덱스는 각 자원 당 하나로 배당되며, 인덱스에 저장되는 값은 이웃피어와의 연관성 확률 값 반영한 값이다. 이러한 초기화 작업을 거쳐서 검색의 과정을 진행하는 Walker를 기반으로 실제적인 검색 과정이 시작되는데, 질의 분석 과정에서 지역 인덱스의 연관성 확률 값을 바탕으로 우위에 있는 이웃피어를 선택하여 질의가 전파된다. 이웃노드에 관한 연관성 확률값 보유 인덱스인 Niindex의 값이 초기값 즉 null 값을 가지고 있는 경우 Random Walks의 규칙에 따라 다음에 진행할 이웃노드를 선택하게 된다. 그렇지 않을 경우는 Niindex의 값을 분석하여 연관성 확률이 가장 높은 값에 해당하는 이웃 피어로 진행된다. 여기서 동일한 연관성 확률값을 가진 인덱스가 여럿 있을 경우는 임의의 이웃을 선택한다. 이와 같은 초기화 작업 후 검색의 과정을 진행하는 Walker를 기반으로 실제적인 검색 과정이 시작되는데, 질의 분석 과정에서 지역 인덱스의 연관성 확률 값을 바탕으로 우위에 있는 이웃피어를 선택하여 질의가 전파된다. 이때 원하는 자원을 발견하였을 시 질의 진행 방향과 역방향으로 연관성 확률 값을 갱신하면서 질의 발생 피어에 도달하게 된다. 즉 질의를 발생한 노드 Qinode와 부합하는 자원을 지닌 노드 Rinode을 찾는 경우 해당 이웃 노드에 대해 연관성 확률 값이 1에 해당하는 OK 인수를 검색 주체 피어에서 요청한 해당 자원이 위치해 있는 피어의 모든 이웃피어에게 부여한다. 이는 연관성 확률 값의 최고치를 부여함으로써 주변 피어에서의 근접율을 높여준다. 이웃 피어를 제외한 질의 진행 역방향에 위치한 해당 피어들은 갱신단위 연관성 확률 값 constantV과 이미 해당 피어에 있는 연관성 확률 값을 나타내는 0과 1사이의 값인 recentTOHitRatio의 곱합 연산에 의해 인덱스 갱신 과정이 진행된다.

### III. 제안된 알고리즘

본 논문에서는 기존의 TO 자원 검색 알고리즘 기반 하에 자원 전송률을 높이기 위한 검색 알고리즘을 제안한다. 본 알고리즘은 피어들의 온라인 상태의 패턴을 분석하여 온라인 상태일 확률값을 기준으로 최적의 피어를 선택하여 해당 자원에 대한 재전송 가능성을 최대한 줄이고자한 검색 기법이다. 먼저 해당 P2P 네트워크를 구성하는 각 피어들에 대하여 일정 기간(Fixed.date)동안 [그림 4]에서 제안한 P2P 시스템 사용시간을 분석하는 과정을 거친다.

```

The_Probabilistic_Proportion_Count(int a) {
Session Start()
Update S(a)= Check.Start_Time;
If(Check.date >= Fixed.date) {
for(b=1;b>=Fixed.num; b++){
sum = sum + S(b);
}
average_start_time = sum/Fixed.date;
}
End
Session Terminate()
Update E(a)=Check.End_Time;
If(Check.date >= Fixed.date) {
for(b=1;b>=Fixed.num; b++){
sum = sum + E(b);
}
average_end_time = sum/Fixed.date;
}
End
}
    
```

그림 4. 사용 시간 분석  
Fig. 4. Analysis for the usage time

해당 시스템에 접속하게 되면 각 사용자의 세션이 시작되고, 해당 시스템에서 로그아웃 되면 세션이 종료하게 된다. 시작 세션에서는 시작 시간을 저장하고, 저장된 세션 시작 시간에 대한 정보들은 일정 기간이 지나면, 전체 시작시간을 합하여 정해진 기간으로 평균을 낸다. 이것을 평균 시작시간으로 한다. 평균 시작 시간을 정하는 방법은 다양하게 결정할 수 있겠지만, 본 논문에서는 전체 시작시간을 합하여 정해진 기간으로 평균을 평균 시작시간으로 단순화 하였다. 그리고 세션이 종료되면 마찬가지로 세션 종료 시간을 일정 기간동안 저장하였다가 해당 시점에 평균 종료 시간을 구하게 된다.

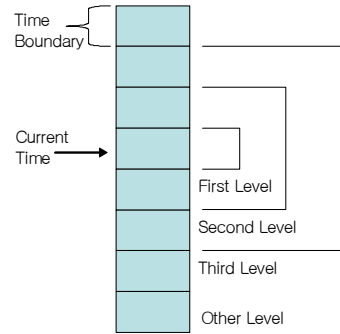


그림 5. 시간 영역 결정  
Fig. 5. Decision of the time boundary

[그림 5]는 [그림 4] 알고리즘에 의해 결정된 각 피어들의 온라인 상태에 있을 확률값을 바탕으로 현재 P2P 시스템에서 해당 자원을 다운받기위해 검색 작업에 들어간 피어의 현재 접속 시간(Current Time)을 나타낸다. 이 현재 접속 시간을 바탕으로 이미 설정된 시간 영역대(Time Boundary)의 기준 영역에 수준을 단계별로 차별화하였다. 첫 번째 단계(First Level)는 해당 피어의 현재 접속 시간이 포함된 동 시간 영역대에 해당하고, 두 번째 단계는 결정된 첫 번째 단계를 기준으로 상하 한 단계씩 추가한 시간 영역대가 된다. 이런 식으로 레벨을 차별화 하는데 있어 본 실험에서는 해당 피어가 소속된 시간 영역은 [그림 4]의 알고리즘을 기준으로 결정된 시작 시간과 끝시간의 중간값으로 결정하였다. 이는 다양한 기준으로 결정 가능한 유연한 영역으로 차후 연구에서는 이러한 시간 영역대 결정을 다양화 하여 최적의 시간대를 알아내고자 한다.

[그림 6]에서 제안한 알고리즘을 보여주는데 TO와 마찬가지로 기본적으로 순수 P2P 모델을 기반으로 하며, P2P 오버레이 네트워크의 토폴로지는 대부분의 P2P 검색 논문에서 적용한 랜덤 그래프 모델을 바탕으로 하며, 질의와 자원의 분포는 Zipfian 이외의 실제 P2P 시스템에서 보여지는 분포와 유사한 다양한 분포 범칙을 적용한다. 기본 입력값, 인덱스, 자원 보유 노드의 수에 따른 네트워크 구조 설정, 노드의 평균 차수 등에 대한 내용은 TO를 따른다. Walker는 질의 발생 노드의 K 수만큼 질의가 발생되어 K 수에 따라 전체 메시지 발생량에 큰 영향을 줌으로 적절한 수의 K를 결정하는데 있어 주요한 역할을 한다. P2P 시스템의 검색 작업이 시작되면서 그 검색 시작 시간을 기준으로 [그림 4]의 기준을 정하게 된다. 이웃노드에 관한 연관

성 확률값 보유 인덱스인 Niindex의 값이 초기값 즉 null 값을 가지고 있는 경우 이웃 노드를 선택하는 기준은 오로지 온라인 상태에 대한 확률값이다.

```

The_Probabilistic_Proportion_Search() {
Initialization : pure P2P, random graph model
Definition : Qinode, Rjnode, i, K, Niindex
For (j = 0; j <= K; j++) {
  Check(current_time);
  If (Niindex = null) {
    For (i = 0; i <= neighbor; i++) {
      If ( Probabilistic.value[i] > Max ) {
        Max = Probabilistic.value[i];
      }
    }
    Neighbor_Select(Max);
  }
  else {
    Top_List(all_of_Niindex);
    Select_Max(Top_List.Probabilistic.value);
    Neighbor_Select;
  }
  If (Qinode = Rjnode) {
    Backward_Forwarding(update_index);
    For (All neighbors of the hit node) {
      TO_hit_ratio_update
      constantV+recentTOHitRatio(0.1)
    }
  }
  When Broadcast_Flooding(kwalker_success);
}

```

그림 6. TO 기반 확률적 비례 검색 알고리즘  
Fig. 6. Probabilistic Proportion Search Algorithm based on TO

해당 이웃노드들의 [그림 4] 사용시간 분석 결과값들을 비교하여 가장 큰 값을 보유한 피어로 검색이 전이된다. 이와 같은 초기화 작업 후 검색의 과정을 진행하는 Walker를 기반으로 실제적인 검색 과정이 시작되는데, 질의의 분석 과정에서 지역 인덱스의 연관성 확률 값을 바탕으로 우위에 있는 이웃피어를 선택하여 질의가 전파되는데, 먼저 지역 인덱스 값을 지닌 피어들 가운데 결정값에 의해 상위 영역 (Top\_List)을 결정하고 다음 기준으로 사용 시간 분석 값의 최고치를 보유한 피어를 다음 피어로 결정한다. 이러한 과정에서 원하는 자원을 발견하였을 시 질의 진행 방향과 역방향으로 연관성 확률 값을 갱신하면서 질의 발생 피어에 도달하게 된다. 즉 질의를 발생한 노드 Qjnode와 부합하는 자원을 지닌 노드 Rjnode을 찾은 경우 역방향으로 연관성 확률 값을 갱신하면서 질의 발생 피어에 도달하게 되는 것이다. 이러한 과정은 TO와 동일하다.

## IV. 평가

이 장에서는 참고문헌[4]에서와 같이 확장한 PeerSim P2P 시뮬레이터를 이용한 일련의 실험들을 통해 제안한 검색 알고리즘의 성능을 평가한다. 제안한 검색 알고리즘의 성공적인 검색 연산을 확인하기 위해 관련 알고리즘 분석을 통해 조사된 Random Walks, TO와의 성능 비교를 통해 제안한 검색 알고리즘의 P2P 약결합 환경에서의 적합성을 검증한다. 본 논문에서는 Window Server 2000 운영체제 하에 J2SDK 1.4.2의 개발 환경을 기반으로 소프트웨어 플랫폼을 구성했고, 시스템 사양은 Intel Pentium III 871MHz, 메인 메모리 256 Mb, 하드디스크 40 Gb 환경에서 실험하였다. 구성된 실험 환경 하에서 실험 대상의 기본 인수들을 설정해야 한다. 시뮬레이션 인수 가운데 P2P 모델은 순수한 P2P 모델, 그래프 모델은 랜덤 그래프 모델, 노드 수는 10000개의 노드, 평균 노드의 차수(degree)는 20, K Walker의 수는 15, TTL은 6, 객체 수는 100, 질의 분포는 Zipf를 기본으로 하고 비교 대상으로 Power-Law 및 Watts-Strogatz 분포를 대상으로 하며, 질의 발생 피어의 수는 1000으로 한다. 그래프 모델로써 랜덤 그래프를 실험 인수로 이용하였는데, 다수의 P2P 관련 검색 알고리즘의 네트워크 토폴로지로서 랜덤 그래프 모델을 선정하여 실험 환경에서 피어 간의 연결 관계를 맺고 있다. 또 질의 분포는 실제 P2P 시스템에서의 분포와 가장 유사하다고 알려져 있는 Zipf 분포를 바탕으로 주요 실험이 이루어졌다. 또한 [그림 4]의 사용 시간 분석 과정에서 나타날 수 있는 유형을 [그림 7]과 같이 패턴화 하였다. 본 알고리즘에 대한 평가를 하는데 있어서 실제 상황이 아닌 시뮬레이션 방법을 선택하였으므로 실험 환경에 맞는 사용 시간 분석 자료가 요구되었고, 이에 [그림 7]과 같은 유형으로 일반화 한 것이다. Y축은 하루 24시간을 나열한 것이고, X축은 1년을 기준으로 계절의 시작인 봄을 알리는 3월부터 2월까지 나열한 것이다. 각 유형에 대해 주해를 달아 보면, 대학생형(A)는 학기중일때와 방학일때로 나누어 시작 시간을 달리 잡았고 전체 이용시간은 5시간으로 획일화 하였다. 올빼미형(B)의 경우는 시스템 시작시간이 오후 5시이나 끝나는 시간이 다음날 새벽 6시로 밤낮이 바뀐 유형이다. 계절형(C)는 우리나라의 6월부터 10월까지 따뜻한 달에는 이용시간이 급격히 늘어나 있는 경우이다. 마지막으로 변덕형(D)의 경우 특정한 요인이 없이 이용 시간 및 시간대가 들쭉날쭉한 유형이다.

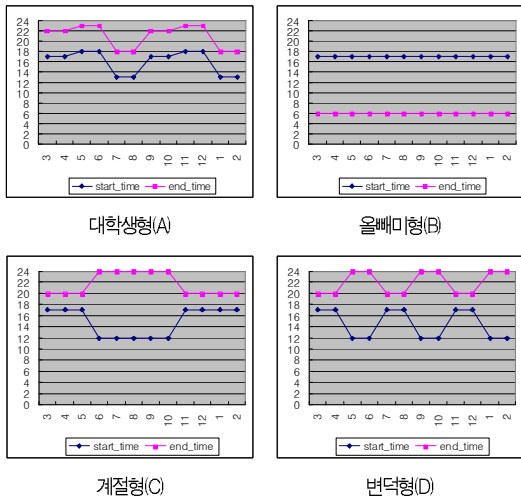


그림 7. 시뮬레이션 값의 패턴  
Fig. 7. Patterns of Simulation Value

표 1. 리플리케이션 기법에 따른 성공률  
Table 1. Success rate according to replication techniques

리플리케이션 기법		사용 시간 유형	평균 성공률 (K=1..15)		
			Random Walks	TO	PTO
확률적 리플리케이션	비례 리플리케이션	(A)	46.6%	78%	84.1%
		(B)			84.5%
		(C)			83%
		(D)			82.7%
	단순 리플리케이션	(A)	40%	72.1%	74.3%
		(B)			74%
		(C)			73%
		(D)			72.1%
자발적 리플리케이션	비례 리플리케이션	(A)	42.1%	73.4%	76%
		(B)			76.2%
		(C)			75.5%
		(D)			75%
	단순 리플리케이션	(A)	37%	65%	68.5%
		(B)			70%
		(C)			68%
		(D)			68.1%

본 실험에서 요청의 자발성과 리플리케이션 될 개수를 정하는 기준에 따라 분류된 리플리케이션의 기법들을 각각 대응시켜, Random Walks, TO, PTO(Probabilistic

TO) 각각과의 조화에 따른 성공률에 대해 <표 1>에 그 결과값을 표현했다. PTO는 현재 본 논문에서 제안한 알고리즘을 나타낸다. 표기된 성공률은 K가 1에서 15까지 각각의 값에 대한 성공률의 평균값을 나타내며, 리플리케이션 될 위치는 자원이 있는 피어에서 3홉 내로 한정하였으며, 자발적 리플리케이션 하에 비례 리플리케이션은 그 비례 정도를 2배로 한정하였다. 이와 같은 환경에서 확률적 리플리케이션 하의 비례 리플리케이션 기법에서 (B) 사용 유형의 조합된 PTO가 가장 좋은 성공률을 보였으며, 자발적 리플리케이션 하의 단순 리플리케이션 기법에 (C) 유형을 조합한 PTO가 해당 검색 기법 가운데 평균 성공률이 가장 낮았다. 그러나 자발적 리플리케이션 하의 단순 리플리케이션 기법의 조합에 의한 Random Walks가 Random walks, TO, PTO 전체 비교에서 가장 낮은 평균 성공률을 보이고 있다. Random walks, TO, PTO 전체 비교 과정에서 PTO의 우월함은 <표 1>의 결과가 말해준다.

### V. 결론 및 향후 연구

약결합 환경인 P2P 시스템에서 자원 제공 피어의 온라인 상태 유지에 대한 정보를 이용하여 검색 성공률을 높일 필요가 있다. 본 연구에서는 기존의 TO P2P 자원 검색 알고리즘 기반 하에 자원 전송률을 높이기 위한 PTO 검색 알고리즘을 제안한다. 본 연구에서는 해당 P2P 시스템을 이용하는 사용자의 평균 이용 시간대의 분석 자료를 자원 제공자 선택의 기준으로 적용하여 자원 전송 보장성을 높이고, 또한 인기도 높은 자원에 대해서 자료 전송 기회를 높여주는 역할의 기존의 분산 객체 리플리케이션 기법들과의 조합에 의한 분산 객체 전송률이 향상된 검색 알고리즘을 제안하였다. 제안한 검색 알고리즘의 성능 평가 결과 모든 유형에 대한 평균 결과값이 75.3%로 Random Walks 41.4%, TO 71.2%를 훨씬 상회했다. 그 중에서도 이와 같은 환경에서 확률적 리플리케이션 하의 비례 리플리케이션 기법에서 (B) 사용 유형의 조합된 PTO가 84.5%로 가장 좋은 성공률을 보였다.

향후 연구과제로 본 검색 알고리즘에서 적용한 시간 영역 결정의 기준에 대해 다양하게 적용하여, 최적의 시간대를 알아내고자 한다. 그리고 해당 P2P 시스템의 사용 시간을 결정하는데 있어 4가지 유형으로 정형화해 실험한 시뮬레이션 값의 패턴에 대해 좀더 실험의 근거를 명확히 할 수 있는 형태로 확장하여 실험할 필요가 있겠다.

### 참고문헌

- [1] 김분희, 이준연, “단계별 OK 기법 기반 효과적인 P2P 검색 알고리즘”, 한국컴퓨터정보학회 논문지, 10권 2호, 2005. 5.
- [2] 김분희, 이준연, “개선된 노드 분산율을 위한 적응적 P2P 검색 알고리즘”, 한국컴퓨터정보학회 논문지, 10권 4호, pp.93-102, 2005. 9.
- [3] Boon-Hee Kim, Young-Chan Kim, “Ptops Index Server for Advanced Search Performance of P2P System with a Simple Discovery Server,” LNCS 3032, Springer-Verlag, pp. 285-291, 2004.4.
- [4] 김분희, 김영찬, “P2P 오버레이 네트워크의 효과적인 평가를 위한 시뮬레이터 설계” 한국인터넷정보학회 2004 추계학술발표논문집 제5권 제 2호 pp.65-68, 2004.11.
- [5] K. Aberer, et al., “Improving Data Access in P2P Systems,” IEEE, vol. 5, no. 1, pp. 58-67, 2002.
- [6] P. Ganesan, et al., “YAPPERS: A Peer-to-Peer Lookup Service over Arbitrary Topology”, INFOCOM’03, pp.1250-1260, 2003.
- [7] D. Tsoumakos and N. Roussopoulos, “Analysis and Comparison of P2P Search Methods,” Technical Report(CT-TR-4451), University of Maryland, Dept. of Computer Science, 2003.
- [8] B. Yang and H. Garcia-Molina, “Improving search in peer-to-peer networks,” ICDCS’02, pp.103-113, 2002.
- [9] Boon-Hee Kim, Young-Chan Kim, “Automatically Generated Ontology System for Peer-to-Peer Networks”, Lecture Notes in Computer Science 3320, Springer-Verlag, pp. 493-496, 2004.12.
- [10] K. Sripanidkulchai and Bruce Maggs, “Efficient Content Location Using Internet-Based Locality in Peer-to-Peer Systems,” INFOCO M’03, pp.238-243, 2003.

### 저자 소개



김분희

2005년 2월 : 중앙대학교 컴퓨터공학과 공학박사  
 2005년~현재 : 동명대학교 멀티미디어공학과 전임강사

관심분야 :

