

자원 방향성 개선을 위한 적응적 P2P 검색 알고리즘

김분희*

*동명대학교 멀티미디어공학과

e-mail : m7515101@nate.com

An Adaptive Peer-to-Peer Search Algorithm to Reform Resource-Directionality

Boon-Hee Kim*

*Dept. of Multimedia Engineering, Tong-Myong University

요 약

기존의 인포드 검색 모델에서 자원 보유 피어의 검색 결과는 검색 과정에서 선택된 피어들 각각의 라우팅 결정 정보로 저장되는데, 자원 적중률에 따라 다음 피어를 선택할 때 주요 결정 요인으로 작용한다. 이러한 라우팅 결정 정보를 이용하여 검색 트래픽과 검색 성공률 등의 측면에서 기존의 Random Walks 등에 비해 효과적인 성능 평가 결과를 보인 LTO(Leveled The Orienteering) 검색 알고리즘은 검색 과정에서의 라우팅 경로가 편중되어 있어 해당 노드들에 검색 부하 집중 경향을 보인다. 이에 본 논문에서는 검색 적중 빈도에 따른 LTO 검색 알고리즘이 검색 과정상에서 라우팅 대상 노드들의 성능을 고려한 적응적 P2P 검색 알고리즘으로 확장 제안한다. 제안한 알고리즘의 성능 평가 결과 노드들의 검색 부하 집중률과 자원 검색 노드의 검색 성공률 측면에서 적절한 타협점에서 동작함을 확인하였다.

1. 서론

컴퓨터 및 통신 기술의 발전과 더불어 네트워크 기반 응용 기술이 컴퓨터 분야의 주요 관심 기술들 중의 하나로 주목받고 있다. 네트워크 기반의 응용 프로그램으로 웹이나 FTP, TELNET 등 제한적인 과거 통신 환경에서와는 달리 P2P(Peer-to-Peer) 응용 프로그램과 같은 네트워크 기반 응용프로그램들이 다양하게 분포되어 있다. 그러나 P2P 시스템의 연구와 활용이 다양화 되면서 피어들의 자원 검색과정 혹은 자원의 리플리케이션 과정에서 발생하는 트래픽의 네트워크 대역폭에 주는 과도한 영향은 더 이상 간과할 수 있는 일이 아니다. P2P 응용 프로그램이 이러한 트래픽 발생의 양을 줄이지 않으면, 다양한 네트워크 응용 프로그램들이 산재해 있는 인터넷 환경에서 원활한 대역폭 사용에 큰 영향을 줄 수 있으며, 많은 네트워크 이용자의 관련 응용프로그램 사용에 불편함을 초래할 수 있다. 또한 P2P 시스템은 피어들의 P2P 오버레이 네트워크에의 연결성이 매우 약한 약결합 시스템으로

검색 단계에서 원하는 자원을 찾는다 하여도 다운로드 단계에서 반드시 그 피어에서 자원을 다운 받을 수 있다는 보장이 없다. 이에 급증하는 P2P 시스템 이용과 더불어 트래픽 량 감소와 약결합 환경에 영향이 적은 시스템에 대한 요구가 증가되고 있다 [1][2][3][4].

앞선 LTO 논문에서는 P2P 시스템을 이용하는 피어의 연결과 종료의 시점이 모호한 약결합 환경을 기반으로 하는 P2P 시스템에서 효과적인 검색 성공률과 트래픽 양을 발생시키는 순수한 P2P 모델 기반의 브로드캐스팅 검색 기법에 대해 제안하였다. 제안한 검색 기법은 관련 시스템에 내재된 문제점을 개선하는 방향으로 설계 및 구현되었으며, 기존의 브로드캐스팅 기반 검색기법과 비교 실험을 통해 제안된 검색 기법의 성능을 평가하였다. 연구 수행 결과는 논문을 위한 선행 단계의 연구로써 P2P 약결합 시스템 기반 검색 기법의 다양한 방법론들을 비교 분석하여, P2P 시스템에 효과적으로 적용 가능한지 검토하였다. P2P 약결합 시스템 기반 검색 기법을 제공하는 기존 시스템을 분

석하여 문제점을 분석하고, 개선 방안을 제시하였다. 기존 연구에 대한 비교 분석 결과 및 도출된 개선 방안을 토대로 P2P 약결합 시스템에 효과적인 검색 기법인 TO(The Orienteering)를 설계하였다.

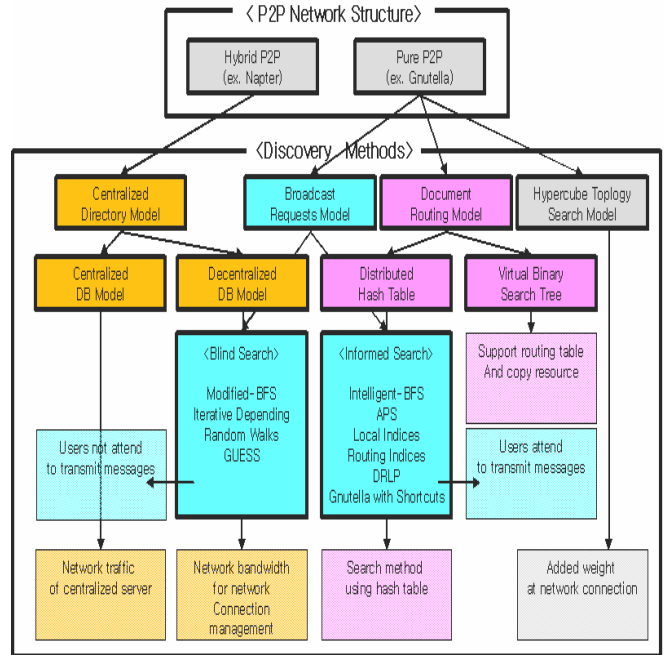
TO의 효과적인 검색 연산을 위해 Random Walks 기반 피어 선택 기법과 질의율에 따른 차별적 리플리케이션 기법을 채택하였다. 제한한 시스템의 성능을 평가하기 위해, 동일한 통신량과 네트워크 환경하에 Zipf, Power-Law, Watts-Strogatz 분포의 실험 모델을 바탕으로 Peersim 환경 하에서 구현한 뒤 성공률, 트래픽 발생률, 노드 분산률을 측정하였다. 그 결과, 대부분의 문제 크기에서 만족할 만한 수준의 오버헤드를 보여주었으며, 시뮬레이션 단계가 증가함에 따라 성공적인 질의 결과를 보여주었고, 질의율에 따른 차별적 리플리케이션 기법의 적용에 따른 유효성을 검증하였다. 또한 검색의 방향성과 통신 발생의 분포 측면에서 적절한 타협점에서 동작함을 확인하였다. 그러나 LTO 검색 알고리즘은 검색 과정에서의 라우팅 경로가 편중되어 있어 해당 노드들에 검색 부하 집중 경향을 보인다. 이에 본 논문에서는 검색 적중 빈도에 따른 LTO 검색 알고리즘이 검색 과정상에서 라우트 대상 노드들의 성능을 고려한 적응적 P2P 검색 알고리즘으로 확장 제안한다.

2. 관련연구

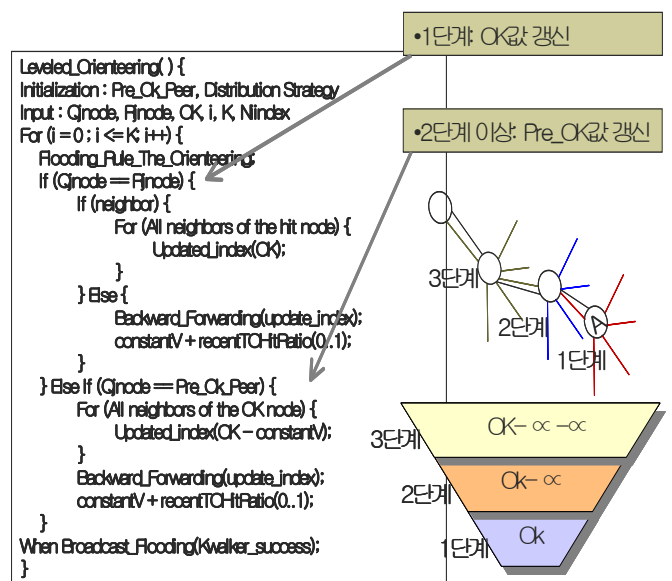
2.1 기존 검색 기법

P2P 시스템에서 자원의 검색 기법은 P2P 네트워크 구조에 따라 분류 할 수 있다. 먼저 P2P 네트워크 구조는 피어간의 네트워킹 방법과 역할에 따라 순수 P2P(Pure Peer-to-Peer), 간단한 조회 기능 서버를 가진 P2P(P2P with a simple discovery server), 조회 서버와 룩업 서버를 가진 P2P(P2P with discovery and lookup servers), 조회/룩업/컨텐츠 제공 기능의 서버를 가진 P2P (P2P with discovery, lookup and content servers) 모델로 분류할 수 있는데, 여기서 순수한 P2P 모델을 제외한 나머지 서버의 기능이 있는 모델을 일반적으로 하이브리드 P2P 로 묶어서 분류한다[9][10]. 하이브리드 P2P 구조에 속한 검색 기법은 중앙집중식 모델로써 대변되는데, 피어의 연결정보를 보유한 중앙 서버가 검색 대상 데이터의 보유 여부에 따라 중앙집중형 데이터베이스 모델과 분산형 데이터베이스 모델로 나뉜다. 순수한 P2P 모델의 경우는 검색 기법의 특징에 따라 브로드캐스트 요청 모델, 도큐먼트 라우팅 모델로 나뉜다. 실제 가장 많이 이용되는 모델인 브로드캐스트 요청 모델은 검색 메시지의 전달에 특정 정보의 이용 여부에 따라 인포드 검색[과 블라인드 검색 기법으로 나뉜다. 순수 P2P 컴퓨팅 환경을 기반으로 하는 검색 기법 가운데, Gnutella[6]는 자원 검색을 위해 사용하기 가장 쉬운 방법으로 불규칙적으로 연결된 네트워크 구조에 적용하기 용이하다. 그러나 검색 기법 가운데 가장 많은 네트워크 트래픽 량을 나타낸다. 변형된 BFS 기법은 Gnutella 알고리즘과 비교해 보면 검색 메시지에 의한 네트워크 트래픽 량이 적지만 인

접한 모든 피어에게 질의 메시지를 전송하는 것이 아니므로 적중률이 낮다. Random Walks 는 검색에 의해 발생하는 메시지의 수를 제한하여 전송하는 방법으로 변형된 BFS 알고리즘에 비해 메시지 발생량이 적다. 그러나 검색 적중률이 낮다. 이러한 검색 알고리즘의 특징은 검색 성공률과 트래픽 량 측면에서 효과적인 결과를 얻고자 연구의 방향이 진행되고 있음을 알 수 있다. 이와 더불어 P2P 환경의 기본 특징인 약결합 환경에 적응적인 기법이 보완되어야 한다.



(그림 1) 기존 P2P 검색 알고리즘



(그림 2) LTO 검색 알고리즘

2.2 LTO(Leveled The Orienteering)

인폼드 검색 방법의 하나인 LTO 는 검색 알고리즘의 근접성을 개선하기 위한 검색 모델로 확장하고자 한다. 검색 진행 메시지의 발생 및 진행의 과정에서 초기 이웃 피어의 선택과정은 K Walks 의 기본 원리를 적용하고 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 값을 반영한 값이다. 이러한 초기화 작업을 거쳐서 검색의 과정을 진행하는 Walker 를 기반으로 실제적인 검색 과정이 시작되며, 질의의 분석 과정에서 지역 인덱스의 연관성 확률 값을 바탕으로 우위에 있는 이웃피어를 선택하여 질의가 전파된다.

검색 주체 피어가 해당 자원이 있는 피어로의 근접성의 정도를 더욱 높이기 위한 정책으로써 단계별 OK 정책을 제안한다. OK 정책의 단계를 정함에 따라 자원에 대한 근접율을 기존의 방법에 비해 향상시키고자 한다. 검색의 과정이 시행되기 이전에 OK 값의 단계를 정하는 작업이 선행되며, 기존의 정책은 단계별 OK 정책을 인용하면 1 단계 OK 정책을 구현한 것이다. 2 단계 이상의 OK 정책을 선택함으로써 비로소 결론으로의 방향성을 거치게 되는 구조이다.

ALTO (K=20, outdegree=20)	Gnutella (outdegree=20)
20	20
20	380
20	7220
20	137180
20	2606420
20	49521980

(그림 3) 알고리즘 복잡도 분석

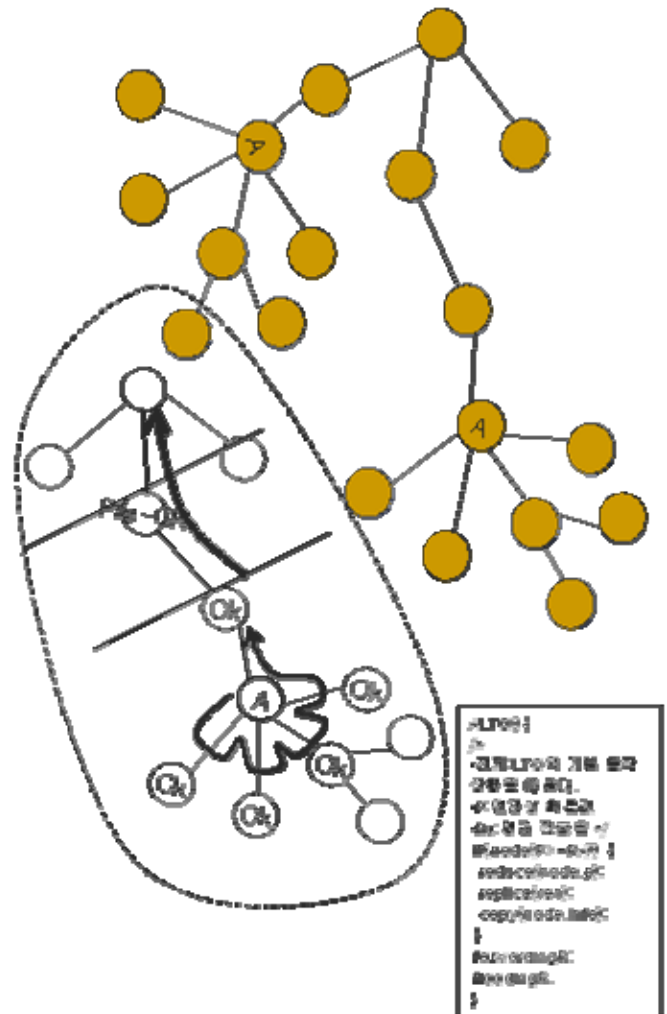
3. 적응적 P2P 검색 알고리즘

기존 LTO 검색 알고리즘이 검색 과정에서의 라우팅 경로가 편중되어 있어 해당 노드들에 검색 부하 집중 경향을 보였다. 이에 본 논문에서는 검색 적중 빈도에 따른 LTO 검색 알고리즘이 검색 과정상에서 라우트 대상 노드들의 성능을 고려한 적응적 P2P 검색 알고리즘을 제안하고자 한다. (그림 4)이 본 논문에서 제안한 적응적 P2P 검색 알고리즘인 ALTO 이다.

ALTO 는 기본적으로 LTO 의 검색 알고리즘의 기본 동작 상황을 따른다. 검색의 방향성을 결정하는 연관성 확률값을 기준으로 특정 노드의 평균 적중률을 대비시킨다. 이러한 비교 결과 한 노드의 연관성 확률값이 전체 노드의 평균 적중률 이상일 경우 해당 노드의 연관성 확률값을 적정 수준으로 줄이고, 해당 역할을 분담할 다른 노드를 선정한다. 해당 역할을 분담 받은 노드는 원래의 노드로부터 해당 검색 대상 자원을 다운받고, 해당 연관성 확률값도 동일하게 설정한다. 부모 노드로부터 자식노드가 모든 특징들을 상속받는 것과 같다.

본 알고리즘의 복잡도를 계산해 보면 (그림 3)과 같다. 문제의 크기를 나타내는 파라미터 n 을 검색의 과정에서 거치게 되는 노드의 수로 정하였다. 아래의 예는 ALTO, Gnutella 모두 동일한 상태에서 6 홉만큼을 모두 거쳤을 경우의 n 을 나타낸다.

n에 관해 가장 빨리 증가하는 항만을 남기고 다른 항은 무시해도 되는 점근적인 복잡도 평가에 의해 ALTO 알고리즘을 평가해 보면 $O(\log n)$ 이다. Gnutella의 브로드캐스팅 검색 기법은 $O(n^n)$ 으로 ALTO 알고리즘의 복잡도가 확연히 낮음을 알 수 있다.



(그림 4) ALTO-적응적 P2P 검색 알고리즘

4. 결론 및 향후 연구

P2P 환경에서 해당 자원을 검색하는데 있어, 기존의 인포드 검색 모델에서 자원 보유 피어의 검색 결과는 검색 과정에서 선택된 피어들 각각의 라우팅 결정 정보로 저장되는데, 자원 적중률에 따라 다음 피어를 선택할 때 주요 결정 요인으로 작용한다. 이러한 라우팅 결정 정보를 이용하여 검색 트래픽과 검색 성공률 등의 측면에서 기존의 Random Walks 등에 비해 효과적인 성능 평가 결과를 보인 LTO 검색 알고리즘은 검색 과정에서의 라우팅 경로가 편중되어 있어 해당 노드들에 검색 부하 집중 경향을 보인다. 이에 본 논문에서는 검색 적중 빈도에 따른 LTO 검색 알고리즘이 검색 과정상에서 라우트 대상 노드들의 성능을 고려한 적응적 P2P 검색 알고리즘으로 확장 제안하였다. 향후 P2P 검색 알고리즘을 근간으로 한 퍼베이션 환경으로의 확장을 피하며, 네트워크 환경의 변화가 검색 알고리즘의 미치는 영향을 연구하고자 한다.

참고문헌

- [1] D. Tsoumakos and N. Roussopolulos, "Analysis and Comparison of P2P Search Methods," Technical Report(CT-TR-4451), Maryland Univ., 2003.
- [2] KT, P2P Traffic Investigation, 2004.
http://news.naver.com/hotissue/daily_read.php?section_id=105&office_id=015&article_id=000071
- [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] D. S. Milojevic, et al., "Peer-to-Peer Computing," Technical Report (HPL-2002-57), Hewlett-Packard Company, 2002.
- [5] Napster website: <http://www.napster.com>.
- [6] S. Daswani and A. Fisk, Gnutella UDP extension for scalable searches (GUESS) v0.1
- [7] Gnutella website: <http://gnutella.wego.com>
- [8] P. Felber, E. Biersack, L. Garces-Erce, K.W. Ross, G. Urvoy-Keller, Data Indexing and Querying in P2P DHT Networks, ICDCS 2004, Tokyo, Japan.
- [9] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and replication in unstructured peer-to-peer networks," ACM (Supercomputing), June 2002.
- [10] Rüdiger Schollmeier, Gero Schollmeier, "Why Peer-to-Peer (P2P) Does Scale: An Analysis of P2P Traffic Patterns," P2P'02.