자원 역 스케쥴링 기반 P2P 검색 알고리즘

김분희*
*동명대학교 멀티미디어공학과
e-mail:bhkim@tu.ac.kr

The P2P Search Algorithm based on Reverse Scheduling about Resources

Boon-Hee Kim*
*Dept of Multimedia Engineering, TongMyong University

요 약

다양한 컴퓨터 시스템이 공존하는 가운데 소요 비용에 비해 해당 성능이 매우 우수한 분산 시스템에 대한 연구가 활발하다. P2P 기술은 이러한 분산 시스템에서 개개인이 보유한 자원을 효율적으로 공유하는 측면에서 활용도 높은 분야이다. 이러한 P2P는 공유된 자원을 보유한 자원 제공 피어의 온라인 상태 유지 여부에 따라 해당 자원에 대한 검색 적중에 기여하게 된다. 이렇게 얻어진 검색 결과를 이용하여 검색 참여 피어는 원하는 자원을 원하는 피어로부터 다운로드하게 된다. 그러나 종종 다운로드 완료 전에 자원 제공 피어의 오프라인 상태로의 변화를 목격하게 된다. 이 상황에서 원하는 자원을 온전히 다운로드 받기 위해서는 주로 재전송의 방법을 이용하게 된다. 이는 P2P 시스템의 성능 저하의 원인이 되므로 이에 대한 해결책이 필요하다. 본 연구에서는 자원 검색 피어가 해당 자원에 대한 다운로드 작업 중 재전송 요구가 발생하였을 때 해당 요구에 효과적으로 대응하기 위한 자원 역 스케쥴링 기법을 제안하여 P2P 시스템의 신뢰도 향상 효과를 얻고자 한다.

1. 서론

자원을 효율적으로 이용할 수 있는 P2P 시스템은 공유된 자원을 보유한 자원 제공 피어의 온라인 상태 유지 여부에 따라 해당 자원에 대한 검색 적중률에 결정적인 기여하게 된다. 이에 P2P 관련 연구 가운데 불완전한 네트워킹 환경을 극복 할 수 있는 분야에 대한 연구가 이슈화 될 수 밖에 없다. 실제로 해당 P2P 환경에서 자원을 찾고 있는 피어는 우선적으로 검색의 과정을 거치게 되고, 이렇게 얻어진 검색 결과를 이용하여 검색 참여 피어는 원하는 자원을 원하는 피어로부터 다운로드하게 된다. 이때 P2P 시스템에서 늘 발생할 수 있는 경우인 다운로드 완료 전 자원 제공 피어의 오프라인 상태로의 변화를 종종 목격하게 된다. 이 상황에서 원하는 자원을 온전히 다운로드 받기 위해서는 주로 재전송의 방법을 이용하게 된다. 그러나 단순히 재전송하는 방법을 이용하게 되면 기본적으로 자원을 검색하는 단계부터 다시 시작되어야 한다. 이는 P2P 시스템의 성능 저하의 원인이 되므로 이에 대한 해결책이 필요하다. 실제 검색 단계에서의 트래픽을 줄이기 위한 연구 또한 활성화 되고 있는 상황에서 이는 상당한 성능 저하 원인이 된다[1][2][3].

본 연구에서는 자원 검색 피어가 해당 자원에 대한 다운로드 작업 중 재전송 요구가 발생하였을 때 해당 요구에 효과적으로 대응하기 위한 알고리즘을 제안한다. 이를 위해 해당 자원에 대한 검색과정에서 자원 역 스케쥴링 기법을 제안하여 재전송 요구시 대응성을 높이고, 이로써 해당 P2P 시스템의 신뢰도 향상에 기여하고자 한다. 본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 3장에서는 본 논문의 핵심 부분인 제안한 검색 알고리즘을 마지막으로 4장에서 결론을 맺는다.

2. 관련연구

P2P 시스템에서 자원의 검색 기법은 그림 1에서와 같이 P2P 네트워크 구조에 따라 분류를 할 수 있다[6]. 먼저 P2P 네트워크 구조는 피어 간의 네트워킹 방법과 역할에 따라 순수한 P2P, 간단한 조회 기능 서버를 가진 P2P, 조회 서버와 룩업 서버를 가진 P2P, 조회/룩업/컨텐츠 제공 기능의 서버를 가진 P2P 모델로 분류할 수 있는데, 여기서 순수한 P2P 모델을 제외한 나머지 서버의 기능이 있는 모델을 일반적으로 하이브리드 P2P로 묶어서 분류한다. 하이브리드 P2P 구조에 속한 검색 기법은 중앙집중식 모델 (Centralized Directory Model)로써 대변되는데, 피어의 연결정보를 보유한 중앙 서버가 검색 대상 데이터의 보유 여부에 따라 중앙집중형 데이터베이스 모델(Centralized DB Model)과 분산형 데이터베이스 모델(Decentralized DB Model)로 나뉜다. 순수한 P2P 모델의 경우는 검색 기법의 특징에 따라 브로드캐스트 요청 모델(Broadcast Requests Model), 도큐먼트 라우팅 모델(Document Routing Model), 하이퍼큐브 토플로지 검색 모델(Hypercube Topology Search Model)로 나뉜다[4][5].

하이브리드 P2P	중앙 집중형 디렉토리 모델	중앙 집중형 데이터베이스 모델
		분산형 데이터베이스 모델
순수 P2P	다큐먼트 라우팅 모델	
	브로드캐스트 요청 모델	인폼드 검색
		블라인드 검색

(그림 1) P2P 검색 알고리즘 분류

3. 제안한 알고리즘

본 논문에서는 자원 재전송에 따른 성능 저하 원인을 예방하여 기존의 TO 자원 검색 알고리즘[6] 기반 동료피어 동반 검색 알고리즘을 기반으로 하는 자원 역 스케쥴링 검색 알고리즘을 제안한다. 본 알고리즘의 동료피어 선정 기준은 이전의 동료피어 동반 검색 알고리즘과 같이 다음의 식을 기반으로 이루어진다.

$$S = Ran(Size(Grid) >= \left(\frac{\sum_{i=1}^{n} Hi(hit)}{n}\right)) \quad \text{--식 1)}$$

$$S = Ran(\left(\frac{\sum_{i=1}^{n} Hi(hit)}{n}\right) > H(1)) \quad \text{--식 2)}$$

식 1)에서 Size(Grid)는 실험 공간의 전체 네트워크 사이즈이다. H는 현재 피어로부터 떨어진 거리를 나타내는 홉 수이다. Hi(hit)는 해당 피어가 이전에 원하는 피어를 찾았던 기록이다. 따라서 이전 기록을 기준으로 히트가 일어난 홉수를 합산하여 전체 횟수 n 만큼 나누어 평균값을 구한다. 식 2)는 임의의 피어를 선정하는데 있어서 최대 홉 수를 이전 히트 기록의 평균수로 하고, 최저 홉 수를 바로 이웃 피어로 하여 검색 피어와의 평균값 보다 가까운 거리의 동료 피어를 선정하는 방법이다.

이러한 동료 피어 동반 검색 알고리즘 기반하에 해당 자원 검색 피어가 다운로드 중에 재전송 요구가 발생할 경우 이 요구에 즉시 대응하기 위한 자원 보유 피어 리스트를 전역 리스트 NRlist에 보관하게 된다. 식3)은 해당 자원 보유 피어 리스트에 해당 자원 보유 피어의 리스트를 우선 순위화 하여 리스트를 만들게 되는 방법이다. TT(Turnaround Time)은 네트워크로 연결된 P2P 환경에서의 해당 자원을 찾음에 대한 결과로써 해당 시스템 내의 응답 시간에 회선 처리 시간과 단말기 처리시간이 합쳐진 응답시간이다. HR(Hit Ratio)은 0과 1사이의 값으로 큰 값이 적중률이 상대적으로 높다. TT와 HR 값을 이용하여 해당 자원 보유 피어의 신뢰도를 측정하게 되고 t을 온라인 평균 유지 시간 대에 따른 가산치 적용에 이용되며, Rank는 이들을 이용하여 해당 자원 보유 피어에 정규화 과정을 거친 우선순위를 부여하여 재전송 요구에 즉시 대응할 후보 리스트의 결정 자료로써 적용되는 것이다.

$$NRlist[i] = Rank(t + R(\frac{TT}{HR})) \quad \text{--식 3)}$$
$$, 0 \leqq HR \leqq 1$$
$$, t(j \leqq D \leqq k)$$

식3)에서 t는 해당 접속 시간대에서의 시스템 접속 시간과 끝의 시간대인 k-j 간격의 너비 정도(D)에 따라 가중치를 부여하여 우선순위 결정의 중요 요소로 적용한다. 이러한 우선순위 기반 재전송 후보 리스트의 선정은 해당 우선순위를 정하는 기준의 적용이라는 복잡한 과정을 거쳐서 선정되므로 계산 시간에서의 추가 비용이 예상되지만 재전송 발생 이전에 이루어지는 작업이므로 재전송 시기의 시간 복잡도에 영향을 미치지 않아 유효하다. 이렇듯 NRlist를 이용하여 재전송 요구에 대응하는 것은 이전 검색 결과값들을 역 추적하여 현재의 우선순위 평가자료로 이용하게 되고, 다양한 결과노드들 간의 효과적인 순위부여 스케쥴링이 적용되는 것이다.

3. 결론

다양한 컴퓨터 시스템이 공존하는 가운데 소요 비용에 비해 해당 성능이 매우 우수한 분산 시스템에 대한 연구가 활발하다. 이러한 분산 시스템 가운데 P2P 기술은 개개인이 보유한 자원을 효율적으로 공유하는 측면에서 활용도 높은 분야이다. 이러한 P2P 시스템에서 공유된 자원을 보유한 자원 제공 피어의 온라인 상태 유지 여부에 따라 해당 자원에 대한 검색 적중에 기여하게 된다. 이렇게 얻어진 검색 결과에서 다운로드 도중에 자원 제공 피어의 오프라인 상태로의 변화를 목격하게 된다. 이 상황에서 원하는 자원을 온전히 다운로드 받기 위해서는 주로 재전송의 방법은 P2P 시스템의 성능 저하의 원인이 되므로 이에 대한 해결책이 필요하다. 본 연구에서는 자원 검색 피어가 해당 자원에 대한 다운로드 작업 중 재전송 요구가 발생하였을 때 해당 요구에 효과적으로 대응하기 위한 자원 역 스케쥴링 기법을 설계하였다.

참고문헌

[1] K. Aberer and Z. Despotovic, "Managing Trust in a Peer-2-Peer Information System," Proc. 10th Int'l Conf. Information and Knowledge Management (2001 ACM CIKM), ACM Press, New York, pp.310-317, 2001.

[2] D. Tsoumakos and N. Roussopolulos, "Analysis and Comparison of P2P Search Methods," University of Maryland, Dept. of Computer Science, 2003.

[3] Boon-Hee Kim, Young-Chan Kim, "Automatically Generated Ontology System for Peer-to-Peer Networks", LNCS3320, Springer-Verlag, pp. 493-496, 2004.12.

[4] D. Tsoumakos and N. Roussopoulos, "A Comparison of Peer-to-Peer Search Methods," WebDB, pp.61-66, 2003.

[5] B. Yang and H. Garcia-Molina, "Improving search in peer-to-peer networks," ICDCS'02, pp.103-113, 2002.

[6] 김분희, "분산 객체의 확률적 비례 검색 기반 전송률 향상 검색 알고리즘", 한국컴퓨터정보학회 논문지, pp.49-56, 2006. 7.