

# Ultrapeer를 사용한 P2P 네트워크에서 동일한 관심을 갖는 Peer들의 그룹화 효과

김기, 옹환기

한동대학교 정보통신공학과

bart@seed.handong.edu, wyong@handong.edu

## The Effect of Grouping by Similar Interested Peers in P2P Network using Ultrapeer

Ki Kim, Whanki Yong

Department of IT, Handong Global University

### 요 약

P2P 네트워크는 사용자에게 보다 효율적으로 많은 자원을 공유하고 사용할 수 있는 방법을 제공해야 한다. 이 논문에서는 ultrapeer와 동일한 관심을 갖는 peer들을 서로 그룹화 함으로써 검색에 필요한 query의 수를 줄이는 방법을 제안한다. 이 방법을 검증하기 위해 세가지 P2P 네트워크 - 비구조적이며 브로드캐스팅으로 검색하는 네트워크, ultrapeer가 존재하며 지역적 특성을 그룹화 기준으로 하는 네트워크, ultrapeer가 존재하며 동일한 관심을 그룹의 기준으로 하는 네트워크로 모델링 하고 각 모델별로 필요한 데이터의 검색과 전송을 시뮬레이션하여 검색 시간과 발생한 query의 수를 비교하여 제시한다.

### 1. 서론

인터넷의 보급이 널리 확대됨에 따라 사용자에게 자유롭게 보다 많은 자원을 공유하고 공유된 자원을 보다 많이 사용하려는 욕구가 늘어나고 있다. 이러한 방법을 제공하기 위해 P2P 네트워크 모델이 기존의 서버/클라이언트 모델을 대신하게 되었다. P2P(peer-to-peer) 네트워크 환경이란 “각 peer들이 가지고 있는 자원 또는 제공하는 서비스를 peer들 상호간의 직접적인 연결을 통하여 교환할 수 있도록 지원하는 네트워크 환경”이라 정의 할 수 있다[1, 2].

P2P 네트워크에서는 분산된 peer들이 자신이 가진 CPU, 디스크, 네트워크, 파일 등의 다양한 자원들을 공유하고, 이러한 자원들을 사용하기 원하는 peer들은 자원을 제공하는 peer들에게 직접 연결하여 서비스를 제공받는다. 따라서 서비스에 필요한 부하를 다수의 peer로 분산시킬 수 있는 장점을 가지며, 상대적으로 가벼운 시스템들을 사용하여 다수의 서비스 요청을 만족시킬 수 있다는 장점이 있다. 그러나 자원의 소유와 관리가 각 peer들에 의해 이루어지며 peer들 사이에 공통적인 플랫폼이 미비하며, peer들의 출입이 잦은 네트워크의 특성을 해결해야 한다는 단점이 있다. 또한 넓은 인터넷 공간에 흩어져 있는 자원들 가운데서 필요한 자원이나 서비스를 찾아내는 방법에 대한 연구가 필요하다.

본 논문에서는 P2P 네트워크 시스템에서 필요한 자원을 효과적으로 검색하기 위한 방법으로 ultrapeer를 사용하고 동일한 관심을 갖는 peer들을 서로 그룹화하여 검색 시 발생하는 query의 수를 줄이는 방법을 제안한다. 이를 위해

P2P 네트워크 환경을 모델링하고 세가지 네트워크 환경하에서 공유자원인 파일의 검색과 전송을 시뮬레이션하고 그 결과를 비교하여 제시할 것이다.

2장에서는 기존에 이루어졌던 P2P 네트워크와 검색 방법에 대한 연구들과 개선방향으로써 제안하는 그룹화 방법을 설명하고, 3장에서는 기존의 방법들과의 비교를 위해 네트워크를 모델링하여 실험하는 방법에 대해 설명한다. 4장에서는 그 실험결과를 보이고 5장에서는 이 논문의 결론을 내리면서 글을 맺는다.

### 2. 기존 연구 및 개선방향

이번 절에서는 기존에 연구되었던 P2P 네트워크의 구조와 검색 기법들과 개선방향 대해서 알아본다.

P2P 네트워크의 구조는 혼합형 구조와 순수형 구조로 나눌 수 있으며, 검색 기법은 중앙집중식 디렉토리 모델과 다큐먼트 라우팅 모델, 브로드캐스트 요청 모델 등이 있다. 이들 중 순수형 네트워크를 가지며 브로드캐스팅 요청 모델을 사용하는 Gnutella 경우에 많은 query가 발생하는 단점을 가지며 모든 peer가 자원을 제공하는 측면에서 동등한 역할을 하고 있지 않다[5]. 이를 해결하기 위해 중앙집중식 방식을 절충하여 주로 파일을 제공하면서 다수의 연결을 감당할 수 있는 peer를 ultrapeer로, 파일을 주로 제공받는 peer를 일반 peer로써 분리하여 사용하는 방법이 대두되었다[3,4].

이러한 ultrapeer를 이용하는 방법에서의 문제점은 ultrapeer의 자격에 대한 기준이 있지만 그룹이 어떠한

peer들로서 구성되어야 한다는 기준이 모호하다는 점이다. 본 논문에서는 그룹의 구성방법에 따라 검색 시에 발생하는 네트워크 트래픽 양과 검색 시간에 차이가 있을 것이라 예상하여, 그룹을 peer가 가지고 있는 content, 혹은 찾고자 하는 content의 성향이 비슷한 것들끼리 묶어주는 방법을 제안한다. 이것의 효과로써 검색 시간과 발생하는 네트워크 트래픽이 감소될 것으로 예상된다. 이는 P2P 네트워크를 구성하는 peer는 자신이 얻고자 하는 자원과 자신이 공유하는 자원에 대해 편향적인 성향을 가질 것이라는 가정에서 추론한 것이다. 이러한 그룹을 만들기 위해서 peer가 속하는 그룹은 P2P 네트워크에 접속할 때 자신이 공유하는 데이터에 따라 1차적으로 결정되며, 2차적으로 운영 중에 얻은 자원의 성향을 파악하여 동적으로 이에 맞는 그룹으로 이동하도록 하는 방법이 필요하다. 이렇게 같은 관심을 갖는 peer들을 그룹으로 묶었을 때, 많은 비율의 검색이 그룹내에서 이루어지고, 다른 그룹으로의 검색도 무작위적으로 이루어지는 것이 아니라 특정 관심을 가지고 있는 그룹부터 검색함으로써 query의 발생을 줄일 수 있을 것이다.

3. 실험방법

실험은 P2P 네트워크 3가지를 모델링하여 Gnutella와 같이 비구조적 네트워크 구조를 가지면서 브로드캐스팅을 검색 방법으로 사용하는 P2P 네트워크, ultrapeer가 존재하며 지역적 특성을 그룹 기준으로 하는 네트워크, ultrapeer가 존재하며 동일한 관심을 그룹 기준으로 하는 세가지 네트워크를 시뮬레이션하여 비교할 것이다. 비교 방법은 시간에 따라서 원하는 데이터를 검색하는데 발생한 query의 개수를 측정함으로써 검색의 효율성과 네트워크에 미치는 부하를 파악하고자 한다.

카테고리	평균크기	표준편차	비율
Document	500 KB	200 KB	10%
Music	4,500 KB	300 KB	30%
Movie	700,000 KB	10,000 KB	15%
Game	650,000 KB	50,000 KB	10%
Application	50,000 KB	20,000 KB	15%
Image	700 KB	300 KB	10%
Book	1,200 KB	300 KB	5%
Cartoon	30,000 KB	4,000 KB	5%

<표 1>공유 파일의 카테고리 및 크기 및 비율

실험의 정책은 다음과 같다. peer가 공유하고 사용하는 자원은 파일로 설정하였으며, 그 종류는 <표1>에서와 같다. 또한 전체에서 해당 카테고리가 차지하는 비율을 조사하여 실험에서 사용할 데이터의 생성에 반영하였다. 파일의 카테고리는 peer들이 공유하는 자원의 성향을 나타내기 위하여 사용되며 peer의 성향은 자신이 가지고 있는 데이터 중에 가장 많은 비율을 차지하는 카테고리가 된다. 데이터 전송 속도는 데이터를 주고받는 양 peer가 가지고 있는 사용가능한 네트워크 대역폭과 양 peer간의 거리에 의해서 결정된다.

그룹을 구성하는 peer 개수의 최대 한도는 없으며 peer의 관심 변화로 그룹 사이에 이동이 일어날 경우 이동 시간은 고려치 않는다. 하나의 peer에 동시에 접속하여 서비스를 제공하고 받을 수 있는 개수를 제한할 수 있다.

설정은 공통적으로 총 160개의 peer가 8가지의 카테고리로 나뉘는 10000개의 데이터를 나누어 가지게 하였으며, 거리 당 속도의 감쇄는 3byte, 동시 upload / download 개수는 각각 30으로 제한하였다. 각 P2P 네트워크에 대한 실험 방법은 다음과 같다.

비구조적이고 브로드캐스팅으로 검색하는 네트워크의 시뮬레이션: Gnutella와 같이 검색을 위해 peer가 가지고 있는 neighbor list를 활용하여 쿼리를 브로드캐스팅하여 검색한다. 전파된 쿼리는 TTL에 따라서 전파될 수 있는 홉의 개수가 결정된다. 검색에 성공한 데이터는 peer간에 커넥터를 직접 연결하여 전송된다.

Ultrapeer가 존재하며 지역적 특성을 그룹의 기준으로 하는 네트워크의 시뮬레이션: 근접한 지역에 따라 그룹이 구성되며, peer들이 소유하고 있는 데이터의 성향과 상관없이 없다. 따라서 시간이 지나도 위치의 변동이 없으므로 피어들의 그룹 이동 또한 없다. query는 자신의 그룹을 담당하는 ultrapeer에게 전송하며, query를 받은 ultrapeer는 1차적으로는 자신의 그룹에서 데이터를 찾고 없으면 다른 그룹의 ultrapeer에게 차례로 query를 던진다. 성공적으로 검색된 데이터는 peer간의 직접적인 연결을 통해 전송된다.

Ultrapeer가 존재하며 동일한 관심을 그룹의 기준으로 하는 네트워크의 시뮬레이션: peer들이 가진 데이터의 성향에 따라 그룹으로 묶여진다. 시간이 지나면서 peer들이 가진 데이터의 성향이 바뀌게 되면 해당 성향을 가진 그룹으로 이동하게 되므로 그룹들 사이에 peer의 이동이 생기게 된다. 검색을 위한 query는 자신의 그룹을 담당하는 ultrapeer에게 전송하며, query를 받은 ultrapeer는 1차적으로는 자신의 그룹에서 데이터를 찾고, 없으면 그 데이터가 가지고 있는 성향을 주된 관심으로 하는 ultrapeer에게 먼저 query를 전달한다. 이 경우에도 없을 경우에 나머지 ultrapeer들에게 차례로 query를 던진다. 성공적으로 검색된 데이터는 피어간의 직접적인 연결을 통해 전송된다.

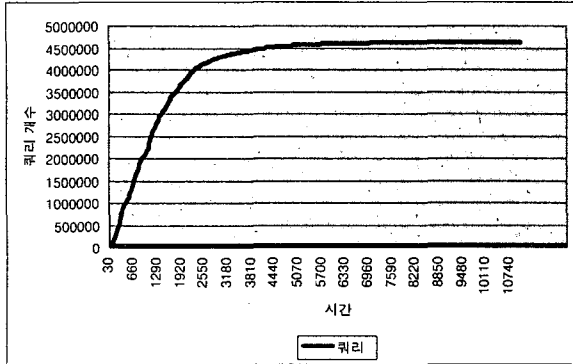
4. 실험결과

세가지 네트워크를 시뮬레이션 한 결과를 <표2>와 <그림 1-3>으로 정리하였다.

네트워크	찾는 파일개수	전송 시간(초)	발생한 query 수
비구조적 브로드캐스팅	478820	11014	4638674
Ultrapeer 지역적 특성		12218	478820
Ultrapeer 동일한 관심		11039	354938

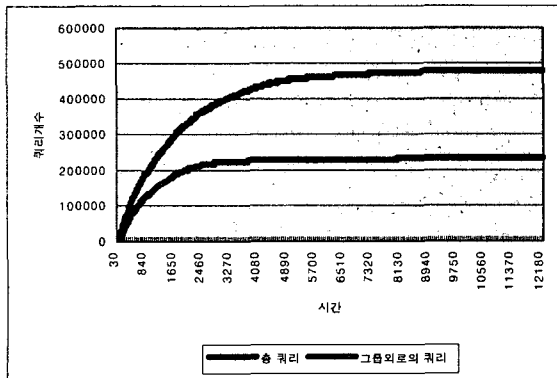
<표1> 세가지 네트워크의 실험 결과

비구조적인 네트워크 구조를 가지고 있으며 브로드캐스팅을 검색방법으로 사용하는 네트워크에서, 각 peer들이 가지고 있는 이웃 peer의 개수는 8개 이며, TTL은 3으로 설정하였다. 시간당 query 발생은 <그림1>에서와 같으며 데이터당 약 20개의 query가 발생하였다.



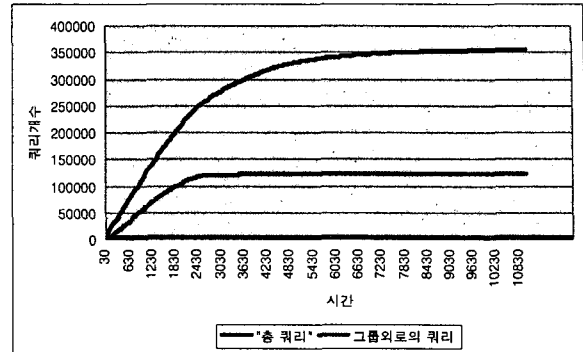
<그림 1> 비구조적이고 브로드캐스팅으로 검색하는 네트워크의 query 발생 수

Ultrappeer가 존재하며 지역적 특성을 그룹을 구성하는 기준으로 하는 네트워크의 실험결과는 <그림2>에서와 같으며, 데이터 당 약 2개의 query가 발생하였고 그룹 바깥으로 전달된 query의 개수가 약 50%를 차지한다. query 발생 수가 위의 실험보다 10%의 수준으로 감소한 것을 볼 수 있다.



<그림 2> 지역적 특성을 그룹 기준으로 하는 네트워크의 query 발생 수

Ultrappeer가 존재하며 동일한 관심을 그룹을 구성하는 기준으로 하는 네트워크의 실험결과는 <그림3>과 같고, 데이터 당 약 1.52개의 query가 발생하였다. 두번째 실험결과보다 75% 수준으로 감소한 이유는 <그림3>과 같이 그룹바깥으로 전달된 query의 개수가 약 1/3을 차지하고 있는 것에서 볼 수 있는 것처럼 대부분의 query가 동일한 관심을 가지는 그룹 안에서 성공적으로 이루어졌기 때문이라고 분석된다.



<그림 3> 동일한 관심을 그룹 기준으로 하는 네트워크의 query 발생 수

위의 결과를 종합하면, ultrapeer 가지고 동일한 관심을 그룹의 기준으로 하는 네트워크가 데이터를 검색하기 위해 발생하는 query의 수를 줄일 수 있는 방법이라는 것을 알 수 있다.

5. 결론

인터넷의 사용자가 자원의 제공자이자 소비자로서의 역할을 하는 P2P 네트워크에서 필요한 자원을 보다 효율적으로 검색할 수 있는 방법에 대한 연구가 필요하였다. 본 논문에서는 네트워크를 이루는 peer들이 일반적으로 편향적인 관심을 가지고 있을 것이라는 가정하에서, ultrapeer를 사용하여 동일한 관심을 갖는 peer들을 서로 그룹화하여 원하는 자원을 검색하기 위해 발생하는 query의 수를 줄이는 방법을 제안하였다. 이를 위해 세가지 P2P 네트워크 환경을 모델링하고 공유자원인 파일의 검색과 이동을 시뮬레이션한 결과를 비교하여 제시하였다. 실험결과 동일한 관심을 그룹 구성의 기준으로 하는 시스템에서 전체 찾고자하는 데이터 개수의 1.52배의 query를 발생하는 결과를 얻을 수 있었다. 향후과제로는 P2P 네트워크에서 공유되는 파일과 같은 자원에서 어떻게 성향을 효과적으로 추출할 수 있을 것인가에 대한 연구가 필요하다.

참고 문헌

- [1] Andy Oram, "Peer-To-Peer", O'Reilly, 2001
- [2] D.S. Milojicic, et. al., "Peer-to-Peer Computing," HP Technical Report, HP Laboratories, Mar, 2002.
- [3] B. Yang, H. Garcia-Molina, "Designing a Super-Peer Network," Proc. of the IEEE International Conference on Data Engineering, Mar, 2003.
- [4] D. Tsoumakos, N. Roussopoulos, "Analysis and Comparison of P2P Search Methods." Technical Report, University of Maryland, Dept. of Computer Science, Nov. 2003.
- [5] 권란, 이경근, "Unstructured Peer-to-Peer 네트워크에서의 파일공유", 정보과학회지 제22권 제3호, 2004