

온라인 커뮤니티 환경에서 P2P를 이용한 대용량 파일 공유 방법에 관한 연구

최지연*, 정강용**, 김원중*

*순천대학교 컴퓨터학과, **순천제일대학 컴퓨터학과
e-mail:kwj@sunchon.ac.kr

The Big-size File Sharing Method using P2P for Online Community Environment

Ji-Yeon Choi*, Gang-Yong Jung**, Won-Jung Kim*

*Dept of Computer Science, Sunchon National University

**Dept of Computer Science, Suncheon First College

요 약

인터넷을 기반으로 하는 동호회의 모임은 매우 활성화되어 있으며, 대부분의 포털 사이트에서도 '카페' 등과 같은 서비스명으로 동호회 모임을 지원하고 있다. 그러나 대부분의 동호회에서는 대규모 파일을 공유하기 위해서는 대규모 데이터 저장 공간의 부족 및 협소한 네트워크 등의 이유로 대규모 파일을 공유하는데 제약이 많다. 이에 본 논문에서는 인터넷 동호회 사이트에서 대규모 파일을 효과적으로 공유할 수 있도록 지원하기 위한 MAFI 시스템을 설계, 구현하였다.

1. 서론

인터넷이 대중화 되기 이전의 1990년대에는 동호회등과 같은 모임은 대부분 오프라인을 중심으로 한 모임이 대부분이었다. 그러나 1990년대 후반부터는 인터넷이 대중화되면서 오프라인 중심의 동호회 모임은 인터넷을 중심으로 하는 동호회 모임으로 변경되었다. 특히 대규모 포털 사이트에서는 인터넷 동호회의 활성화가 포털 사이트의 발전에 도움이 된다고 판단되자 게시판, 대화방, 자료실 등의 다양한 기능들을 제공하여 온라인 동호회를 지원하고 있다. 특히 자료실의 경우 대규모 용량의 다양한 자료를 공유할 수 있도록 용량이 증가되고 있다. 동호회 내부에서 대용량의 공유 자원이 증가되는 경우, 포털 사이트는 해당 자료를 동호회 회원들이 저장하고 다운로드 받을 수 있도록 지원하기 위하여 인터넷 디스크 용량의 증설 및 네트워크 대역폭 확보 등에 대한 지속적인 투자가 이루어져야 한다. 또한 사용자는 해당 자료를 다운로드 받기 위해서는 해당 사

이트를 통해서만 가능하므로 업로드 시간 및 다운로드 시간만큼의 지연 시간이 발생하는 문제점이 있다. 이에 본 논문에서는 온라인 커뮤니티 사이트 등에서 대용량의 자료를 공유할 수 있도록 하기 위하여 P2P를 이용한 공유 방법을 제안한다.

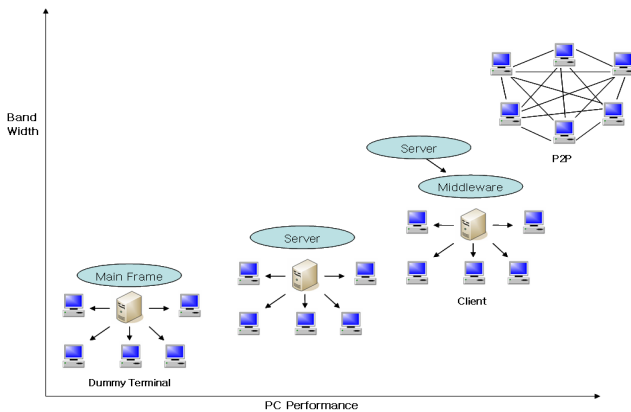
2. P2P

P2P는 최근 각광받는 인터넷 비즈니스 모델로, Peer to Peer의 약어다. P2P는 인터넷 혹은 네트워크상에서 각 컴퓨터가 동등한 입장에서 저장해둔 데이터와 파일을 별도의 중앙 서버를 거치지 않고 다른 사용자의 컴퓨터에 자유롭게 들어가 필요한 자료를 공유하는 것을 말한다.

P2P에서 피어(peer)의 사전적인 의미는 동등한 동료를 의미한다. 즉, P2P의 피어는 거대 네트워크상에서 클라이언트와 서버에 대응되는 개념으로 클라이언트와 서버를 특별히 구별하지 않고 네트워크에 참여하고 있는 구성원들 모두가 동등한 지위를 갖는

것을 말한다. P2P 프로그램에서는 기존의 클라이언트/서버구조에서 어느 구성원이나 전부가 클라이언트나 서버가 될 수 있기 때문에 컴퓨터를 클라이언트와 서버로 나누는 것은 무의미해졌다고 할 수 있다. P2P프로그램을 이용하여 사용자들은 별도의 서버를 구축하지 않고 디지털 정보자원을 상호교환 할 수 있게 되었고 궁극적으로 서로의 파일을 공유할 수 있게 되었다.

P2P의 발전은 (그림 1)과 같이 메인프레임에서 클라이언트/서버, 3-Tier 형태의 웹 환경, P2P로 발전해 왔다. 이러한 배경은 개인용 컴퓨터 성능이 서버와 비슷한 수준으로 급속히 발전함과 동시에 초고속 인터넷 환경이 제공되고 사용자들의 요구가 더욱 많아지면서 사용자간의 컴퓨터 시스템 리소스 및 데이터를 자유롭게 공유하기 위해 출현하게 되었다[1].



(그림 1) Peer-to-Peer의 발전

P2P는 피어들의 낮은 서버 의존도와 서버로부터 중앙 집중제어를 받지 않는다는 점과 클라이언트/서버 구조보다 피어에게 더 많은 제어 권한을 생성해주며 피어들이 정보를 공유하기 위해 서버의 도움이 필요 없이 자유롭게 파일을 주고 받을 수 있는 장점이 있다.

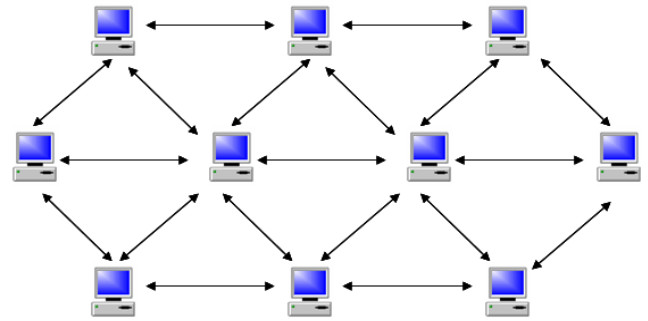
응용 분야에는 대표적으로 파일공유와, 분산 컴퓨팅이 있으며 분산 컴퓨팅은 자원공유, 즉 하드웨어의 자원을 공유함으로써 그 효율성을 극대화하기 위한 것이다. 분산 컴퓨팅은 대용량 데이터의 고속 연산처리가 가능하도록 무수히 많은 피어들의 CPU를 공유하는 모델이다. 또한 각 Peer들에게 담겨진 문서 중 가장 최신의 것이 채택되는 실시간 업데이트 형 분산 검색 엔진이 있으며 그밖에 협업, 전자상거래, 온라인 교육, 그룹웨어 등으로 다양하다.

P2P 서비스 모델은 크게 중앙 서버의 존재 여부에 따라 순수(Pure) P2P와 하이브리드(Hybrid) P2P

로 구분될 수 있다. 순수 P2P는 비슷한 성능을 가진 PC들 간의 연결 형태이고, 하이브리드 P2P는 PC들 간의 인터랙션을 원활하게 해주는 서버가 개입되는 형태이다[2].

2.1 순수(Pure) P2P 모델

순수P2P 모델 (그림 2)는 각각의 컴퓨터가 서버와 클라이언트의 모든 역할을 가지는 것이다.



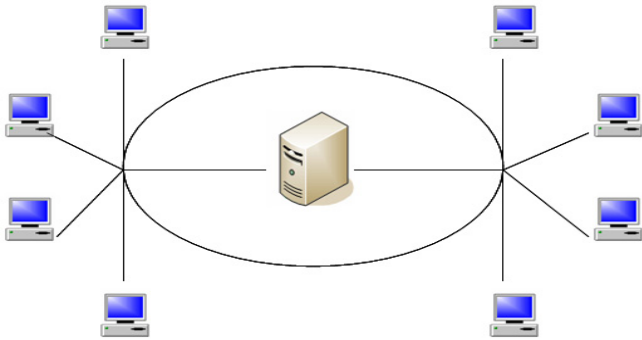
(그림 2) Pure Peer-to-Peer 모델

순수 P2P는 네트워크에 연결된 피어를 동적으로 찾고, 피어가 중앙서버 없이 작동하며 검색엔진은 개별 클라이언트들이 구동한 프로그램에 설치해서 익명성이 보장된 형태로 자료나 리소스를 공유한다. 순수 P2P는 클라이언트/서버의 관습적인 통신방법을 탈피하여 사용자가 규칙을 지정한다. 사용자는 자신의 네트워크 환경을 설정하여 상호 대칭적 의사소통을 이루며 파일을 검색하거나 파일을 다운로드할 때 모든 컴퓨터가 서버와 클라이언트로서의 역할을 하게 된다. 즉 별다른 중앙 서버를 둘 필요가 없는 장점이 있다[3]. 하지만 컴퓨터를 그물망으로 묶어서 검색명령을 재전송 하게 되어 검색시간이 길어지고 작은 네트워크 망을 형성할 경우 네트워크 부하가 걸릴 수도 있다는 단점을 가지고 있다. 또한 서버가 없는 동시에 시스템 전체에 대한 관리가 필요 없고 각 피어에서 개별적으로 인증하는 방법을 사용하여 서버 부하가 없다. 그러나 사용자에 따른 사용권한을 설정할 수 없으므로 인증 그룹마다 사용권한을 설정하여 사용함에 따라 고도의 보안 유지를 기대할 수 없으며 데이터의 생성과 색인 정보를 각 피어자신이 관리하고 검색하므로 데이터의 검색 결과 전체에 대한 빠른 응답을 기대하기 어렵다는 단점이 있다. 순수 P2P모델을 사용하는 대표적인 서비스에는 그누텔라(Gnutella)와 소리바다가 있다.

2.2 하이브리드(Hybrid) P2P 모델

하이브리드 P2P 모델은 피어간 상호 효율적으로

통신을 하고 필요한 정보를 전달하기 위해 중앙에 서버를 두어서 사용자들의 공유파일 목록 등을 제어하는 방법이다.



(그림 3) Hybrid Peer-to-Peer 모델

서버는 접속하는 피어에게 이미 접속된 피어의 정보를 제공하는 것에만 한정하며 접속을 수립하는 것과 통신을 수행하는 것은 피어들에 의해 이루어진다. 접속된 컴퓨터는 파일의 전송에 대해서만 서버와 클라이언트의 역할을 하게 된다. 하이브리드 P2P를 이용하면 검색의 속도나 검색 성공률을 높일 수 있다. 하지만 이 경우 동시 접속자가 늘어나면 서버측에 걸리는 부하도 늘어난다. 또한 하이브리드 P2P 모델의 특징은 검색용 색인관리를 위한 서버에 따라 시스템에 대한 계정 관리자가 필요하고 인증이나 이용 권한을 설정할 수 있어 서버를 통한 보안 유지가 가능하다. 그러나 서버 시스템의 운용관리가 요구되며 정기적 서버의 점검이나 주변 장치 및 통신 회선 등의 장애로 정지될 가능성이 존재한다. 하이브리드 P2P모델을 사용하는 대표적인 서비스에는 냅스터(Napster)와 이동키(eDonkey)가 있다.

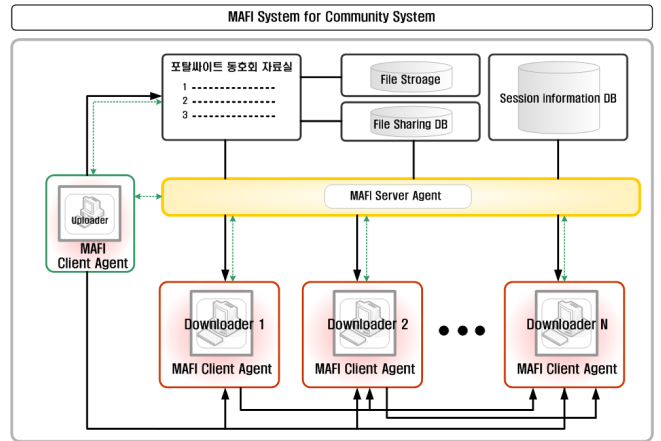
3. MAFI

MAFI(Sharing Tools of Mass File using P2P)는 포털 사이트의 동호회 등에서 사용 가능한 하이브리드 P2P 형태의 파일 공유 시스템이다. (그림 4)는 MAFI의 구조로서 크게 MAFI Server Agent, MAFI Client Agent, File Sharing DB, Session Information DB 등으로 구성된다.

3.1 MAFI의 파일 공유 개념

MAFI에서의 대용량의 파일을 공유할 수 있는 방법은 크게 두 가지로 나눌 수 있다. 첫째, 파일 업로더가 대용량 파일을 업로드 하고 있을 때, 해당 파일을 다운로드 하고자 하는 사용자가 서버에 업로드가 완료되기 전에 업로더의 PC로부터 P2P를 이용하여 직접 파일의 다운로드가 가능하다. 둘째, 서버 및

다른 사용자의 PC를 통하여 해당 파일을 이미 전송 받은 사용자는 또 다른 파일 서버가 되어 다른 사용자도 해당 파일을 다운로드할 때 해당 파일을 제공함으로써 해당 사용자는 파일 전송 시간을 단축시킬 수 있으며, 커뮤니티 사이트의 경우도 네트워크 대역폭 사용량을 감소시킬 수 있는 장점이 있다.



(그림 4) MAFI의 구조

3.2 MAFI Server Agent

MAFI Server Agent의 역할은 크게 두 가지로 나눌 수 있다.

첫째, MAFI Server Agent는 Uploader의 MAFI Client Agent로부터 업로드를 하겠다는 메시지를 받게 되면 동호회 자료실의 해당 자료에 대하여 “업로드중”이라는 메시지와 함께 “P2P 다운로드” 메시지를 같이 표시하도록 한다. “P2P 다운로드”를 다운로드가 선택하게 되면 해당 업로더로부터 직접 파일에 대한 부분 전송이 가능하다.

둘째, 다운로드가 대용량 파일을 다운로드 하고자 하는 경우 동호회 자료실의 해당 자료를 업로드 한 사용자나 이미 다운로드 한 사용자들의 세션 정보를 분석하여 다운로드가 가능한 피어가 있는 경우 다운로드의 MAFI Client Agent에게 해당 정보를 전송하여 해당 다운로드가 파일을 동호회의 자료실 및 전송 가능한 클라이언트들로부터 파일을 다운로드할 수 있도록 한다.

3.3 File Sharing DB

File Sharing DB는 동호회의 자료실에 저장되어 있는 파일들에 대한 메타 정보가 저장되어 있는 공간이다.

해당 파일을 최초로 업로드 한 사용자와 해당 파일을 이미 다운로드 한 사용자의 정보등과 해당 파

일에 대한 무결성을 검증하기 위한 각종 해쉬 정보들이 저장되어 있다. 특히 전송 가능한 피어 수에 따른 분할된 부분 조각들에 대한 각각의 해쉬 정보가 저장됨으로서 특정 파일을 다운로드 할 때 무결성을 검증할 수 있도록 한다. <표 1>과 <표 2>는 File Sharing DB에 저장되어 있는 데이터들의 일부이다.

<표 1> 파일 메타 정보

File Name	File Size	Hash Info	Uploader
1.avi	503,128,549	cc46045613c dc3ea4a3c04 ba6bd8f8cd8 378d34a	hyun43
2.wmv	128,430,394	2560f2da328 20143e8286a fb188f7c36d 905a735	choi
3.zip	364,304,234	dc46045613c dc3ea6a3c04 ba6bd8f8cd8 378d89a	jung

<표 2> 파일 슬라이스별 메타 정보

File Name	File Size	Slice Num	Hash Info
1.avi	503,128,549	3	31fdc2d7a1f1709aa0 2c9ea5854015645bd6 9504,a7be748fdb720 44acf0afab9c54bbc5 2f0a49b0f,a4feb5807 a406565f632cad9453 1d0bc7a57fd42
2.wmv	128,430,394	4	afdb5151e7bda76084 328bce46d5d91d010 4b9f8,038d38a29a6f5 db182d600d2f4b7ba 831b9d6c6c,f560f26a 32820143e8286afb18 8f7c36d905a735,4d2 000474c3f59beace97 ed5af15020a2630f2a 2

3.4 Session Information DB

동호회 회원 중에 다운로드로부터 연결이 가능한

피어들의 정보가 저장되어 있는 저장소(Repository)다. 다운로드들은 File Sharing DB와 Session Information DB의 정보를 이용하여 동호회에서 자료를 다운로드 받고자 하는 경우 사용 가능한 피어들의 정보를 수집한 후, 피어들 중에 연결 가능한 피어들과 통신을 하여 해당 파일을 다운로드 한다. 해당 파일을 다운로드할 때는 다운로드가 가능한 연결된 피어의 수가 증가함에 따라 전송 시간의 단축 및 서버 부하를 감소시킬 수 있다.

3.5 MAFI Client Agent

MAFI Client Agent는 사용자가 파일을 업로드하거나 다운로드하고자 할 때 필요한 구성 요소로서 사용자가 파일을 업로드 하고 있을 경우 해당 파일을 직접 다운로드하고자 하는 사용자가 있는 경우 해당 사용자와 직접 연결하여 데이터를 전송하는 기능을 수행한다. 또한 자료를 다운로드 하는 경우 동호회의 자료실에서 다운로드를 수행하면서 해당 파일을 제공 가능한 피어가 있는 경우, 분할해서 해당 파일을 다운로드 한다.

4. 결론 및 향후 연구

MAFI는 인터넷을 기반으로 동호회 사이트에서 네트워크 대역폭의 협소 등과 같은 많은 제한을 가지는 환경에서 효과적으로 대용량 파일등의 자료를 공유할 수 있도록 지원할 수 있는 시스템이다. 또한 사용자 측면에서도 대용량의 자료를 전송받을 때 동호회 사이트 외에 다른 사용자들로부터 관련 자료를 다운로드 받음으로써 자료 다운로드에 필요한 전송 시간을 절약할 수 있는 장점이 있다.

향후 연구 방향으로는 인터넷 동호회 사이트와 연동이 되는 폐쇄적인 형태의 순수 P2P 프로그램을 개발하고자 한다.

참고문헌

- [1] 최현우, "Next Generation Peer-to-Peer", 펜타시스템테크놀로지(주) 연구보고서, 2002.
- [2] Andy Oram, "Peer-to-Peer: Harnessing the Power of Disruptive Technology", O'Reilly, March 2001.
- [3] Sandvine Inc, "Peer-to-Peer File Sharing : The Effects of File Sharing on a Service Provider's Network," White Paper, Sandvine Inc., Jul.2002.