Proceedings of KFIS Spring Conference 2006 Vol. 16, No. 1.

신뢰 정보를 이용하는 하이브리드 P2P 파일 공유 시스템

A Hybrid P2P File Sharing System using Trust Information

손봉기[1], 김학준[2]

[1] 서원대학교 컴퓨터정보통신공학부
E-mail: bksohn@seowon.ac.kr
[2] 호원대학교 멀티미디어정보학과
E-mail: hjkim@howon.ac.kr

요 약

P2P 파일 공유 시스템은 중앙 서버를 거치지 않고 피어 컴퓨터들 간의 직접적인 자원 공유로 특징지워지는 P2P 기술의 대표적인 응용으로, 사용자는 인터넷을 통해 원하는 파일을 쉽게 검색, 획득, 유포할 수 있다. 그러나 악의적인 사용자가 허위 파일이나 바이러스를 유포할 수 있어, 허위파일로 인한 재전송이 빈번해짐으로써 네트워크 트래픽이 증가하고, 바이러스로 인한 피해가 발행할 수 있다. 또한, 자신의 파일에 대한 공유없이 다른 사용자들의 공유 파일만 다운로드하는 free riding 문제가 발생할 수 있다. 이 논문에서는 사용자의 파일 공유 트랜잭션 정보에 기반한 신뢰 정보를 이용하여 P2P 파일 공유 시스템의 문제점을 보완할 수 있는 시스템을 제안한다. 제안한 시스템에서 중앙 서버는 사용자의 파일 공유 트랜잭션 피드백을 이용하여 사용자 신뢰도를 구하여 허위파일이나 바이러스를 유포하는 악의적인 사용자나 free rider에 대해 시스템 사용을 제한한다. 사용자는 파일 다운로드 시에 중앙 서버로부터 사용자 메타데이터를 수신하여 자신의 선호도를 반영한 파일 신뢰도를 이용함으로써 허위파일이나 바이러스를 다운로드할 위험을 줄인다.

Key Words : P2P File Sharing System, Trust, Inauthentic File, Virus, Free Rider

1. 서 론

P2P 파일 공유 시스템은 중앙 서버를 거치지 않고 피어 컴퓨터들 간의 직접적인 자원 공유로 특징지워지는 P2P 기술의 대표적인 응용으로, 사용자는 인터넷을 통해 원하는 파일을 쉽게 검색, 획득, 유포할 수 있다[1]. 이에 반해, P2P의 개방성과 익명성은 유통되는 파일에 대한 책임을 물을 수 없어 악의적인 사용자들이 P2P 네트워크를 남용할 수 있다. 악의적인 사용자는 의도적으로 바이러스를 유포하여 다른 사용자에게 피해를 줄 수 있고, 허위 파일을 제공함으로써 재전송을 유발시켜 네트워크 트래픽을 증가시킬 수 있다[2]. 또한, 자신의 파일에 대한 공유 없이 다른 사용자들의 공유 파일을 다운로드만 하는 "free riding"이 빈번히 발생해 파일을 공유하는 사용자의 시스템 성능이 저하될 수 있다[3].

따라서, P2P 파일 공유 시스템에서 사용자가 제공받는 파일의 질에 대한 정보를 알 수 있는 방법과 free riding 문제 해결 방법이 필요하다.

이 논문에서는 하이브리드 파일 공유 시스템에서 사용자의 파일 공유 트랜잭션에 기반한 신뢰 정보를 이용하여 악의적인 사용자와 free riding을 행하는 "free rider"로부터의 위험을 줄일 수 있는 시스템을 제안한다. 제안한 시스템에서 중앙 서버는 사용자의 공유 파일 목록, 허위 파일 유포, 바이러스 유포에 대한 스코어와 다른 사용자에 대한 피드백 회수 등에 대한 사용자 메타데이터를 유지하고, 이 정보를 이용한 신뢰도로 악의적인 사용자나 free rider 구분하여 시스템 사용을 제한한다. 사용자는 파일을 다운로드할 상대를 선택할 때, 중앙 서버에서 제공하는 사용자 메타데이터를 이용하여 파일 신뢰도를 이용한다. 파일 신뢰도는 파일 소유자에 대한 회선 속도, 허위 파일 및 바이러스 유포 스코어에 대해 선호도를 반영하여 구한다. 사용자는 신뢰 정보에 의해 파일을 다운로드한 후, 이에 대한 피드백을 서버로 전송한다.

이 논문의 구성은 다음과 같다. 2장에서는 P2P 파일 공유 시스템과 신뢰 정보를 이용한 시스템에 대해 알아본다. 3장에서는 제안한 시스템을 소개하고, 4장에서 결론을 맺고 향후 과제를 제시한다.

2. 관련 연구

3.1 P2P 파일 공유 시스템

P2P 파일 공유 시스템은 중앙 서버가 등록된 사용자의 공유 파일 목록과 회선 속도 등에 대한 메타데이터를 유지하면서 사용자간의 직접적인 파일 공유를 연결해주는 하이브리드(hybrid) 방식과 중앙 서버 없이 직접 파일을 공유하는 순수(pure) 방식이 있다[4]. 순수 방식은 중앙 서버가 불필요하다는 장점이 있지만 파일을 검색할 때마다 네트워크의 모든 컴퓨터를 직접 접속하여 검색하는 과정을 반복하기 때문에 과도한 트래픽일 발생할 수 있다. 하이브리드 방식은 중앙 서버에 대한 단일 지점 오류가 발생할 수 있지만 서버에서 파일을 검색하기 때문에 빠른 검색이 가능하다. 또한, 사용자의 접속여부, 회선 속도, 과거 트랜잭션 정보 등을 관리할 수 있기 때문에 악의적인 사용자나 free rider를 구분할 수 있는 정보를 제공할 수 있다.

3.2 신뢰 정보를 이용한 파일 공유 기법

신뢰(trust)는 전자상거래, P2P, 유비쿼터스 컴퓨팅과 같은 개방된 환경에서 트랜잭션의 불확실성과 위험을 감소시킬 수 있는 수단으로 평가되고 있다[5]. 악의적인 사용자나 free rider로부터의 위험을 줄이기 위해 신뢰 정보를 이용한 P2P 파일 공유 시스템에 대한 몇 가지 연구가 있었다.

Damiani 등[6]은 순수 방식에서 분산된 polling 알고리즘을 기반으로 피어의 평판(reputation)에 관한 정보를 공유하기 위한 방법으로 신뢰할 수 있는 파일 제공 피어를 선택할 수 있지만 평판을 수렴하는 과정에서 통신량이 많다.

Kamvar 등[2]은 P2P 네트워크에서 인증되지 않은 파일의 다운로드 수를 감소시킬 수 있는 알고리즘으로 과거 트랜잭션 정보를 이용해 피어에 대한 유일한 전체 신뢰값(global trust value)을 할당하여 악의적인 피어를 식별할 수 있지만, 전체 신뢰값을 구하는데 통신량이 증가한다.

[3]에서는 하이브리드 방식에서 평판 정보를 이용한 접근 제어 기법을 제안하였지만, 평판 계산량이 많고 사용자의 선호도를 반영하지 않고 있다.

악의적인 사용자는 허위 파일이나 바이러스를 무수히 만들어낼 수 있기 때문에 허위 파일 자체를 알아낼 수 있는 방법보다는 허위 파일을 제공하는 악의적인 사용자를 구별하는 것이 더 우선적이라 할 수 있다[2]. P2P 파일 공유 시스템에서 free rider는 사용자가 참여한 트랜잭션 정보를 통해 알아낼 수 있다. 즉, 사용자가 일정 기간동안 참여한 파일 공유 트랜잭션 수가 작다면 그 사용자는 free rider로 간주할 수 있다. 악의적인 사용자나 free rider는 성능 향상을 위해 중앙 서버에서 시스템 이용을 제한할 수 있어야 한다. 이 논문에서는 사용자간의 파일 공유 트랜잭션 정보를 기반으로 하여 악의적인 사용자와 free rider를 구분하고, 이들로부터의 위험을 줄일 수 있는 하이브리드 파일 공유 시스템을 제안한다.

3. 신뢰정보를 이용하는 하이브리드 P2P 파일 공유 시스템

제안하는 시스템의 모든 사용자는 중앙 서버에 계정을 가지고 있으며, 시스템을 이용하기 위해 중앙서버에 인증 과정을 거친다. 인증 과정을 거친 사용자는 서버에 자신의 공유 파일 목록을 등록한다.

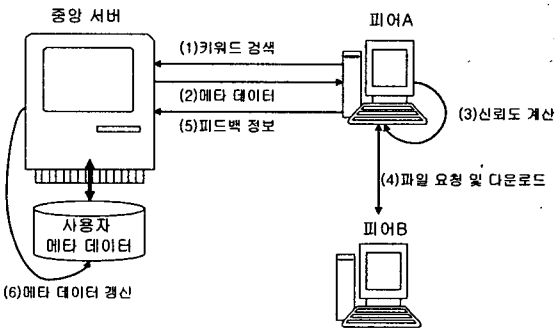

그림 1. 시스템 구성 및 동작 시나리오

3.1 시스템 구성 요소 및 동작 시나리오

그림 1은 제안한 시스템의 구성 요소 및 동작 시나리오를 나타낸 것이다. 중앙 서버는 사용자에 대한 메타데이터를 유지하고, 이를 이용해 사용자 신뢰도(user trust degree)를 구하여 악의적인 사용자와 free rider에 대해 시스템 사용을 제한하고, 사용자로부터의 질의에 응답한다. 피어 컴퓨터는 중앙 서버에 대해 파일 목록 및 사용자 정보를 요청하고, 이를 이용해 파일 다운로드 대상을 결정하고 파일을 공유한다. 시스템의 동작 시나리오는 다음과 같다.

(1) 사용자는 파일 검색에 대한 키워드 질의를 중앙 서버에 전송한다.
(2) 중앙 서버는 검색 키워드와 일치하는 파일에 대한 사용자의 메타데이터를 해당 피어로 전송한다.
(3) 피어는 중앙 서버로부터 수신한 메타데이터에 대해 사용자의 선호도를 반영하여 파일 신뢰도를 계산한다.

(4) 피어는 가장 높은 파일 신뢰도의 사용자 피어 컴퓨터에 대해 파일을 요청하고 다운로드한다.
(5) 사용자는 다운로드한 파일을 확인하고, 중앙서버로 피드백 정보를 전송한다.
(6) 중앙 서버는 피드백 정보를 이용해 사용자 메타데이터를 갱신한다.

3.2 사용자 메타데이터

중앙 서버는 사용자에 대한 메타데이터를 유지하면서 악의적인 사용자와 free rider를 판별하여 시스템 사용을 제한하고, 피어로부터의 질의에 응답한다. 사용자의 메타데이터에 유지하는 속성은 표 1과 같다. 사용자 메터데이터에는 사용자 ID, 회선 속도, 공유 파일 목록과 같은 사용자에 대한 기본적인 정보와 신고 회수, 허위 파일 신고 회수, 바이러스 신고 회수, 피드백 회수와 같은 파일 공유 트랜잭션 정보, 허위 파일 유포 스코어, 바이러스 유포 스코어, 신뢰도 정보가 포함된다.

파일 공유 트랜잭션 정보는 사용자 U_i가 다른 사용자에 대해 피드백을 제공한 회수와 제공한 파일을 다운로드한 다른 사용자가 제공한 피드백 정보로 얻어진다. 피드백 정보는 '허위파일', '바이러스', 'Good'의 경우 중 하나로 제공된다. U_i에 대한 피드백 정보는 허위 파일 유포 스코어, 바이러스 유포 스코어, 신뢰도 계산에 사용된다.

1) 허위 파일 유포 스코어

U_i에 대한 허위 파일 유포 스코어는 사용자 i가 공유한 파일에 대해 시간 구간 $[t-dt, t]$동안 다른 사용자가 다운로드한 후 허위 파일이라고 피드백한 결과를 정량화한 것으로 식 (1)과 같이 정의한다.

$$S_i^{inauthentic} = \frac{N_i^{inauthentic}[t-dt, t]}{n_i^{feedback}[t-dt, t]} \quad (1)$$

허위 파일 유포 스코어가 높을수록 사용자 i는 자주 허위파일을 유포했다는 것을 의미한다.

2) 바이러스 유포 스코어

사용자 i가 시간 구간 $[t-dt, t]$ 동안 바이러스를 유포한 것을 정량화한 것으로 바이러스 유포 스코어는 식 (2)와 같이 정의한다.

$$S_i^{virus} = \frac{N_i^{virus}[t-dt, t]}{n_i^{feedback}[t-dt, t]} \quad (2)$$

3) 사용자 신뢰도

사용자 신뢰도는 중앙 서버가 사용자 i의 피드백 정보를 수용하거나 거절할 때 사용되는 측도이다. 또한, 악의적인 사용자나 free rider에게 시스템 사용을 제한할 때도 사용된다. 사용자 i의 신뢰도는 식 (3)과 같이 정의한다.

표 1. 사용자 메타데이터

속성	표기법	설명
사용자 ID	U_i	사용자 i의 ID
회선 속도	U_i^{speed}	사용자 i의 회선 속도
공유파일 목록	U_i^{files}	사용자 i의 공유 파일 목록
신고 회수	$n_i^{otherfd}$	사용자 i가 다른 사용자에 대해 피드백한 회수
허위파일 신고 회수	$N_i^{inauthentic}$	사용자 i에 대한 다른 사용자의 허위파일 유포 신고 회수
바이러스 신고 회수	N_i^{virus}	사용자 i에 대한 다른 사용자의 바이러스 유포 신고 회수
피드백 회수	$n_i^{feedback}$	사용자 i에 대한 다른 사용자의 피드백 회수
허위파일 유포 스코어	$S_i^{inauthentic}$	사용자 i의 허위파일 유포 스코어
바이러스 유포 스코어	S_i^{virus}	사용자 i의 바이러스 유포 스코어
신뢰도	T_i^{user}	사용자 i의 신뢰도

$$T_i^{user} = S_i^{inauthentic} + S_i^{virus} + \frac{1}{n_i^{feedback}} + \frac{1}{n_i^{otherfd}} \quad (3)$$

사용자 i가 허위파일과 바이러스를 자주 유포하거나 free riding을 할수록 T_i^{user}가 높아져 신뢰도가 낮아지게 된다. 다른 사용자로부터의 피드백 회수가 작다는 것은 사용자 i가 free riding한 것으로 볼 수 있기 때문에 피드백 회수의 역수를 취하여 신뢰도에 반영한다. 어떤 사용자는 파일 다운로드 후 피드백을 하지 않을 수 있다. 따라서 다른 사용자에 대한 피드백 회수에 역수를 취하여 신뢰도에 반영함으로써 피드백 정보를 중앙 서버로 전송해야만 자신의 사용자 신뢰도가 높아지게 한다. 중앙 서버는 사용자 i의 T_i^{user}가 제한된 임계값 θ_r를 넘으면 식 (4)와 같이 사용자 i로부터의 피드백 정보를 반영하지 않거나 시스템 사용을 제한할 수 있다.

$$\begin{cases} T_i^{user} \geq \theta_r : \text{시스템 사용제한, 피드백 미반영} \\ T_i^{user} < \theta_r : \text{시스템 사용, 피드백 반영} \end{cases} \quad (4)$$

3.3 파일 다운로드를 위한 파일 신뢰도 계산

사용자의 키워드 검색에 대한 결과로 중앙 서버는 키워드와 일치하는 파일의 사용자메타

데이터를 피어 컴퓨터로 전송한다. 피어는 다운로드 파일 선택 시에 사용자 선호도를 반영한 파일 신뢰도(file trust degree)를 구하여 가장 적합한 사용자로부터 파일을 다운로드한다.

파일 신뢰도 계산의 평가 기준으로 파일 공유자의 회선 속도, 허위파일 유포 스코어, 바이러스 유포 스코어를 고려한다. 사용자는 각 평가 기준에 대한 선호도를 반영하여 신뢰도를 구하는데, 사용자에 따라 파일 다운로드 속도, 키워드와 일치하는 파일, 안전한 파일을 가장 우선시 할 수 있기 때문이다. 사용자 i가 공유한 파일 신뢰도는 식 (5)와 같이 정의한다.

$$T_i^{file} = S_i^{inauthentic} \cdot w_i + S_i^{virus} \cdot w_v + \frac{1}{U_i^{speed}} \cdot w_s \quad (5)$$

식 (5)에서 w_i, w_v, w_s는 각각 허위파일 유포 스코어, 바이러스 유포 스코어, 회선 속도에 대한 가중치를 의미한다. 회선 속도는 빠를수록 선호하기 때문에 역수를 취하여 신뢰도에 반영한다. T_i^{file}가 낮을수록 높은 신뢰 정도를 나타낸다.

3.4 피드백 및 사용자 메타데이터 갱신

사용자는 파일 신뢰도가 가장 높은 사용자의 피어 컴퓨터로부터 파일을 다운로드한 후 중앙 서버로 피드백 결과를 전송한다. 식 (3)과 같이 사용자가 피드백 정보를 중앙 서버로 전송하지 않으면 사용자 신뢰도가 낮아지게 되기 때문에 사용자는 피드백 정보를 전송하게 된다.

사용자로부터의 피드백 정보를 수신받은 중앙 서버는 식 (4)를 이용해 피드백 반영 여부를 결정하고 사용자 메타데이터를 갱신한다.

5. 결론 및 향후 연구

최근 P2P 기반의 파일 공유 시스템이 많은 인기를 끌고 있다. 그러나 P2P 시스템의 개방성과 익명성을 이용한 악의적인 사용자가 허위 파일이나 바이러스를 유포하거나 free riding과 같은 문제가 발생하고 있다. 이로 인한 문제는 P2P 기술 발전을 위해 반드시 해결되어야 할 문제이다.

이 논문에서는 P2P 파일 공유 시스템을 사용하는 사용자의 파일 공유 트랜잭션 정보에 기반한 신뢰 정보를 이용하여 P2P 파일 공유 시스템의 문제점을 보완할 수 있는 시스템을 제안하였다. 제안한 시스템에서 중앙 서버는 사용자 과거 트랜잭션 정보를 이용한 사용자 신뢰도를 이용해 악의적인 사용자나 free rider에게 시스템 사용을 제한할 수 있다. 또한, 사용자는 파일 다운로드 시에 중앙 서버로부터 사용자 메타데이터를 수신하여 자신의 선호도를 적용한 파일 신뢰도를 계산하여 사용하기 때문에 허위 파일이나 바이러스를 다운로드 받을 위험을 줄일 수 있다.

향후 연구로는 제안한 시스템 구현을 통해 기존 신뢰 정보를 이용한 시스템과의 비교 실험이 필요하다.

감사의 글: 이 연구는 호원대학교 교내학술연구비 지원을 받아 수행한 것입니다. 연구비 지원에 감사드립니다.

참 고 문 헌

[1] S. Androutsellis-Theotokis, D. Spinellis, "A Survey of Peer-to-Peer File Sharing Technologies," White Paper, Athens University, 2002.

[2] S. D. Kamvar, M. T. Schloser, H. Garcia-Molina, " The EigenTrust Algorithm for Reputation Management in P2P Networks," In Proceedings International WWW Conference, 2003.

[3] 신정화, 신원, 이경현, "P2P 파일 공유 시스템에서 평판 정보를 이용한 접근 제어," 정보처리학회논문지, 제12-A권 제 6호, pp.493-498, 2005.

[4] 이인, 황종인, "P2P 환경에서 피어 관리 기법을 이용한 효과적인 다운로드 방법," 한국정보처리학회 춘계학술발표대회논문집, 제12권 제1호, pp.959-962, 2005.

[5] P. Dasgupta, "Trust as a Commodity," In D. Gambetta, editor, Trust:Making and Breaking Cooperative Relations, pp.49-72, Blackwell, 1998.

[6] E. Damiani, C. Vimercati, and et al, " A Reputation-Based Approach for Choosing Reliable Resources in Peer-to-Peer Networks," Conference on Computer and Communications Security archive Proceedings of the 9th ACM, pp.207-216, 2002.

[7] F. Cornelli, E. Daminani, and et al, "Choosing Reputable Servants in a P2P Network," In Preceedings of the 11th WWW Conference, 2002.