논문 2014-1-5

토렌트에서의 콘텐츠 저작권보호 방안

이규대*

A Copyright of contents on the Torrent protocol

Kyu-Tae Lee*

요 약

토렌트는 개인 간의 파일교환에 있어서 운영의 관리 주체를 배제하는 방법으로 등장한 기술로 불법 콘텐트의 유통시, 최초 유포자를 찾기가 어렵고, 공유대상 파일이 조각 화일로 전송되기 때문에 전체 화일을 불법으로 다운로드 하였는지 구분이 불가능한 특징을 보이고 있다. 따라서 토렌트를 이용하는 콘텐트의 공유는 콘텐트 산업의 발전을 저해하는 심각한 사회문제로 부각되고 있다. 콘텐츠의 보호와 토렌트의 활성화를 위한 방안을 찾는 연구가 저작권보호 차원에서 필요한 상황으로 인식되고 있다. 본 연구에서는 토렌트 기술의 특징을 분석하고, 토렌트 환경에서 유통되는 콘텐트를 보호하기위한 방안을 제시한다.

Abstract

Torrent is the new technology for exchanging the files between each private person without handling of system operator. But it has a harmful disadvantages to digital contents owner cause the contents file could be transmitted with many divided files pieces even illegal copyright content. Also it could not be traced who is the heavy loader. By this reason, Torrent is known as the illegal content provide technology and many service provider like webserver manager, is now protecting the service to users. This paper shows the way of safe usages to contents for both of owner and torrent service provide.

한글키워드 : 토렌트, 웹하드, 콘텐트 보호

1. 서 론

인터넷은 네트워크로 연결된 모든 컴퓨터간의 정보교환이 시간적, 공간적 제약을 받지 않고, 언제 어디서나 어떤 사람과 정보교환이 가능한 환경을 제공하는 정보기술이다. 또한 스마트폰의 보급으로 이동상황에서도 자유롭게 인터넷에 접근이 가능해져 다양한 콘텐츠의 활용이 가능해지고 있다. 그러나 음악, 영화, 소설 등의 저작권을 보호받아야하는 정보 콘텐트가 디지털화하면서 파일로 작성된 콘텐트가 인터넷을 통해 무제한적으로 공유될 수 있고, 무료로 모든 콘텐츠를 사용하려는 사용자들로 인해 저작권보호에 어려움을 가중하는 문제를 나타내고 있다.[1] 또한

* 공주대학교 정보통신공학부
(email: ktlee@kongju.ac.kr)
접수일자: 2014.6.2 수정완료: 2014.6.27

인터넷 상에서 개인적 파일공유의 목적으로 사용되고 있는 유료 웹사이트가 불법적인 콘텐츠의 유통 수단으로 활용되면서 다양한 방법의 규제가 시행되고 있다. 비슷한 목적으로 활용되고 있는 P2P 방식은 서버의 관리자가 필요 없는 방법으로 개인 간의 파일 공유 목적으로 제작되어 활용되고 있으나, 이 또한 불법 콘텐트의 통신에 악용되고 있다.

운영의 관리 주체를 배제하는 방법으로 등장한 토렌트는 불법 콘텐트의 유통시, 최초 유포자를 찾기가 어렵고, 조각 파일로 전송되는 기술적 특징으로 전체 파일을 불법으로 다운로드 하였는지 구분이 불가능한 특징을 보이고 있다.[2] 따라서 토렌트를 이용하는 콘텐트의 공유는 콘텐트 산업의 발전을 저해하는 심각한 사회문제로 부각될 수 있기 때문에 저작권보호 차원의 연구가 집중되고 있다.

본 연구에서는 토렌트 기술의 특징을 분석하고, 토렌트 환경에서 유통되는 콘텐트를 보호하기위한 방안을 제시하였다.

2. 토렌트 시스템

비트토렌트(BitTorrent)는 개인들 간의 네트워크를 P2P(peer-to-peer)로 연결하여 대용량 파일을 빠른 속도로 전송하는 목적으로 만들어진 프로토콜이며, 응용소프트웨어의 이름이다. [3]

이 프로토콜은 2001년 미국의 Brahm Cohen이 창안 하였으며, BitTorrent Inc. 회사를 설립하여 관리 배포하고 있다. 이 방법의 특징은 공유대상 파일을 직접적으로 다운로드하는 기존의 전통 P2P방식과 달리, 공유대상 파일을 여러 개의 조각(piece)으로 분리하여, 병렬로 다운로드가 가능하도록 하는 기술로, 트랙커라 불리는 서버가 사용자들 간의 파일 업/다운 정보만을 제공하고, 실제의 파일은 전송하지 않는 원리로 운영된다.

2.1 시스템 구성

비트토렌트는 5개의 구성요소로 공유대상 파일을 전송한다. 여기서 시더(seeder)는 공유대상 화일의 모든 조각을 가지고 있는 클라이언트이고, 리처(leecher)는 공유대상화일의 일부 조각만을 보유하고 있는 클라이언트, 피어(peer)는 시더와 리처를 모두 가리킨다. 트래커(tracker)는 공유대상화일의 피어정보를 관리하는 서버이고, 스웜(swarm)은 해쉬(hash)값과 공유화일을 소유하고 있는 피어들로 트래커에 의해 관리된다.

구성요소	내용
MetaInfo File (*.torrent)	· 토렌트 파일에 대한 모든 정보를 포함 · announce, announce-list, SHA1 해쉬코드, # of piece, created by, creation date, file name/size, piece size
Data	· single file or multiple file(directory) · 데이터는 같은 크기의 조각(piece)들로 분리 · common piece size : 256KB, 512KB and 1MB
Tracker	· swarm[1]을 구성 · 클라이언트에게 파일 소스의 위치를 추적 · 각 클라이언트 간, 파일 업/다운로드 조정 · HTTP/HTTPS protocol(포트: 6969)
Peers	· seeder[2] and leecher[3] · Peer와 TCP 통신(포트: 6881~6889)
Client	· 사용자 PC에서 설치되어 있는 다운로드 프로그램 · BitTorrent, µTorrent, deluge, KTorrent, Vuze 등

그림 1. 비트토렌트 구성요소

2.2 다운로드 프로세스

비트토렌트는 기존의 P2P와 같이 독립적으로 파일을 다운로드하지 않는다. 공유화일을 다운로드 하고자 하는 사용자는 토렌트 파일을 공유하는 웹사이트로부터 토렌트 파일을 다운로드하고, 클라이언트 프로그램을 이용하여, 다운로드한 토렌트화일을 실행함으로써 공유대상 파일의 다운로드가 가능하다.

클라이언트 프로그램은 비트토렌트 외에 다양한 프로그램이 등장하였다. 즉, 비트코멧·비트스피리트·비트토란도·델류즈·K토렌트·라임와이어·ML동키·q비트토렌트·쉐어아자·트랜스미션·µ

토렌트·뷰즈(이전에는 Azureus)·Xunlei 등이 사용되고 있으며, uTorrent 가 많이 사용되고 있는 것으로 알려져 있다.

즉, 사용자가 클라이언트 프로그램을 이용하여, 공유대상파일의 다운로드를 시작하면, 토렌트 화일 내의 트래커에 연결되고, 트래커는 관련 공유화일의 동일 토렌트화일(해쉬코드)을 가지고 있는 사용자 리스트 정보를 제공한다.

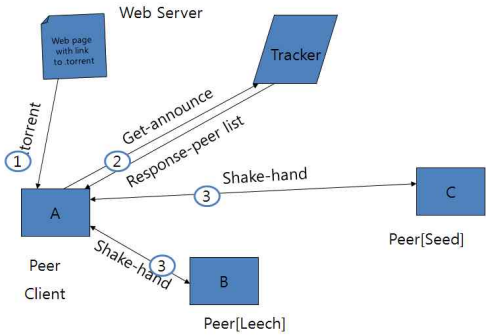

그림 2. 비트토렌트시스템 파일공유 프로세스

이렇게 연결된 사용자 들을 swarm 이라 한다. 이러한 방식은 기존의 P2P 네트워크가 서버중심의 파일공유 시스템인 것과 달리 공유화일을 중심으로 네트워크가 형성되는 차이가 있다. 트래커도 단순 정보제공 용으로 사용될 뿐, 네트워크의 중심역할을 하는 것이 아니다.

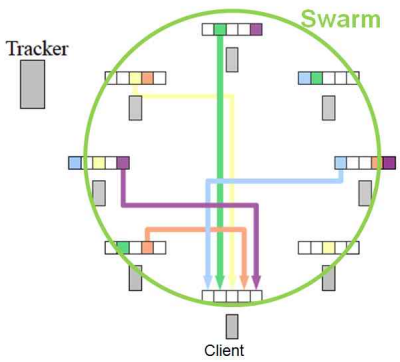

그림 3. 조각(piece)파일 다운로드 프로세스

사용자는 다른 사용자의 IP, Port 정보 등을 이용하여 네트워크를 구성하고, 해당 공유파일을 조작파일 단위로 다운로드와 업로드를 수행한다. 일반적으로 조각의 크기는 256kb이며, 공유대상 화일의 크기가 클수록 조각 파일의 크기도 커진다. 전송효율을 최대화 하는 방법으로 트래커는 각 클라이언트에게 파일정보를 제공할 때, 한 클라이언트에 집중되지 않도록 배분하는 방법을 사용한다.

2.3 프로토콜 특징

비트토랜트는 분산파일을 위한 프로토콜로, URL이 지정하는 콘텐츠를 나타내며, 웹을 통해 연속적으로 집적화(integrated)되도록 설계된다. 일반적인 HTTP에 대한 장점은 동일한 파일의 다중 다운로드가 동시에 발생할 때, 다운로더는 서로 업로드하여, 파일소스가 적절한 증분만으로 많은 수의 다운로더를 지원할 수 있다는 것이다.

비트토렌트의 파일분산은 다음의 엔터티로 구성된다.

- 일반적인 웹서버
- 메타인포(metainfo) 파일
- 비트토렌트 트렉커(tracker)
- original 다운로더
- 사용자 웹브라우저(end user)
- 사용자 다운로더(end user)
- 한 파일에 대한 이상적으로 많은 사용자 (end user)

비트토렌트 서비스에서 호스트는 다음의 단계를 따라 시작된다.

- 트랙커를 실행한다.
- 웹서버를 실행한다.(apache 와 같은)
- 웹서버에서 (mimetype application/ x-bittorrent)를 확장자 (*.torrent) 와 연결한다.
- metainfo(.torrent) 파일을 생성한다. 트래커

의 URL과 서비스될 최종 파일이다.
- metainfo 파일을 웹서버에 복사한다.
- metainfo 파일을 몇 개의 다른 웹페이지에 링크한다.
- 전체화일을(origin) 가지고 있는 다운로더를 실행한다.

사용자가 다운로드를 하기위해서는 다음의 절차를 수행한다.
- BitTorrent를 설치한다.
- 웹페이지를 찾아 (.torrent) 파일에 링크된 페이지를 클릭한다.
- 파일이 저장되어있는 곳을 선택하고, 다운로드가 완료될 때 까지 기다린다.
- 다운로더가 종료되었음을 알린다.

bencoding
- 문자열(strings)은 콜론과 문자(string)으로 연속된 10개의 고정길이로 작성된다.
- 예를들어 4:spam 은 ' spam ' 과 관련되고, 정수는 'i'와 십진수, 그리고'e'가 추가된다.

즉, i3e 는 정수'3'이고, i-3e 는 정수 -3을 나타낸다. 정수는 크기제한이 없으며, 단, iI-0e 의 표현은 허용되지 않는다. 또한 0이 따라오는 정수 표현은 허용되지 않는다(i03e). 즉, i0e 의 표현이 0을 나타낸다.

리스트(Lists)는 'l'로 시작하고 'e'로 종료하는 문자열로 인코드 된다. 즉 'l4:spam4:eggse' 의 표현은 'spam','eggs' 로 인코드된다.

사전은(Dictionaries)은 'd' 로시작하고 'e'로 종료하는 선택키를 갖는 문자열로 나타낸다. 즉, 'd3:cow3:moo4:spam4:egge' 는 {'cow':'moo', 'spam': 'eggs'}를 나타낸다. 또한 'd4:spam l1:a1:bee'는 {'spam': ['a','b']}를 나타낸다. 키는 문자열이어야 하며, 저장된 순서로 표시된다.

2.4 peer 프로토콜

비트토렌트의 피어 프로토콜은 TCP 또는 uTP에서 운영된다.

피어 연결은 대칭적이며, 양방향으로 전송되는 메시지는 유사하고, 데이터는 어떤 방향으로도 이동할 수 있다.

피어 프로토콜은 메타인포(metainfo) 파일에서 설명한 것처럼 0으로 시작되는 인덱스에 의해 파일의 조각들을 나타낸다. 피어가 한조각의 다운로드를 종료하고 , 해쉬의 매칭을 확인하면, 모든 피어들에게 다운로드된 조각을 가지고 있다는 것을 알리게 된다.

연결(connection)은 양끝 단에 두 비트의 상태를 포함한다.(choked or not) 그리고 (interested or not). 쵸킹(choking)은 unchoking 이 발생할 때 까지 데이터가 전혀 송신되지 않았음을 알리는 것이다.

데이터 전송은(data transfer) 한쪽이 interested 이거나, 다른 쪽이 choking 이 아닐 때마다 발생한다. 인터레스트상태(Interest state)는 모든 시간마다 수정되어야 한다. 다운로더가 피어에 요청한 것이 없을때마다 choked 되었음에도 인터레스트가 없다는 것을 보여야 한다.

이러한 상태를 적절히 구현하는 것은 기술이 필요하다. 그러나 다운로더에게 피어가 바로 다운로드를 시작한다는 것을 알려줄 수 있게 된다.

3. 저작권보호와 토렌트 활성화

3.1 토렌트 파일에서의 저작물 검색

토렌트는 웹하드나 P2P와 같이 콘텐츠 파일을 직접 다운로드하는 형식이 아니고, 사용자 컴퓨터에 설치된 토렌트 클라이언트 프로그램을 이용하여, 웹서버로부터 공유대상 파일의 정보를 갖고 있는 토렌트 파일로부터 실제 파일이 존재하는 컴퓨터를 찾아가는 방식이다. 따라서 이를 추

적하기 위해서는 토렌트 파일을 다운로드하는 부분과, 실제 콘텐츠를 다운로드하는 부분으로 분리하여 검색하는 방법을 사용하여야 한다.

3.2 토렌트에 의한 저작권 침해

2012년도 한국저작권단체 연합회에서 조사한 유통경로별 불법 복제물 증감률은 웹하드의 경우 15.0%, 포탈의 경우 3.6% 감소를 보이고, 토렌트의 경우는 29.3%의 증가로 나타나, 토렌트를 이용한 유통이 확대되고 있음을 보이고 있다. 이는 토렌트의 특성상 불법 파일의 유포자가 확인되기 어려운 점을 활용한 결과로 보인다.

즉, 익명네트워크 기술을 사용한 Tor 기반의 Easy-Tor 의 등장으로 불법 복제물을 유통시키는 헤비시더(heavy seeder: 대용량 파일 제공자)의 추적이 어려운 환경을 활용한 결과로 보인다.

그림 4. 불법복제 유통현황(2012 저작권보호연차보고서)

토렌트의 특징을 살리면서 인터넷상에서 파일공유를 자유롭게 하기위해서는 불법 콘텐츠의 유통을 제한하는 기술이 적극 개발 활용되어야 하는 이유이다.

3.3 토렌트활성화 방안

토렌트의 활성화 방안의 핵심은 저작권보호를 보장하는 방법이다. 즉, 불법 콘텐츠의 유통이 토렌트환경에서 발생시, 이를 차단하거나 적발해 낼 수 있는 방법이 개발되고, 검증될 경우 웹하드 서비스와 같이 합법적인 프레임속에서 공유자료 파일의 원활한 소통이 가능해지고, 웹하드와 같은 새로운 토렌트 사업영역의 확장도 기대할 수 있게 된다.

- 토렌트 패킷분석에 의한 차단

토렌트 패킷을 NetFPGA 등의 실시간 패킷검사장비를 이용하여, 패킷을 추출하고, 패킷내의 해쉬코드를 실시간 검색방법으로 매핑하는 방안이다. 이 방법은 보호대상 콘텐츠의 해쉬코드를 데이터베이스로 저장하고, 입력되는 모든 패킷의 해쉬코드를 비교하는 방법이다.

Parameter	Description
info_hash	20-byte SHA1 hash of the value of the info key from the Metainfo file.
peer_id	20-byte string used as a unique ID for the client, generated by the client at startup
port	The port number that the client is listening on. Ports reserved for BitTorrent are typically 6881-6889.
uploaded	The total amount uploaded so far, encoded in base ten ascii.
downloaded	The total amount downloaded so far, encoded in base ten ascii.
left	The number of bytes this client still has to download, encoded in base ten ascii.
event	If specified, must be one of started, completed, or stopped. If not specified, then this request is one performed at regular intervals.
-started	The first request to the tracker must include the event key with the started value.
-stopped	Must be sent to the tracker if the client is shutting down gracefully.
-completed	Must be sent to the tracker when the download completes. However, must not be sent if the download was already 100% complete when the client started.
ip	Optional. The true IP address of the client machine, in dotted quad format or rfc3513 defined hexed IPv6 address.
numwant	Optional. Number of peers that the client would like to receive from the tracker. This value is permitted to be zero. If omitted, typically defaults to 50 peers.

그림 5. 트래커 프로토콜(request)

Key	Description
failure reason	If present, then no other keys may be present. The value is a human-readable error message as to why the request failed (string).
interval	Interval in seconds that the client should wait between sending regular requests to the tracker
Tracker id	String that the client should send back on its next announcements. If absent and a previous announce sent a tracker id, do not discard the old value; keep using it.
complete	number of peers with the entire file, i.e. seeders (integer)
incomplete	number of non-seeder peers, aka leechers (integer)
peers	The value is a list of dictionaries, each with the following keys
-peer id	peer's self-selected ID, as described above for the tracker request (string)
-ip	peer's IP address (either IPv6 or IPv4) or DNS name (string)
-port	peer's port number (integer)

그림 6. 트래커 프로토콜(response)

서버 또는 라우터등의 기기를 보유하고, 서비스가 가능한 사업자가 주기적인 데이터베이스 관리를 통해 콘텐츠보호가 가능하게 된다. 또한 저작권보호기관이나, 중요 콘텐츠관리 기관에서

는 주기적인 해쉬코드정보를 제공함으로써 효율적인 불법콘텐츠 유통을 방지하면서 토렌트 서비스를 사용할 수 있을 것으로 보인다.

4. 결 론

개인 간의 파일공유를 원활하게 제공할 목적으로 만들어진 토렌트 기술이 저작권보호의 영역에서 불법 콘텐츠의 유통방법으로 활용되는 문제점이 제기되어, 많은 시스템 관리자들이 콘텐츠의 서비스를 차단하고 있다. 본 논문은 콘텐츠의 저작권보호를 유지하면서 토렌트 기술이 개인 간의 정보교환의 수단으로 원활히 활용되기 위한 방안을 제시하였다. 콘텐츠의 디지털화는 사용의 용이성, 운반의 용이성, 시공간의 제약을 받지 않는 특징으로 사용자나 배포자가 선호하고 있는 방법이며, 대부분의 콘텐츠를 디지털화 방법으로 제작 보급되고 있다. 디지털 콘텐츠의 활발한 보급과 토렌트의 기술이 공존하는 방법에 대한 연구는 계속 수행되어야 하는 과제로 판단된다.

참 고 문 헌

[1] 이정현, 안관준, 박원형, 임종인, “익명네트워크를 이용한 사이버공격에 대한 대응방안 연구”, 한국사이버테러정보전 학회, 2011.6
[2] BitTorrent Protocol Specification v1.0, http://wiki.theory.org/BitTorrentSpecification
[3] NMC consulting group, "Netmanias 기술문서:BitTorrent 프로토콜의 동작원리“, 2011.5
[4] “2012 저작권 보호 연차보고서”, 한국저작권단체연합회 저작권보호센터, 2012.5
[5] http://www.torproject.org
[6] http://www.bittorrent.org/introduction.html
[7] http://netfpga.org/main.html

저 자 소 개

이규대(Kyu-Tae Lee)

1984 고려대 전자공학과 졸업
1986 고려대 전자공학과 석사
1991 고려대 전자공학과 박사
2001 미 죠지아텍 교환 교수
2006 미 일리노이주립대 교환 교수
2007~2009: 한국전자통신연구원 이동통신연구소 초빙연구원
1992.3~현재 : 공주대 정보통신공학부 교수

<관심분야> 회로 및 시스템, 신호처리, 콘텐츠 저작권보호기술