토렌트 프로토콜의 응용 별 분석

권재범, 유종현, 심규석, 김명섭

고려대학교 컴퓨터정보학과

{lhmagic , dididaboa , kusuk007, tmskim}@korea.ac.kr

요 약

최근 토렌트 사용량의 증가에 따라 토렌트 트래픽은 전체 트래픽의 상당한 부분을 차지하고 있다. 토렌트 트래픽은 대부분 대용량 파일을 전송하기 때문에 전체 네트워크에 많은 부담을 준다. 이러한 부담은 네트워크의 가용성을 감소시키고, 데이터 전송의 지연을 초래한다. 때문에 토렌트 트래픽에 대한 분석이 필요하다. 하지만 토렌트 프로토콜을 사용하는 응용 마다 패킷의 유형이 다르기 때문에 여러 종류의 토렌트 응용프로그램에서 발생하는 트래픽을 탐지할 수 있어야 한다. 본 논문에서는 여러 토렌트 응용을 탐지하는 방법론을 제안한다. 그리고 실험을 통하여 본 논문에서 제안한 방법론의 성능을 입증한다.

1. 서론

오늘날 네트워크에서는 여러 네트워크 응용의 등장으로 인해 트래픽 사용량이 급속도로 증가하고 있다. 네트워크의 자원은 한정되어 있는데, 토렌트 트래픽이 상당한 양을 차지하고 있어 우선시 되는 일의 효율성을 낮추고 있다. 영국의 시장조사기관 인비져널(Envisional)의 2011 년도 조사 결과에 따르면 토렌트 트래픽이 전 세계 인터넷 트래픽의 17.9%를 차지하고 있다[1]. 이 때문에 토렌트 트래픽을 차단하거나 우선순위를 낮추어서 업무에 지장을 주지 않는 네트워크 관리 정책 수립이 요구된다.

최근 많은 토렌트 프로토콜 응용 프로그램이 서비스 되고 있다. 이러한 응용들은 동일한 토렌트 프로토콜을 사용하지만 트래픽 발생량과 유형이 서로 다르다. 때문에 네트워크 관리 정책의 적용을 위해서는 토렌트 트래픽을 기존의 토렌트 프로토콜 단위의 분류가 아닌, 응용프로그램 별로 분석하는 방법이 필요하다. 따라서 본 논문에서는 토렌트 트래픽을 응용별로 분류하는 방법을 제안한다.

본 논문의 구성은 서론에 이어 본론에서는 토렌트의 동작 원리에 대해 언급하고, 토렌트를 응용 프로그램 별로 분류하는 방법을 제안한다. 또한, 분류되지 않는 토렌트 트래픽에 대한 분류 방법을 제시한다. 마지막으로 결론 및 향후 연구를 언급한다.

2. 본론

기존 Torrent 에 대한 연구는 모든 토렌트 트래픽을 응용프로그램별로 구분 하지 않았다. 하지만 토렌트 종류마다 트래픽 발생량에 차이가 있고, 유형이 다르기 때문에 응용별로 토렌트 트래픽을 분류하는 방법론이 요구된다. 따라서 토렌트 응용 별 분류 방법론을 제안한다.

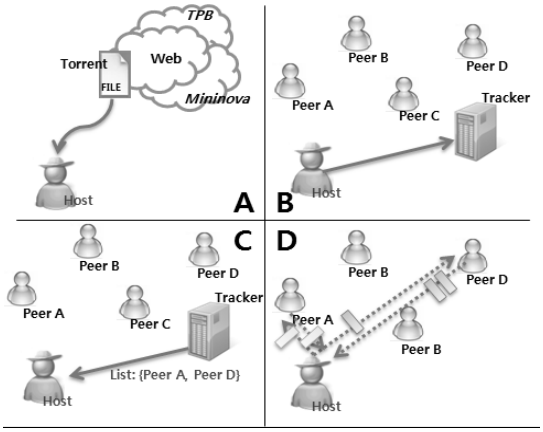

그림 1. Torrent 동작 과정 모식도

그림 1 은 토렌트의 동작 과정 모식도이다[2]. 파일 배포자는 파일의 정보를 담은 torrent 파일을 생성하고, tracker 와 사람들이 사용할 수 있도록 TPB[3] 또는 Mininova[4] 와 같이 잘 알려진 웹사이트에 등록한다. 등록된 파일에 관심을 갖고 있는 호스트는 해당 웹사이트로부터 torrent 파일을 받는다. 그림 1 의 A 에서 해당 호스트는 torrent 파일에 포함되어 있는 tracker 의 주소로 그림 1 의 B 와 같이 접속을 한다. 이 때, tracker 는 그림 1 의 C 에서 처럼 자신이 관리하고 있는 해당 파일을 공유하는 peer 들 중에서 일부의 peer 를 선택하여 그들의 주소가 담긴 리스트를 호스트에게 알려준다. 이때, 보통 50 개의 peer 정보를 호스트에게 알려준다. 마지막과정으로 호스트는 그림 1 의 D 처럼 tracker 로부터 획득한 peer 들의 주소를 바탕으로 연결을 시도하고 연결된 peer 에 한해서 자신이 원하는 파일의

이 논문은 2012 년 정부(교육과학기술부)의 재원으로 한국연구재단(2012R1A1A2007483) 및 2013 년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보.컴퓨팅기술개발사업(2010-0020728)의 지원을 받아 수행된 연구임.

각기 다른 piece 를 요청하게 된다.

본 논문에서는 토렌트 트래픽을 응용 별로 분류하기 위한 각 응용 프로그램 별 페이로드 시그니쳐를 추출하였다. 페이로드 시그니쳐란 개별 응용을 타 응용과 구분지을 수 있는 본연의 특징을 페이로드에서 추출한 것이다. 토렌트 시그니쳐 중 가장 대표적인 것은 “BitTorrent protocol”로써, 본 논문에서는 대표 시그니쳐가 포함된 패킷의 시그니쳐와, 포함되지 않은 패킷 중 TCP/UDP 프로토콜 별 시그니쳐에 대해 각각 기술한다.

표 1.대표 시그니쳐 포함 패킷의 응용 별 시그니쳐

Utorrent	BitTorrent protocol.-.14c.$3.cx….-UT341
Bittorrent	BitTorrent protocol.-.14c.$3.cx….-BT790
qBittorrent	BitTorrent protocol....t...fs…[...=..-qB3190
Azureus	BitTorrent protocol.-.4v…...cx….-AZ5300

표 1 은 대표 시그니쳐를 포함하는 토렌트 패킷의 응용 별 시그니쳐를 나타낸다. 해당 패킷들은 응용을 나타내는 특정한 단어인 UT, qB, BT 그리고 AZ 를 포함한다. 따라서 이러한 단어들은 각 응용을 대표하는 시그니쳐로 사용할 수 있다.

표 2 . TCP 패킷의 응용 별 시그니쳐

Utorrent	Host : bench.utorrent.com user-Agent : BtwebClient/3300(29544)
Bittorrent	Host : bench.bittorrent.com user-Agent : BtwebClient/7910(30739)
qBittorrent	user-Agent : qBittorrent/v3.1.9
Azureus	user-Agent : Azureus 5.3.0.0

표 2 는 TCP 패킷의 페이로드에 포함되어 있는 Host 와 user-Agent 정보이다. 해당 정보를 시그니쳐로 사용하여 토렌트 응용을 구분 할 수 있다 . Azureus 와 qBittorrent 토렌트 응용은 user-Agent 만으로 응용을 분류 할 수 있지만, Utorrent 와 Bittorrent 는 user-Agent 정보 뿐만 아니라 Host 정보를 사용하여 토렌트 응용 분류가 가능하다.

표 3. UDP 패킷의 응용 별 시그니쳐

Utorrent	34 3A 55 54 73 68 31 3A 79 31 3A 71 65 4 : U Ts h 1 : y 1 : q e
Bittorrent	34 3A 55 54 78 13 31 3A 79 31 3A 71 65 4 : U T x . 1 : y 1 : q e
qBittorrent	34 3A 4C 54 00 10 31 3A 79 31 3A 71 65 4 : L T . . 1 : y 1 : q e
Azureus	34 3A 55 54 5A C3 31 3A 79 31 3A 71 65 4 : U T Z Ã 1 : y 1 : q e

표 3 은 UDP 패킷의 페이로드에 명시 되어 있는 토렌트 트래픽의 응용 별 특징이다. 각 응용별로 나타내는 특징이 다르기 때문에 표 3 시그니쳐를 이용하여 토렌트의 응용 별 분류가 가능하다.

표 3 의 방법론은 표 1,2 보다 중요하다. 토렌트 트래픽은 UDP 트래픽을 많이 발생시키기 때문이다. 표 3 의 시그니쳐를 사용하기 전에는 토렌트 트래픽을 30%밖에 분석하지 못했다. 하지만 해당 시그니쳐를 적용한 후 70%이상의 토렌트 트래픽을 구분할 수 있었다. 40%에 가까운 토렌트 트래픽이 대표 시그니쳐를 포함하지 않은 UDP 플로우이기 때문이다.

분류할 수 없는 토렌트 트래픽도 존재한다. 이러한 트래픽은 분류 된 토렌트 트래픽의 IP 주소와 Port 정보를 비교하여 분류할 수 있다. 따라서 대부분의 토렌트 트래픽은 응용별로 분류할 수 있다.

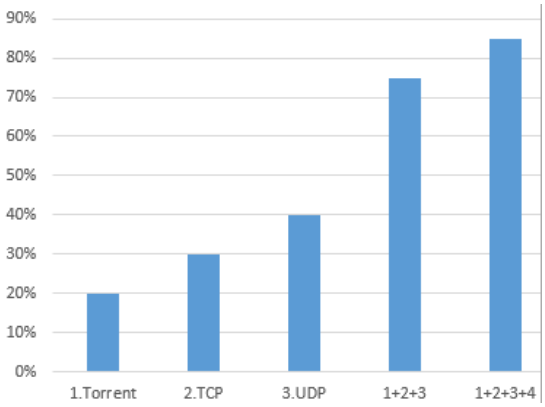

그림 2. 프로토콜에 따른 트래픽 응용 분석률

그림 2 는 방법론에 따른 트래픽 응용 분석률을 나타낸다. 1 은 대표 시그니쳐를 이용한 분류 결과이고 2 와 3 은 각각 대표 시그니쳐를 포함하지 않은 패킷 중 TCP, UDP 시그니쳐를 적용한 분류 결과이다. 마지막으로 4 는 시그니쳐로 분류할 수 없는 토렌트 트래픽을 분류된 트래픽의 IP/Port 정보와 비교하여 분석한 결과이다. 1, 2, 3, 4 를 모두 적용하여 분석한 결과 총 85%를 넘는 분석률을 보였다.

3. 결론 및 향후 연구

토렌트를 응용별로 분석하는 이유는 응용별로 토렌트 트래픽 발생량과 트래픽의 유형이 다르기 때문이다. 이는, 핵심프로토콜은 동일하나 응용 별로 사용하는 부가적인 프로토콜이 다르기 때문이다. 본 연구에서는 토렌트 트래픽의 응용 별 분류를 위해 응용 별 시그니쳐를 추출하였고, 이를 적용한 결과 전체 토렌트 트래픽의 85%를 분석할 수 있었다.

향후 연구에서는 토렌트 프로토콜의 응용 별 차이점에 대하여 연구할 것이다.

4. 참고 문헌

[1] http://news.mk.co.kr/newsRead.php?year=2011&no=769288

[2] 정태중, 한진영, 김현철, 권태경, 최양희, “Unchoked Peer 개수에 따른 BitTorrent 성능 분석, 한국통신학문논문지 Vol. 35, No.8 , pp.1197 – 1203 2010 년 8 월

[3] The pirate bay. http://thepiratebay.org

[4] Mininova. http://www.miniova.org