

BitTorrent 내 공유집단의 토폴로지 메저먼트 기법 연구

정태중, 한진영, 김승배*, 권태경**, 최양희**

서울대학교 컴퓨터 공학부

{tjchung, jyhan}@mmlab.snu.ac.kr, *sbkim@itc.kaist.ac.kr

**{tkkwon, yhchoi}@snu.ac.kr

A Study on How to Measure the Topology of Swarm in BitTorrent

Taejoong Chung, Jinyoung Han, Seungbae Kim, Ted "Taekyoung" Kwon, Yanghee Choi
School of Computer Science and Engineering, Seoul National University.

요 약

BitTorrent는 파일 공유를 목적으로 하는 Peer-to-Peer 방식의 소프트웨어로써 현재 인터넷 트래픽의 약 27~55%를 차지할 정도로 많이 사용되고 있다. 이렇게 BitTorrent가 많이 사용되는 여러 가지 이유 중 가장 큰 이유는 같은 파일을 여러 호스트들이 같이 공유하게 되는 공유집단의 효율적인 형성과정에 있다. 이러한 중요성으로 인하여 현재까지의 연구는 보다 빠르고 효과적으로 공유 하기 위한 공유집단의 형성에 초점이 맞추어져 진행되어 왔다. 하지만 이러한 공유집단의 중요성에도 불구하고, 공유집단의 구조적 특성에 대한 연구는 진행되기 어려운 점이 많았다. 이는 P2P의 특성상 각각 공유를 하는 호스트들의 커넥션 및 트래픽을 알기 힘들다는 이유에 기인하는데, 이로 인하여 실제적으로 공유집단이 어떠한 구조 및 특성을 가지며 파일을 공유하는 지에 대해서는 알려지지 않고 있다. 이에 따라 본 논문에서는 BitTorrent에서 사용되고 있는 프로토콜 중 한가지인 PEX (Peer Exchange Protocol)을 이용하여 공유집단의 토폴로지 정보를 효과적으로 수집할 수 있는 메저먼트 기법을 제안하였다. 본 논문에서 제안된 메저먼트 기법이 적용된 메저먼트 클라이언트를 서울대 캠퍼스 망에 총 16대 설치하여 2010년 4월 6일부터 5월 9일까지 총 27,391개의 토렌트에 대해 토폴로지 정보를 수집하였으며 이에 따른 결과로 총 3,010,186개의 호스트의 IP 및 이에 따른 12,693,674개의 커넥션 정보를 수집하였다. 이를 통하여 본 논문에서 제안된 PEX를 통한 공유집단 토폴로지 분석이 효과적으로 호스트 구조 정보를 수집할 수 있음을 보였으며 이는 추후에 진행될 BitTorrent 공유집단 분석의 초석이 될 것으로 기대 된다.

I. 서 론

BitTorrent[1]는 파일을 공유하기 위한 peer-to-peer(이하 P2P) 소프트웨어로써, IPOQUE에서 최근 발간한 보고서에 따른 현재 인터넷 트래픽의 약 27~55%를 차지하고 있다[2]. 이렇게 BitTorrent가 널리 사용되는 여러 가지 이유 중, 가장 큰 장점은 파일을 효과적으로 공유할 수 있고 또한 tit-for-tat으로 대변 될 수 있는 자연적으로 확장 및 축소 가능한 공유집단의 효율적인 메커니즘에 있다[3].

이러한 공유집단의 중요성으로 지금까지의 연구들은 좀 더 효과적으로 공유 집단을 형성하게끔 유도하는 알고리즘이 제시되어 왔으나[4, 5] 본질적으로 공유집단의 구조적 특성에 대한 연구는 진행되어 있지 않다. 즉 실질적으로 공유가 이루어지는 공유집단에 대한 분석이 이루어지지 않고 있는데, 이에 대한 가장 큰 원인으로 분산되어 파일을 공유하는 BitTorrent (P2P)의 특성상 개별 호스트의 커넥션정보 및 트래픽을 알기 힘들다는 문제점을 들 수가 있다.

따라서 본 논문에서는 BitTorrent 프로토콜 중 하나인 PEX (Peer Exchange Protocol)을 이용하여 BitTorrent 내에서의 공유집단의 커넥션 및 토폴로지 정보를 수집할 수 있는 메저먼트 메커니즘을 제시하고자 한다. 이는 추후에 이루어 질 수 있는 공유집단의 구조분석 및 실제

공유집단이 어떻게 형성되는지 알 수 있는 연구에 대한 초석이 될 것으로 기대된다.

II. 메저먼트 기법 설계

2.1 PEX 프로토콜

BitTorrent에서 원하는 파일을 가지고 있는 호스트들을 찾는 방법에는 (1) Tracker을 이용하는 방법, (2) DHT (Distributed Hash Table)를 이용하는 방법 그리고 (3) PEX (Peer Exchange Protocol)을 이용하는 방법이 있다. 본 논문에서는 공간 제약상 (3)번만 설명하고자 한다.

각 호스트들의 유기적인 공유집단의 형성을 위하여 BitTorrent는 커넥션을 맺은 호스트 (Peer) 끼리, 자신이 맺고 있는 모든 호스트들의 IP 주소 정보를 교환하는 (Exchange) 프로토콜을 이용하며 이를 PEX (Peer Exchange Protocol)이라고 일컫는다.

즉 그림 1에서 처럼, 커넥션을 맺은 (가), (나) 호스트에 대해 (가) 호스트는 자신과 연결되어 있던 호스트 a, b, c, d 주소 정보를, 그리고 (나) 호스트는 자신과 연결되어 있던 호스트 e, f, g, h의 주소정보를 서로 교환하게 된다.

이에 따른 메커니즘으로 PEX는 좀 더 넓은 공유집단의 형성을 가능하게 할 뿐만 아니라, 현재

커넥션을 맺고 있는 호스트보다 더 나은 환경을 가진 호스트들과 연결 할 수 있는 기회도 제공하여 준다.

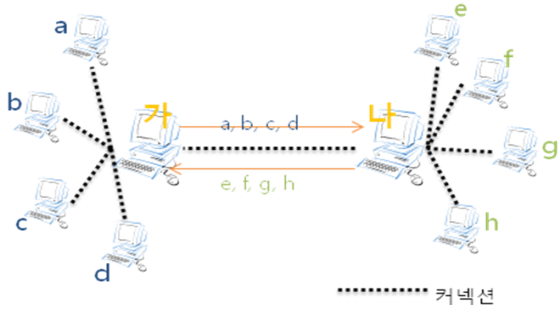


그림 1. PEX (Peer Exchange Protocol) 개념도

2.2 메저먼트 기법

연결된 호스트끼리 자신과 연결된 또 다른 호스트들의 주소를 주고받는 PEX 를 이용하면 공유집단의 토폴로지를 추측 할 수가 있다. 예를 들어 호스트 a, b, c, d, e 가 그림 2 의 (가) 와 같이 토폴로지 형태를 가지고 있다고 가정하자. 이 때 메저먼트 클라이언트 M 은 Tracker 를 통해 미리 알고 있는 호스트 a 와 커넥션을 맺은 후, a 가 현재 맺고 있는 커넥션 정보 (b, c)를 PEX 를 통해 수집 하게 된다. 이에 따라 현재 메저먼트 클라이언트 M 이 수집한 토폴로지 라우팅 정보 또한 그림 2 에 기록되어 있다. (그림 2 의 가의 테이블). PEX 를 하기 전 메저먼트 클라이언트 M 이 가지고 있는 정보는 호스트 a 에 대한 주소 정보만 있음을 주의하자.

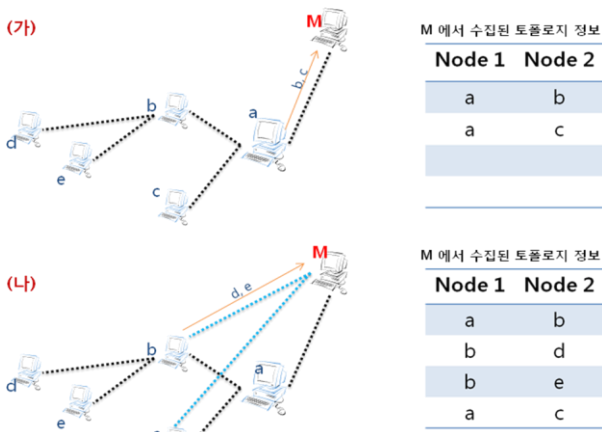


그림 2. PEX 를 이용한 토폴로지 메저먼트 시나리오와 이에 따른 메저먼트 라우팅 테이블 엔트리 정보

메저먼트 클라이언트 M 은 a 와의 PEX 를 통하여 알게 된 호스트 b 와 c 의 주소를 이용하여 호스트 b 와 c 에게 커넥션을 맺게 되고, 이에 따라 또 다시 b 와의 PEX 를 통하여 호스트 d 와 e 의 주소 또한 알 수 있게 된다. 이러한 방식으로 공유집단 전체에 대한 토폴로지 정보를 추측 할 수가 있는데, 처음 알게 된 하나의 호스트를 중심으로 그 호스트에 대한 모든 정보에 우선권을 두어 커넥션을 맺는 DFS (Depth First Search) 방식과, 주위의 모든 호스트 들에 대해 차례 차례 커넥션을 맺은 후 그 모든 호스트들에 대해 PEX 를 통해 토폴로지를 추측하는 BFS (Breadth First Search)방식이 있다. 하나의 예로 PEX_BFS 알고리즘이 표 1 에 기술되어 있다.

```

1 procedure PEX_BFS(Given_Address start_address):
2   create a queue Q
3   enqueue start_address onto Q
4   mark source
5   while Q is not empty:
6     dequeue an item from Q into v,
7     Make a connection and gather IP from PEX
8     for each connection e from host h:
9       let w be the other host's address from h.
10    if w is not searched from measurement:
11      save routing information of w
12    enqueue w onto Q
    
```

표 1. BFS (Breadth First Search)방식의 PEX 토폴로지 메저먼트 알고리즘

III. 결론

본 논문에서는 BitTorrent 의 프로토콜 중 하나인 PEX (Peer Exchange Protocol)을 이용하여 공유집단의 정보를 수집할 수 있는 메저먼트 기법을 제안하였다. 본 연구팀은 본 알고리즘을 적용시킨 메저먼트 클라이언트 16 대를 설치하여 대표적인 메타 토런트포털인 TPB (The Pirate Bay)[6]에서 발행된 총 27,391 개의 토런트에 대하여 2010 년 4 월 6 일 부터 2010 년 5 월 9 일까지 약 35 일의 기간 동안 공유집단 정보를 수집하였다. 이에 따라 총 3,010,186 개의 호스트의 IP 와 이들에 대한 12,693,674 개의 커넥션 정보를 성공적으로 수집하였다.

이를 통하여 제안된 메저먼트 기법이 효과적으로 공유집단 정보를 수집할 수 있음을 보였으며 본 연구팀에서는 수집된 토폴로지 정보를 바탕으로 공유집단의 구조적 특성에 대한 연구를 진행 할 예정이다.

ACKNOWLEDGMENT

본 연구는 한국방송통신전파진흥원의 차세대 통신 네트워크원천기술개발사업 (11-913-05-002: 이름 주소 기반 네트워킹을 위한 내재 캐칭 및 라우팅 원천기술 연구)의 일환으로 수행되었음.

참 고 문 헌

[1] B. Cohen, " Incentives build robustness in bittorrent." 1st Workshop on Economics of p2p Systems, 2003.

[2] " The impact of p2p file sharing, voice over ip, instant messaging, oneclick hosting and media streaming on the internet." <http://www.ipoque.com/resources/internetstudies/internet-study-2008-2009>

[3] Arnaud Legout et. al. "Rarest first and choke algorithms are enough", IMC' 06 Proceedings of the 6th ACM SIGCOMM conference on Internet Measurement.

[4] Shansi Ren et. al. TopBT: A Topology-Aware and Infrastructure-Independent BitTorrent Client

[5] D. R. Choffnes and F. E. Bustamante, "Taming the torrent: a practical approach to reducing cross-isp traffic in peer-to-peer systems", ACM SIGCOMM, 2008

[6] "The Pirate Bay", <http://www.thepiratebay.org>.