서버 캐시 기반 응용 트래픽 분류 방법론의 성능 향상

박준상, 김명섭
고려대학교

{junsang_park, tmskim}@korea.ac.kr

Performance Improvement of Server Cache-based Application-level Traffic Classification Method

Jun-Sang Park, Myung-Sup Kim
Korea Univ.

요 약

응용 레벨 트래픽 분류에 있어서 페이로드 시그니쳐 기반 응용 레벨 트래픽 분류 방법은 고속 링크의 대용량의 트래픽의 분석 과정에서 높은 부하를 발생시키며 처리 속도가 느린 단점을 갖는다. 이러한 문제를 해결하기 위해서 서버 IP, Port 캐시 기반 트래픽 분석 시스템이 제시되었다. 하지만 캐시의 관리 방법에 따라서 분류 시스템의 처리 속도와 분석률에 영향을 미친다. 따라서 본 논문에서는 분석 시스템의 처리 속도와 분석률을 향상시킬 수 있는 캐시 관리 기법을 제안한다. 제안하는 방법을 학내 망의 실제 트래픽 분석에 적용하여 그 타당성을 증명한다.

Ⅰ. 서 론

네트워크의 효율적 운용과 관리를 위한 응용 레벨의 트래픽의 모니터링과 분석은 네트워크 사용현황 파악과 확장계획 수립 등의 다양한 분야에서 필요성이 커져가고 있다. 이를 위해서는 다양한 종류의 응용 레벨 트래픽을 정확하게 분류할 수 있는 방법과 고속 링크에서 발생하는 대용량의 트래픽을 실시간으로 처리하는 방법이 요구된다.

응용 레벨 트래픽 분류 방법에 있어 페이로드 시그니쳐 기반 분석 방법은 패킷의 헤더 정보나 통계 정보를 이용하는 다른 분석 방법들에 비해 상대적으로 높은 분류 정확성과 분석률을 보인다[1,2]. 하지만 분류 시스템의 처리 속도에 있어 현재의 고속 네트워크 상에서 발생하는 대용량 트래픽을 실시간으로 처리하기에 부적합한 방법이다. 응용의 수와 대용량의 트래픽을 발생시키는 응용의 사용이 증가하고 있는 추세를 고려했을 때 페이로드 기반 분석 방법의 처리 속도 문제는 반드시 해결되어야 하는 과제이다. 이를 해결하기 위해 기존의 연구에서는 페이로드 시그니쳐 기반 분석 방법론으로 분석된 플로우의 서버 IP, Port 정보를 캐싱하여 이를 기반으로 분석하는 방법론을 제안하였다. 하지만 이러한 방법은 캐시에 유지되는 데이터의 관리 방법에 따라 분석률, 분류 정확도, 분석 시간 등의 분류 시스템의 성능이 결정되기 때문에 캐시 데이터의 관리 방법에 대한 연구가 선행되어야 한다. 따라서 본 논문에서는 서버 IP, Port 캐시의 관리 방법을 제시하여 분류 시스템의 성능을 향상 시킨다.

* 이 논문은 정부(교육과학기술부)의 재원으로 2010 년도 한국연구재단-차세대정보컴퓨팅기술개발사업(20100020728) 및 2012 년도 한국연구재단(2012R1A1A2007483)의 지원을 받아 수행된 연구임

본 논문의 구성은 다음과 같다. 본 장의 서론에 이어, 2 장에서는 관련연구에 대해 기술하고, 3 장에서는 실험 결과를 바탕으로 제안하는 방법을 기술한다. 4 장에서는 제안하는 방법을 분류 시스템에 적용하여 그 타당성을 증명한다. 마지막으로 5 장에서는 결론 및 향후 연구에 대해 기술한다.

Ⅱ. 관련연구

서버 IP, Port 캐시 기반 트래픽 분류 방법론의 타당성을 확인하기 위해서 학내망에서 하루 동안 발생하는 전체 TCP 트래픽을 대상으로 동일한 서버 IP, Port 의 분포를 확인하였다.

TCP 전체 플로우 (850,000) 의 80%가 10,000 개 이하의 SSIP 로 접속하는 것을 확인할 수 있었다. 서버의 IP, Port 는 응용 프로그램의 특정 서비스를 연결하는 주소로 사용된다. 따라서 10,000 개 이하의 서버 IP, Port 가 제공하는 응용의 분류 결과를 분석 시스템에서 유지할 수 있으면 동일한 서버 IP, Port 로 접속하는 플로우를 페이로드 시그니쳐 기반 분석없이 식별할 수 있기 때문에 페이로드 시그니쳐 기반 분석 방법의 처리 속도를 크게 향상 시킬 수 있다.

M. Baldi.는 서버 IP, Port 캐시 기반 트래픽 분석 방법론을 통해 분류 시간을 향상 시킬 수 있는 방법론을 제시하였다[1]. 서버 캐시를 통해 처리 시간을 향상 시킬 수 있었지만 캐시 데이터의 관리 방법을 제시하지 못하고 있다. 캐시에 영구적으로 정보를 저장하고 분석하면 탐색하는 정보의 양이 증가하여 처리 속도가 늦어지는 문제점이 발생한다. 반대로 캐시 데이터를 일정한 주기로 삭제하는 방법은 탐색 속도 문제를 해결할 수 있지만 캐시에 의해서 분석될 수 있는 트래픽을 분석하지 못하는 문제점이 발생한다.

따라서 본 논문에서는 서버 IP, Port 캐시 기반 분석 방법론의 분석 시간, 분석률을 고려한 캐시 관리 방법론을 제시한다.

Ⅲ. 제안하는 방법

본 장에서는 서버 IP, Port 캐시 기반 트래픽 분석 방법론의 처리 속도와 분석률을 향상 시킬 수 있는 캐시 관리 방법론 제시하며 실험적으로 증명한다.

P2P 와 같은 응용은 다양의 플로우를 발생 시키며 Peer 의 서버 IP, Port 정보가 캐시에 저장되기 때문에 캐시 검색 양을 증가 시켜서 분석 속도를 저하 시킨다. 또한 1 개의 서버 IP, Port 에서 제공하는 여러 가지의 응용 프로그램의 기능 중 페이로드 시그니쳐가 추출되지 않은 기능에 대해서도 해당 서버 IP, Port 로 분석될 수 있기 때문에 무작정 캐시 데이터를 삭제하는 방법은 분류 시스템의 분석률을 감소 시킬 수 있다. 따라서 처리속도와 분석률을 향상 시킬 수 있는 캐시 관리 방법이 요구된다.

캐시 데이터 관리 방안으로는 캐시에 데이터가 등록되는 시간(RT)과 마지막으로 분류에 사용된 시간(RUT)을 기준으로 적용하는 방법을 실험적으로 평가하였다. 이때 캐시 데이터의 유지 시간은 선행 연구를 통해 증명한 3 시간으로 고정하여 적용한다[4].

그림 1 은 등록 시간 기준 방법과 최근 사용 시간 기준 방법을 적용하여 플로우 단위 분석률을 측정한 결과이다.

최근 사용 시간 기준 삭제 방법이 등록 시간 기준 삭제 방법 보다 더 높은 분석률을 나타내는 것을 알 수 있다. 이는 등록 시간 기준 기반 삭제 방법은 최근 사용 유무를 반영하지 않고 캐시에서 제거되기 때문에 제거된 서버 IP, Port 정보로 분석할 수 있는 트래픽을 분석하지 못하고, 페이로드 시그니쳐 기반으로 분석된 후 캐시에 재등록되기 때문이다.

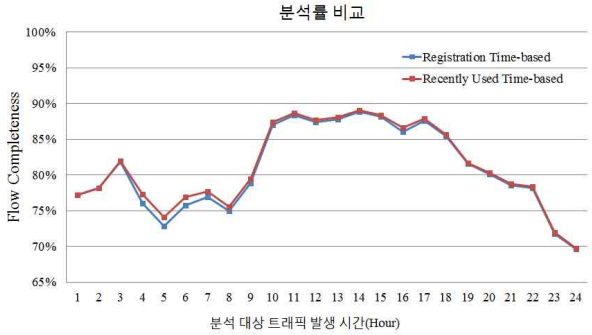

그림 1. RT vs. RUT 분석률 비교

그림 2 는 등록 시간 기준 방법과 최근 사용 시간 기준 방법을 적용하였을 때 캐시에 삽입, 제거되는 횟수의 합을 1 시간 단위로 측정한 결과이다. 캐시의 교체 횟수는 분류 시스템의 분석 시간에 영향을 미치는 요소로 작용한다.

RT 와 RUT 기반 캐시 관리 방법에 등록된 레코드의 평균 개수는 각각 73,743, 53,214 로 RT 기반 관리 방법이 평균적으로 20,000 여개의 서버 IP, Port 정보를 더 많이 유지하고 있는 것을 알 수 있다.

분류 시스템이 실행된 직후부터 캐시 데이터 유지 시간인 3 시간 동안은 캐시에 데이터 삽입만 수행되기 때문에 교체 횟수가 동일하게 나타나지만 3 시간 이후 RUT 의 교체 횟수가 RT 의 교체 횟수보다 적은 것을 알 수 있다. 이는 RT 가 3 시간이 지난 캐시 정보를 모두 제거하고, 페이로드 시그니쳐 분석 후 재등록하는 과정을 반복하는 반면에, RUT 는 최근 3 시간 내에 분석에 사용된 캐시 정보는 교체 작업을 수행하지 않기 때문이다. 2 장의 관련연구에서 언급하였듯이 학내망에서 발생하는 TCP 플로우의 80%는 10,000 개 이하의 동일한 서버 IP, Port 로 접속하기 때문에 RUT 기반 캐시 관리 방법을 적용하면 캐시의 교체 작업없이 자주 사용되는 서버 IP, Port 정보를 지속적으로 유지할 수 있다.

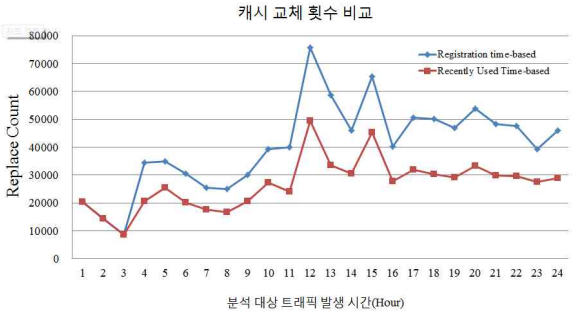

그림 2. RT vs. RUT 캐시 교체 횟수 비교

분석률과 분류 시스템의 분석 시간을 고려하였을 때 최근 분석에 사용된 여부를 기준으로 캐쉬를 관리하는 방법이 효과적임을 알 수 있다.

Ⅴ. 결론 및 향후 과제

본 논문에서는 페이로드 시그니쳐 기반 응용 레벨 트래픽 분류 시스템의 처리 속도 향상을 위해서 제안된 서버 IP, Port 기반 트래픽 분석 방법의 분석률과 분류 속도를 향상 시키기 위해서 최근 사용 시간 (RUT) 기반 캐시 관리 방법을 제시하였다. 등록 시간(RT) 기반 관리 방법에 비해 높은 분석률과 적은 캐시 교체 횟수를 보이는 것을 확인할 수 있었다.

본 연구에서는 모든 응용에 대하여 일괄적인 캐시 교체 정책을 적용하였다. 향후 연구로는 응용의 트래픽 발생 특징을 반영하여 응용 별 캐시 관리 방법에 대하여 연구를 진행할 계획이다.

참 고 문 헌

[1]M. Baldi, A. Baldini, N. Cascarano, and F. Risso. “Service-based traffic classification: Principles and validation”, In Sarnoff Symposium, SARNOFF'09. IEEE, 2009 pp. 1-6.

[2]F. Risso, M. Baldi, O. Morandi, A. Baldini, and P. Monclus, "Lightweight, Payload-Based Traffic Classification An Experimental Evaluation," IEEE International Conference on Communications, Beijing, China, May. 19-23, 2008, pp. 5869-5875.

[3]Sung-Ho Yoon, Jin-Wan Park, Young-Seok Oh, Jun-Sang Park, and Myung-Sup Kim, "Internet Application Traffic Classification Using Fixed IP-port," APNOMS 2009, LNCS, Jeju, Korea, Sep. 23-25, 2009, pp. 21-30.

[4]박준상, 윤성호, 박태영, 김명섭, “서버 캐시 기반 트래픽 분류 방법에 관한 연구”2013 년도 한국통신학회 KNOM 학술발표회, 계명대학교, 대구, May. 9-10, pp.xx-xx