응용 레벨 트래픽 분류를 위한 시그니쳐 관리 시스템 개발

박준상, 김명섭
고려대학교 컴퓨터정보학과
e-mail : {junsang_park,tmskim}@korea.ac.kr

Development of Signature Management System for Application-level Traffic Classification

Jun-Sang Park, Myung-Sup Kim
Dept. of Computer and Information Science, Korea University

요 약

응용 레벨 트래픽 분류를 위한 다양한 방법 중 페이로드 시그니쳐 기반 분석 방법은 높은 정확성과 분석률을 보인다. 하지만 현재의 인터넷 기반의 응용 프로그램은 사용자의 요구사항을 만족시키고 안정적인 서비스를 제공하기 위해 빠른 속도로 변화하고 있어서 지속적으로 높은 분류 성능을 보장할 수 없다. 따라서 본 논문에서는 페이로드 시그니쳐 기반의 분석 방법을 기반으로 응용 프로그램의 변화, 출현에 유연하게 대처 가능한 시그니쳐 관리 시스템을 제안한다. 또한 시그니쳐 관리 시스템을 학내망에 적용하고 실시간으로 트래픽을 분석하여 그 타당성을 증명한다.

1. 서론

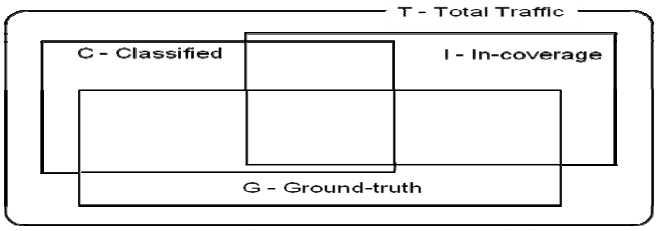
기존의 다양한 연구[2,3,4]에서 페이로드 시그니쳐 기반의 응용 레벨 트래픽 분석 방법을 제시하고 그 타당성을 증명하고 있지만, 시그니쳐의 정확성을 지속적으로 유지하기 위한 방법을 제시하지 못하고 있다. 응용 프로그램에 대한 사용자의 기호 변화와 데이터 전송 기술의 발전은 응용프로그램의 변화, 새로운 응용의 출현으로 이어지고 응용 프로그램 시그니쳐의 정확성을 감소시킨다.[1] 따라서 본 논문에서는 응용 프로그램 시그니쳐의 정확도를 지속적으로 유지하기 위한 시그니쳐 관리 시스템을 제시하고 그 타당성을 증명한다.

본 논문은 다음과 같은 순서로 구성된다. 2 장에서는 문제의 범위를 명확하게 정의하고, 3 장에서는 시그니쳐 관리 시스템의 구조 및 방법에 대해 기술한다. 4 장에서는 관리 시스템의 적용 결과에 대해 설명하고, 5 장에서는 결론을 맺고 향후 연구에 대하여 기술한다.

2. 문제의 범위

우리는 선행 연구[5]에서 응용프로그램 시그니쳐 자동 생성 시스템을 제안하고 이를 기반으로 94 개의 응용 프로그램에 대해 시그니쳐를 생성하여 실시간으로 학내망의 응용 레벨 트래픽을 분석하였다. 그 결과 95%의 이상의 정확도와 전체 트래픽의 75% 이상의 분석률을 보였지만, 시간이 지남에 따라 정확도와 분석률이 감소하는 현상을 보였다. 정확도와 분석률의 감소 원인은 응용의 변화에 따른 시그니쳐의 충돌, 새로운 응용의 출현으로 분석되었다.

그림 1 은 분류 대상 전체 트래픽을 분류한 트래픽, 분류 가능한 트래픽, Groud Truth 를 갖는 트래픽으로 구분하고, 시그니쳐 관리 시스템의 적용 범위를 표현하고 있다.



(그림 1) 관리 시스템의 적용 범위

시그니쳐 관리 시스템의 목적은 시그니쳐의 유효성 판단에 따른 시그니쳐 수정, Groud Truth 에 나타나는 새로운 응용과 알 수 없는 응용에 대한 시그니쳐 생성으로 나누어진다.

$C \cap G$: 시그니쳐의 유효성 판단
$G-(C \cup I)$: 새로운 응용에 대한 보고
$T-(C \cup G \cup I)$: 알 수 없는 응용에 대한 보고

[1] 이 논문은 2007 년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임.(KRF-2007-331-D00387)

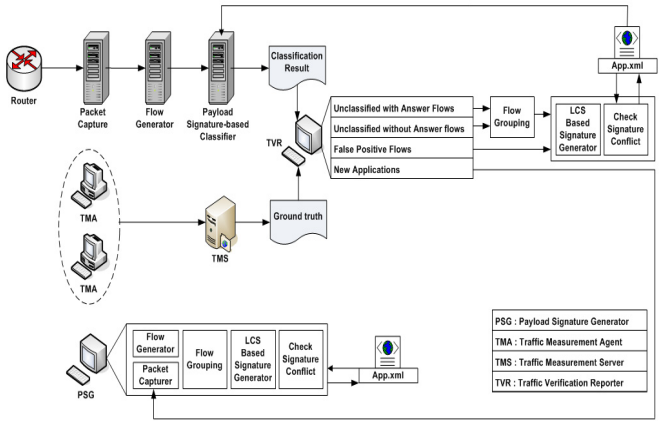
3. 시그니쳐 관리 시스템 구조 및 방법

본 장에서는 응용 프로그램 시그니쳐 관리 시스템의 구조 및 방법에 대해 기술한다.

시그니쳐의 관리를 위해 학내망에서 발생하는 모든 트래픽을 대상으로 실시간 검증 시스템을 구축하였다. 검증 시스템은 TMA 를 기반으로 응용 프로그램 단위로 결과를 추출하며, 검증된 결과를 바탕으로 오분류되거나 미분류되어진 트래픽을 수집한다.

관리 시스템의 구조는 그림 2 와 같이 트래픽 수집 시스템, 페이로드 기반 트래픽 분류 시스템, TMA (Traffic Measurement Agent), TMS(Traffic Measurement Server), TVR(Traffic Verification Reporter), 시그니쳐 추출 시스템으로 구성된다.

시그니쳐를 기반으로 분류된 결과는 관리자에게 Accuracy, Completeness, Coverage 를 제공하며, 오분류되거나 미분류되어진 플로우를 수집하여 시그니쳐 관리 시스템의 입력 데이터로 제공하여 플로우 그룹핑 단계, LCS 기반의 시그니쳐 추출 단계, 시그니쳐 유효성 평가 단계를 거쳐 응용프로그램의 시그니쳐를 갱신한다. 또한 TMA 를 기반으로 수집된 응용의 정보에 기초하여 새로운 응용의 출현에 대한 리포팅을 통해 시그니쳐 추출 대상 응용프로그램을 선정하고 시그니쳐 생성 시스템을 기반으로 시그니쳐 생성하고 등록하게 된다.



(그림 2) 검증 및 관리 시스템 구조

4. 결과

실험 결과는 2009 년 2 월 1 일을 기준으로 시그니쳐 관리 시스템의 적용 전과 적용 후 현재까지의 Coverage, Accuracy, Completeness 를 나타낸다.

표 1 은 관리 시스템의 적용 전후 Coverage 결과를 나타낸다. 관리 시스템에 의해서 새로운 응용을 추가한 결과 응용 프로그램의 수가 2 배 이상 증가하였다.

<표 1> 관리 시스템 적용 전후 Coverage 결과

분류 단위	적용 전	현재
응용프로그램	81	175
프로토콜	13	13
프로세스	153	260
시그니쳐	537	823

표 2 는 관리 시스템의 적용 후 Accuracy, Completeness 의 변화를 나타낸다. 응용프로그램의 Coverage 는 2 배 이상 증가하였지만 Completeness 는 5%~6%만의 증가율을 보였다. 이는 다양한 응용이 존재하고, 사용자의 기호에 따라 서로 다른 응용을 사용하기 때문인 것으로 분석되었다. FP, FN 으로 발생하는 플로우에 대해 시그니쳐 관리 시스템 적용한 결과 시그니쳐의 정확도가 향상되어 99% 이상의 높은 Accuracy 를 보이는 것을 확인할 수 있다.

<표 2> 관리 시스템 적용 전후 분류 성능

검증 단위		적용 전	현재
Completeness	Flow	86.06%	93.49%
	Byte	76.75%	81.19%
	Packet	75.72%	82.97%
Accuracy	Flow	96.63%	99.45%
	Byte	98.35%	99.77%
	Packet	98.34%	99.88%

5. 결론 및 향후 연구

본 논문에서는 응용 프로그램의 변화와 새로운 응용의 출현에 대해 효율적인 시그니쳐 관리가 가능하도록 페이로드 시그니쳐 관리 시스템을 구축하고, 그 실효성을 증명하였다. 관리 시스템은 시간의 흐름과 공간적인 차이에 따라 다른 분류 결과를 보이는 분류 방법의 문제점을 해결할 수 있을 것으로 기대된다.

관리 시스템의 적용 결과 응용의 수와 시그니쳐의 수가 증가되는 것을 알 수 있다. 이는 분류 시스템의 과부하를 초래하기 때문에 사용되지 않는 응용과 시그니쳐를 관리하는 방법에 대한 연구가 요구된다.

참고문헌

[1] Wei Li, Marco Canini, Andrew W, Moore, Raffaele Bolla, Efficient application identification and the temporal and spatial stability of classification schema, Computer Networks, vol. 53, no. 4, pp 181-192, Apr 2009

[2] Subhabrata Sen , Oliver Spatscheck , Dongmei Wang, Accurate, scalable in-network identification of p2p traffic using application signatures, World Wide Web 2004, New York, NY, USA, May 17-20, 2004

[3] Byung-Chul Park, Young J. Won, Myung-Sup Kim, James W. Hong, "Towards Automated Application Signature Generation for Traffic Identification", NOMS 2008, Salvador, Bahia, Brazil, Apri. 7-11, 2008, 160-167

[4] Patrick Haffner , Subhabrata Sen , Oliver Spatscheck , Dongmei Wang, ACAS: automated construction of application signatures, Proceeding of the 2005 ACM SIGCOMM, Philadelphia, Pennsylvania, USA August 26-26, 2005.

[5] 박준상, 박진완, 윤성호, 오영석, 김명섭, "응용 레벨 트래픽 분류를 위한 시그니쳐 생성 시스템 및 검증 네트워크의 개발", 정보처리학회 춘계학술발표대회 (KIPS), 부산, 한화리조트, Apr. 23-24, 2009, 제 16 권 제 1 호, pp. 1288-1291.