실시간 네트워크 관리를 위한 페이로드 시그니쳐 관리 시스템

심규석, 윤성호, 김명섭

고려대학교

{kusuk007, sungho_yoon, tmskim}@korea.ac.kr

The Payload Signature Management System for Network Management on Real-Time

Kyu-Seok Shim, Sung-Ho Yoon, Myung-Sup Kim

Korea Univ.

요 약

오늘날 네트워크를 사용하는 다양한 응용이 생산되고 있으며 기존 응용 또한 암호화 및 트래픽 발생 패턴 변화로 인해 응용 트래픽을 탐지할 수 있는 기존 시그니쳐가 무의미해 지면서 네트워크 관리에 어려운 점이 있다. 따라서 네트워크 관리자는 네트워크 관리 원활을 위해 지속적으로 응용에 대한 시그니쳐를 추출하고, 적용해야만 한다. 하지만 이와 같은 작업은 매우 어렵고, 시간 비용이 많이 소비되는 작업이다. 따라서 본 논문에서는 실시간에서 작동하며, 응용에 대한 페이로드 시그니쳐를 자동으로 업데이트 해주는 실시간 시그니쳐 관리 시스템을 제안한다. 본 시스템은 시그니쳐 자동 생성 시스템을 이용하여 기존 시그니쳐에서 무의미한 시그니쳐를 삭제하고, 새로 의미 있는 시그니쳐를 추가함으로써 네트워크 관리를 원활하게 한다. 또한, 본 시스템은 실시간에 적용하여 지속적인 시그니쳐 관리가 가능하다.

Ⅰ. 서 론

오늘날 네트워크를 사용하는 다양한 응용이 생산되고 있으며 기존 응용은 암호화 및 트래픽 발생 패턴 변화되어가고 있다. 네트워크 관리를 위해서는 이러한 응용 및 서비스의 발생 패턴을 인지하여 탐지할 수 있어야한다. 따라서 응용 및 서비스를 탐지할 수 있는 고유한 패턴인 시그니쳐는 필수적이다. 기존 네트워크 관리자가 각 응용 및 서비스에 대한 시그니쳐를 보유하고 있다 하더라도, 현재 각 응용 및 서비스를 잘 탐지할 수 있는 옳은 시그니쳐인지는 보장할 수 없다. 따라서 시그니쳐에 대한 검증 및 업데이트는 매우 중요한 작업이다.

하지만 시그니쳐 추출은 쉬운 작업이 아니다. 해당 응용 및 서비스의 트래픽을 수집하고, 트래픽의 모든 부분을 눈으로 확인한 뒤 공통으로 발생하는 패턴을 추출해야 한다. 추출하더라도 특정 응용 및 서비스에서만 발생하는 고유한 패턴인지 검증단계도 필요하다. 이러한 작업은 매우 많은 시간을 소비하며, 시그니쳐를 추출하는 추출자에 따라서 시그니쳐도 변화할 수 있다는 단점을 가지고 있다. 또한, 눈으로 보고 추출하는 작업은 트래픽의 패턴에 대한 인지를 잘 할 수 없기 때문에 트래픽 상관관계를 필요로 하는 시그니쳐는 추출되기 힘들다.

따라서 본 논문에서는 시그니쳐 관리 시스템을 제안한다. 본 시스템은 여러 호스트에서 발생하는 다양한 응용의 트래픽을 수집하여, TMA를 이용하여 트래픽을 분류하고 분류된 트래픽을 이용하여 시그니쳐를 관리한다. 시그니쳐 관리는 기존 관리자가 보유한 시그니쳐 중 사용 불가한 시그니쳐는 삭제하고, 사용가능한 시그니쳐는 남기며, 새로운 시그니쳐를 추가하는 작업을 의미한다. 따라서 본 시스템은 실시간에서 사용가능하며 관리자는 따로 시그니쳐를 추출하는 작업을 필요로 하지 않는다.

본 논문의 구성은 본 장 서론에 이어, 2장에서 해당 시스템에 대한 관련 연구에 대해 언급하고, 3장에서 본 시스템의 방법론과 한계점에 대해 제안한다. 마지막으로 결론 및 향후연구에 대해 기술하고 본 논문을 마친다.

Ⅱ. 관련 연구

네트워크 관리자가 사용하는 시그니쳐는 트래픽의 고유한 특징을 기반으로 하는 다양한 종류로 존재한다. 시그니쳐 종류는 트래픽의 헤더정보 중 포트번호를 이용하여 트래픽을 분석하는 포트번호 기반 시그니쳐, 트래픽의 크기, 위치, 시간 등 통계적인 정보를 이용한 통계 기반 시그니쳐, 패킷의 데이터 부분인 페이로드 정보를 이용한 페이로드 기반 시그니쳐 등이 연구되고 있다.

포트 기반 시그니쳐의 경우 현재 네트워크를 사용하는 많은 응용들은 방화벽 및 IPS 장비를 통과하기 위해 포트 번호를 임의로 설정하기 때문에 더 이상 포트 기반 시그니쳐는 무의미하다. 통계기반 시그니쳐[1]는 시그니쳐를 추출하기 어려울 뿐만 아니라 정확성을 기대하기 힘들기 때문에 그에 맞는 연구가 진행되고 있다. 페이로드 시그니쳐[2]는 분석률과 정확도가 가장 높기 때문에 많은 연구가 진행되고 있지만, 시그니쳐의 추출하는 작업이 매우 힘들고, 많은 시간을 필요로 한다.

따라서 페이로드 시그니쳐 자동 추출 방법에 대한 연구가 이어지고 있다. 시그니쳐 자동 추출 방법은 패킷의 페이로드 내용을 기반으로 공통 문자열을 추출하여 시그니쳐화 하는 방법이다. 현재가지는 LCS 알고리즘[3], Smith-Waterman 알고리즘[4] 등 다양한 방법으로 연구되고 있다. 본 논문에서 사용되는 시그니쳐 자동 추출 방법은 데이터 마이닝에서 사용되는 순차 패턴 알고리즘 방법을 사용한다.

본 논문은 BK21플러스 사업+ (No. T1300572) 및 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보·컴퓨팅기술개발사업(2010-0020728)의 지원을 받아 수행된 연구임

Ⅲ. 시그니쳐 관리 시스템 구조

본 논문에서 제안하는 시스템은 실시간으로 동작이 가능한 시그니쳐 관리 시스템이다. 본 시스템은 여러 호스트에서 발생한 트래픽을 이용하여, TMA를 이용한 정답지 트래픽 수집부, 기존 시그니쳐를 이용하여 분석하는 트래픽 분석부, 분석하지 못한 트래픽을 이용하여 시그니쳐를 자동으로 생성하는 시그니쳐 자동 생성부[5], 그리고 생성된 시그니쳐를 검증 및 기존 시그니쳐를 삭제하는 시그니쳐 관리부로 구분하여 동작한다.

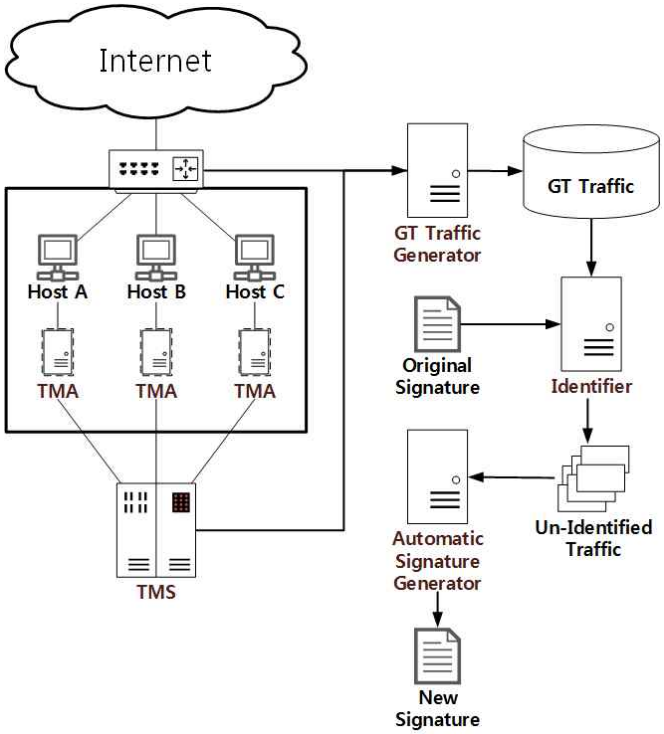

그림 1. 시그니쳐 관리 시스템 구조

TMA(Traffic Measurement Agent)를 이용한 정답지 트래픽 수집부는 TMA가 설치된 호스트에서 실행중인 프로세스들이 사용하는 소켓정보를 주기적으로 수집하여 지정된 TMS(Traffic Measurement Server)로 제공하여 정답지 트래픽을 수집할 수 있는 정보를 제공한다. TMA는 표1과 같은 정보를 제공한다. 정답 트래픽 생성기는 다음과 같은 정보를 이용하여 정답 트래픽을 생성한다. TMS에서 수집된 TMA 정보와 패킷 단위로 수집된 네트워크 트래픽을 매칭하여 정답 트래픽을 수집한다. 수집된 트래픽은 호스트 별 트래픽 파일로 저장된다.

표1. TMA 정보

Process name
IP address (local, remote)
Port number (local, remote)
State (start, continue, end, server)
Protocol
Path

트래픽 분석부는 분류된 정답지 트래픽을 기존 시그니쳐로 분석하는 역할을 한다. 관리자가 보유하고 있던 시그니쳐를 이용하여 트래픽을 분석하고, 현재 트래픽에 적합하지 않은 기존 시그니쳐는 삭제한다. 적합하지 않은 시그니쳐는 기존에는 트래픽 분석이 가능한 시그니쳐이었으나, 현재 트래픽 발생 패턴의 변화로 인해 분석이 불가능한 시그니쳐를 의미한다. 다음과 같은 과정을 거치면 현재 사용가능한 시그니쳐만 남게된다.

기존 시그니쳐로 분석되지 않은 트래픽은 분석기에서 출력된다. 출력된 트래픽은 응용 별로 구분되어 출력되며, 출력된 트래픽을 이용하여 새로운 시그니쳐를 추출한다. 새로운 시그니쳐를 자동으로 추출하기 위해 본 논문의 방법론은 순차 패턴 마이닝을 사용하여 공통 문자열을 추출할 뿐만 아니라 트래픽의 단위를 기준으로 패킷 형태의 시그니쳐, 플로우 형태의 시그니쳐를 추출한다. 본 방법은 트래픽의 상관관계를 시그니쳐에 포함하여 정확도를 향상한다. 다음과 같이 생성된 시그니쳐는 기존 시그니쳐 리스트에 추가되어 짐으로써 자동으로 시그니쳐 업데이트가 가능하다.

패킷 시그니쳐는 여러 개의 공통 문자열이 지속적으로 하나의 패킷에서 검출되면, 여러 개의 시그니쳐로 존재하는 것이 아닌 하나의 시그니쳐로 생성된다. 또한 플로우 시그니쳐는 여러 개의 패킷 시그니쳐가 하나의 플로우에서 검출되면, 여러 개의 패킷 시그니쳐로 존재하는 것이 아닌 하나의 플로우 시그니쳐로 생성된다. 다음과 같은 과정은 시그니쳐의 개수를 줄이면서 트래픽 분석 시스템의 부하를 감소시킬 뿐만 아니라, 해당 응용만을 분석하는 시그니쳐의 정확도를 향상할 수 있다.

마지막으로 시그니쳐 검증부는 기존 시그니쳐에서 사용이 불가능한 시그니쳐를 삭제하고, 새롭게 생성된 시그니쳐를 기존 시그니쳐 리스트에 추가한다. 다음과 같은 과정을 지속적으로 수행하면서 네트워크 관리자의 시그니쳐 리스트는 지속적으로 업데이트되고, 만약 기존에 없던 새로운 응용이 발견될 시 자동으로 시그니쳐 리스트가 생성된다.

Ⅳ. 결론 및 향후 연구

본 논문은 트래픽을 응용 및 서비스 별로 탐지할 수 있는 고유한 패턴인 시그니쳐를 삭제, 생성, 관리하는 시스템을 제안한다. 트래픽 발생 패턴 변화에 맞게 시그니쳐 중 분석하지 못하는 시그니쳐는 삭제하고, 새로운 시그니쳐를 자동으로 생성한다. 본 시스템에서 시그니쳐 자동 생성은 패킷 단위, 플로우 단위로 생성됨으로써 트래픽 분석 시스템의 부하를 감소시킬 뿐만 아니라, 시그니쳐의 정확도를 향상한다.

향후 본 시스템은 실시간으로 적용하여, 지속적인 시그니쳐 관리로 인해 효율적인 네트워크 관리가 가능하게 해야 한다. 또한 생성된 시그니쳐를 효율화 시켜 시스템 향상을 해야 한다.

참 고 문 헌

[1] 안현민, 함재현, 김명섭, "통계 정보 기반 트래픽 분석 방법론의 성능 향상", 정보처리학회 논문지 컴퓨터 및 통신시스템 Vol.2 No.8, Aug. 2013, pp.335-342

[2] 박준상, 윤성호, 안현민, 김명섭, "페이로드 시그니쳐 기반 인터넷 트래픽 분류", 2014년 통신망운용관리 학술대회 (KNOM 2014), 충남대학교, 대전, May. 15-16, 2014, pp.10-14.

[3] Byung-Chul Park, Young J. Won, Myung-Sup Kim, James W. Hong, "Towards Automated Application Signature Generation for Traffic Identification," Proc. of the IEEE/IFIP Network Operations and Management Symposium (NOMS) 2008, Salvador, Bahia, Brazil, Apri. 7-11, 2008, 160-167

[4] Feng, Xuepeng, et al. "Automatic traffic signature extraction based on Smith-waterman algorithm for traffic classification." Broadband Network and Multimedia Technology (IC-BNMT), 2010 3rd IEEE International Conference on. IEEE, 2010.

[5] 심규석, 윤성호, 이수강, 김성민, 정우석. "네트워크 트래픽 분석을 위한 Snort Content 규칙 자동 생성." 한국통신학회논문지 Vol.40 No.04. Apr. 2015.pp666-677