자동 생성된 페이로드 시그니쳐의 정확도 향상을 위한 연구

심규석, 윤성호, 김명섭
고려대학교

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

A Study on Accuracy Improvement of Automatically Generated Payload Signature

Kyu-Seok Shim, Sung-Ho Yoon, Myung-Sup Kim
Korea Univ.

요 약

오늘날 네트워크 기능을 사용하는 다양한 응용이 새롭게 생성되거나 기존에 있던 응용의 트래픽 패턴들이 변화하는 추세가 이어지고 있다. 따라서 네트워크 관리자 입장에서 시그니쳐에 대한 빠른 업데이트를 통해 즉각적인 대응을 하기 위해 시그니쳐 자동 생성에 대한 연구가 활발히 이루어지고 있다. 그러나, 이러한 시그니쳐 자동 생성을 통해 추출된 시그니쳐는 각 응용에 대한 정확도 향상이 필요하다. 본 논문에서는 시그니쳐 자동 생성에 대해 간략히 언급한 뒤 자동으로 추출된 시그니쳐에 대한 정확도를 향상시키기 위한 시스템을 제안한다. 본 시스템을 거친 결과로 5 개 응용에서 모두 FP(False Positive) 수치가 감소함으로써 제안하는 시스템의 성능을 증명하였다.

Ⅰ. 서 론

오늘날 네트워크 기능을 사용하는 다양한 응용들이 새롭게 생성되고 있다. 또한 기존 네트워크 기능을 사용하는 응용들도 개인정보침해 및 데이터 유실을 방지하기 위해 암호화 및 트래픽 패턴 변화가 일어나고 있다. 이러한 변화는 네트워크 관리자 입장에서 네트워크 트래픽 모니터링 및 QoS(Quality of Service) 정책 설정에 있어 작업을 힘들게 하고 있다. 이러한 변화에 대응하기 위해 네트워크 관리자는 지속적으로 응용에 대한 시그니쳐를 업데이트 해야 한다. 다양한 시그니쳐 중 페이로드 시그니쳐는 정확도가 높기 때문에 가장 널리 쓰이는 시그니쳐의 종류이다.[2] 그러나 다양한 응용의 페이로드 시그니쳐를 지속적으로 업데이트하는 것은 매우 어려운 일이다. 현재 다양한 응용 트래픽에서 시그니쳐를 추출하기 위해서는 네트워크 관리자가 많은 시간을 투자해야 하는 수동적인 방법을 수행해야 한다.

따라서 시그니쳐 자동 생성에 관한 연구[1,3]는 활발하게 이루어 지고 있다. 시그니쳐를 자동으로 생성함으로써 응용 트래픽 패턴의 변화나 새로운 응용이 발생하였을 때 즉시 파악하고, 수동적인 방법이 아니기 때문에 빠른 시간 안에 시그니쳐를 추출할 수 있기 때문이다. 본 논문에서는 기존 연구인 순차 패턴 알고리즘을 이용한 시그니쳐 자동 추출 시스템에 대해 간략히 소개한다.

그러나 응용 트래픽 시그니쳐는 정확도에 대한 검증 단계가 필요하다. 본 논문에서는 FP(False Positive) 수치를 이용한다. FP 는 특정 응용 트래픽을 제외한 트래픽을 특정 응용 시그니쳐로 분석하였을 때, 나타나는 수치이다. 특정 응용에 대한 시그니쳐가 추출되었을 때, 해당 시그니쳐가 특정 응용 트래픽만 분석하지 않고, 다른 응용 트래픽을 분석하는 것이 가능하다면, 해당 시그니쳐는 특정 응용을 분류할 수 없기 때문에 의미가 없어진다.

본 논문은 1 장 서론에 이어, 2 장 본론에서 순차 패턴 알고리즘을 이용한 시그니쳐 자동 추출 시스템에 대해 간략히 소개하고, 본 시스템에서 추출된 시그니쳐를 검증하는 방법에 대한 언급한다. 3 장 실험결과에서 추출된 시그니쳐의 정확도와 검증단계를 거친 시그니쳐의 정확도를 비교한다. 마지막 4 장에서 결론 및 향후 연구에 대해 언급하고 본 논문을 마친다.

Ⅱ. 본론

본 장에서는 기존 연구인 순차 패턴 알고리즘을 이용한 시그니쳐 자동 생성 시스템의 간략한 소개와 본 시스템에서 추출된 시그니쳐 검증 시스템에 대해 다룬다. 시그니쳐 자동 생성 시스템은 특정 응용에 대한 트래픽을 입력으로 순차 패턴 알고리즘을 이용하여 시그니쳐를 출력하는 시스템이다. 또한, 시그니쳐 검증 시스템은 추출된 시그니쳐와 특정 응용에 대한 트래픽을 입력으로 해당 시그니쳐 별로 정확도를 출력하는 시스템이다.

시그니쳐 자동 생성 시스템은 시그니쳐를 추출하기 위해 순차 패턴 알고리즘 중 Aprioriall 알고리즘을 트래픽 패턴에 맞게 수정하여 사용한다. 시그니쳐를 자동으로 생성하는 단계는 그림 1 과 같다. 먼저 트래픽의 페이로드를 추출하여 sequence 를 생성한다. 생성된 sequence 들의 집합을 사용하여 길이 1 의 content 를 생성하게 된다. 길이 1 의 content 는 최소 지지도 검사를 진행한 후, 최소 지지도 값 이상으로 sequence 에

본 논문은 BK21 플러스 사업(No. T1300573) 및 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원(B0101-15-233, 스마트 네트워킹핵심기술개발)을 받아 수행된 연구임

포함되어 있는 content 를 제외하고 삭제된다. 다음 단계로 남겨진 길이 1 의 content 는 서로 결합하여 길이 2 의 content 가 된다. 길이 2 의 content 는 다시 최소 지지도 검사를 진행한다. 이러한 단계를 지속적으로 수행하면 최종적으로 최소 지지도 값을 만족하는 공통 문자열이 추출된다.

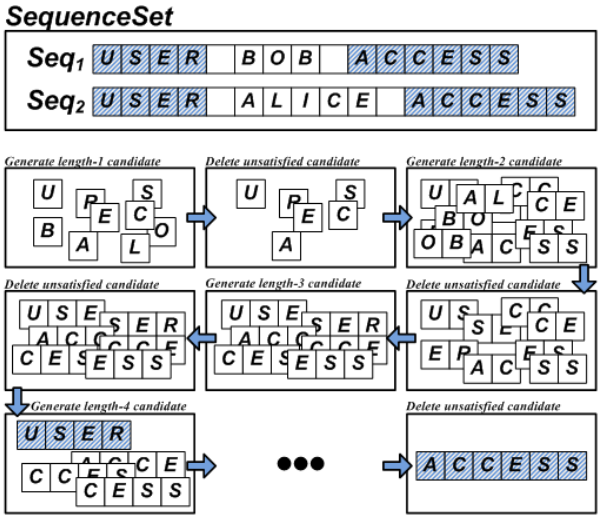

그림 1. 시그니쳐 자동 생성 과정

이러한 과정을 수행하여 추출된 시그니쳐는 검증 시스템으로 입력된다. 시그니쳐 검증 시스템은 특정 응용에 대한 시그니쳐를 추출했을 때, 추출된 시그니쳐가 해당 응용 이외에 다른 응용 또한 분석할 수 있다면, 트래픽 모니터링 및 네트워크 관리에서 의미 없는 시그니쳐이기 때문에 삭제되어야 한다. 시그니쳐 검증 시스템의 구조는 그림 2 와 같다. 먼저 수집된 트래픽을 TMA(Traffic Measurement Agent)를 이용하여, 정답지 트래픽을 생성한다.

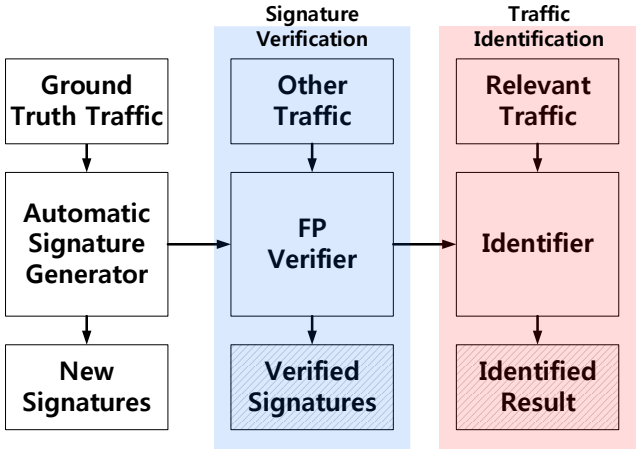

그림 2. 시그니쳐 검증 시스템 구조

다음 단계로 시그니쳐 검증 시스템은 시그니쳐 자동생성단계에서 추출된 새로운 시그니쳐와 특정 응용 트래픽을 제외한 다른 트래픽을 이용하여 FP 수치를 계산하기 위한 시그니쳐 검증부, 검증된 시그니쳐와 특정 응용 트래픽을 이용하여 트래픽 분석을 하기 위한 트래픽 분석부로 나누어 진다. 시그니쳐 검증부에서는 FP(False Positive) 수치를 각 시그니쳐 별로 계산하여 만약 FP 가 있는 시그니쳐는 본 단계에서 삭제된다. 물론 기존 TP 수치도 낮아질 수 있지만, 네트워크 관리 측면에서 분석률보다 정확도를 더 우선시 하기 때문에 TP 수치가 감소하는 것은 감안한다. 본 검증부에서 최종 시그니쳐를 출력한다. 트래픽 분석 단계에서는 분석률을 계산한다.

이러한 단계를 거쳐 최종적으로 특정 응용만 분석할 수 있는 높은 질의 정확한 시그니쳐가 선별된다. 다음의 실험 결과는 5 가지 응용을 이용한 실험결과이다. 응용은 Youtube, Naver, Utorrent, Dropbox, Facebook 을 선정하였다. 실험결과는 TP(True Positive), FP(False Positive)을 나타내어 시그니쳐의 분석률과 오탐률을 나타낸다. 표에 나타내는 값은 트래픽의 Flow 단위이다. 표 1 은 다음의 결과이다.

본 실험결과에서 모든 응용에 대해 FP 수치를 0 으로 감소시켰음에도 불구하고, TP 수치는 많이 감소하지 않았다. 특히, Dropbox 응용은 FP 수치를 0 으로 감소하였지만 TP 수치는 유지시키는 효과를 볼 수 있었다. 따라서 본 시스템을 통해 시그니쳐의 정확도를 향상시킬 수 있다.

표 1. 시그니쳐 정확도 향상 결과

	미 검증 시그니쳐			검증된 시그니쳐		
	TP	FP	개수	TP	FP	개수
Naver	2,068 /2,128	904 /3,582	542	1,929 /2,128	0 /3,582	508
Youtube	642 /645	2,458 /5,065	112	417 /645	0 /5,065	100
uTorrent	2,138 /2,613	2,591 /3,097	631	2,108 /2,613	0 /3,097	608
Dropbox	31 /51	5 /5,659	5	31 /51	0 /5,659	4
Facebook	273 /273	1,344 /5,437	37	212 /273	0 /5,437	22

Ⅲ. 결론

본 논문에서는 자동 생성된 페이로드 시그니쳐에 대해 정확도를 검증하여 향상시키는 시스템을 제안하였다. 본 시스템은 시그니쳐 생성에 많은 어려움을 해결하고자 시그니쳐를 자동으로 생성하는 연구의 문제점인 정확도 판단을 시그니쳐 검증을 통해 정확도가 높은 시그니쳐만 선별하여 정확도를 향상하였다. 본 시스템을 거친 실험 결과로 5 개 응용에서 FP(False Positive) 수치가 모두 0 으로 감소하였고, 대부분 응용의 TP(True Positive) 수치는 거의 유지시킴으로써 증명하였다.

참 고 문 헌

[1] Kyu-Seok Shim, Sung-Ho Yoon, Mi-Jung Choi, and Myung-Sup Kim, "Signature Management System to cope with Traffic Changes in Application and Service," Proc. of the Asia-Pacific Network Operations and Management Symposium (APNOMS) 2015, Busan, Korea, Aug. 19-21, 2015, pp.192-197.

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

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