# 프로토콜 필터를 이용한 페이로드 시그니쳐 자동 생성 시스템에 관한 연구

박철신, 박준상, 김명섭
고려대학교

{iandyoy, junsang_park, tmskim}@korea.ac.kr

# Study on Automatic Payload Signature Generation System using Protocol Filter

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

요 약

페이로드 시그니쳐 기반 분석 방법에서 페이로드 시그니쳐의 생성은 응용 발생에 대한 대처 및 전문인력 확보와 같은 시그니쳐 추출 오버헤드가 발생된다는 것이 단점으로 지적되고 있다. 이런 한계를 극복하기 위해 페이로드 시그니쳐 자동 생성 시스템이 다양한 방법으로 연구 되고 있으나 이 또한 시그니쳐의 생성 방법의 한계로 단일 응용의 멀티 프로토콜에 대한 시그니쳐 생성시 추출 성능 저하에 한계를 가지고 있다. 본 논문에서는 기존 연구의 단점을 해결하고자 프로토콜 필터를 적용하여 다양한 분석 방법에 적용될 수 있는 페이로드 시그니쳐를 자동생성 할 수 있는 시스템을 제안하고자 한다. 프로토콜 필터를 통해 멀티 프로토콜 기반의 응용에 대해 각 프로토콜 별 시그니쳐를 자동 생성 할 수 있음을 보이고, 이렇게 생성된 시그니쳐를 트래픽 분석을 통해 시스템의 타당성을 보였다.

## Ⅰ. 서 론

트래픽 분석에 있어 페이로드 시그니쳐 기반 분석 방법은 분석 성능 및 정확도 측면에서 매우 큰 장점을 가지고 있다. 하지만 트래픽의 정확한 분석을 위해서는 정확도가 높은 시그니쳐의 보유 여부가 시스템의 성능을 좌우 한다. 이렇게 정확도가 높은 시그니쳐를 생성하기 위해서는 두 가지 방법이 이용된다. 수동 생성 방법은 프로토콜을 분석 할 수 있는 전문가가 시그니쳐를 추출하는 방법으로 높은 성능의 시그니쳐를 추출하는데 효과적이나, 전문가의 확보 및 매우 빠르게 출현하는 응용에 대처하기가 힘들다는 한계점을 가지고 있다. 이러한 문제를 보완 하기 위해 자동 생성 방법이 연구되어 오고 있다. 자동 생성 방법은 트래픽의 페이로드로부터 다양한 알고리즘을 통해 공통 스트링을 추출한 후 이것을 시그니쳐로 활용 하는 방법이다. 본 논문에서는 LCS 알고리즘을 이용한다.

본 논문에서는 페이로드 시그니쳐 자동 생성을 위해 프로토콜 필터 기반 시그니쳐 자동생성 시스템을 제안 한다. 본 논문은 다음과 같은 순서로 구성된다. 2 장에서는 페이로드 시그니쳐 자동 생성 시스템에 대한 관련 연구를 알아 보고 3 장에서는 프로토콜 필터 기반 시그니쳐 자동 생성 시스템의 프로토콜 필터 구조에 대해 설명 하고 마지막으로 4 장에서는 결론 및 향후 연구에 대하여 기술한다.

## Ⅱ. 관련연구

선행 연구[1-2]의 페이로드 시그니쳐 자동 생성 시스템에서는 다음과 같은 한계점이 있다.

첫째 단일 응용의 멀티 프로토콜 기반 응용에 대한 페이로드 시그니쳐 추출 시 다양한 프로토콜에서 자주 나타나는 키워드 들이 공통 스트링으로 추출되는 단점이 있다. 이는 공통 스트링의 반복 빈도를 고려한 자동 생성 시스템에서 추출 성능 저하의 원인이 된다. 둘째 기존 연구는 응용을 프로토콜과 같은 분류 체계를 적용한 시그니쳐를 생성한다. 이렇게 되면 다양한 분류체계를 지원하기 위한 시그니쳐 생성에 한계가 있다. 따라서 다양한 분류체계[3]를 지원하기 위해서는 응용과 프로토콜이 다른 분류 기준으로 적용되어야 하며 시그니쳐 자동 생성 시스템 또한 하나의 응용에서 발생되는 다양한 프로토콜에 대한 프로토콜 종속 시그니쳐를 자동 생성 할 수 있어야 한다.

본 논문에서는 위에서 제시한 단점을 보완 할 수 있는 효율적인 프로토콜 필터 기반 시그니쳐 자동 생성 시스템을 제안 한다.

## Ⅲ. 프로토콜 필터 기반 시그니쳐 자동 생성 시스템

프로토콜 필터 기반 시그니쳐 자동 생성 시스템은 표 1 과 같이 5 개의 모듈로 구성된다. 각 모듈은 서로의 출력을 이용하여 독립 적으로 운영 될 수 있는 구조이다. 5 개의 모듈 중 FTG, SGE 에서 각각 프로토콜 필터가 적용 된다.

| 이름 | 역할 |
|---|---|
| **AGT (Application Ground Truth)** | 응용별 정답지 생성 |
| **FTG (Flow Type Grouping)** | 그룹핑 알고리즘을 이용한 트래픽의 기능별 분류 |
| **SGE (Signature Extractor)** | 공통 스트링을 갖는 응용 시스니쳐 추출 |

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

| | |
|---|---|
| **FFS (Flow Filter by Signature)** | 재생성 및 분석을 위한 시그니쳐 발생 플로우 제거 |
| **SGV (Signature Verifiers)** | 생성된 시그니쳐 검증 및 분석 |

**표 1 자동 생성 시스템의 모듈별 역할**

그림 1 은 두 모듈에서 사용되는 프로토콜 필터의 구성도다. Protocol Recognition Part 의 경우 프로토콜의 빠른 인식을 통해 한 응용의 플로우 집합을 프로토콜 별로 그룹핑하기 위한 모듈로서, 플로우를 다양한 방법으로 그룹핑하기 위한 모듈인 FTG 모듈 에서 활용된다. Protocol Feature Analysis Part 의 경우 프로토콜의 특징 분석을 적용한 효율적인 시그니쳐 생성에 초점이 맞추어져 있으며 공통 스트링 추출 모듈인 SGE 모듈에서 활용된다. 그림 2 는 특징 분석 파트를 HTTP 프로토콜에 적용한 HTTP 프로토콜 필터의 흐름도 이다. 최근 HTTP 트래픽의 증가[4]에 따라서 HTTP 필터의 적용은 시그니쳐 자동생성 시스템에 있어 필수적이다.

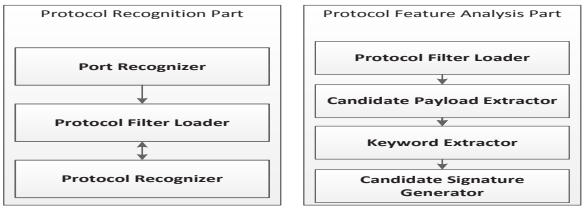


**그림 1 프로토콜 필터 구성도**

그림 2 의 Parse HTTP Head Field 에서 키워드 추출을 위해 HTTP Header 추출 및 Response Error Check 처리를 거쳐 대상 필드를 추출 한다. 이것은 의미 없는 데이터를 제거 하여 시그니쳐 추출 대상을 최적화 하기 위해 필요하다. Create Ordered Table 에서는 각 필드에 대한 우선 순위를 적용 함으로써 최종 추출된 공통 스트링이 정렬되어 출력 되도록 한다. 이렇게 정형화된 Payload 를 바탕으로 각 필드 별 공통 스트링을 추출 하는 Extract Common Sub String 단계를 거쳐 필드 별 후보 시그니쳐가 생성되며 최종 시그니쳐로 활용되기 위해 분리되어있는 후보 시그니쳐 필드를 Merge Processed Fields 단계를 거쳐 하나의 시그니쳐 파일로 생성 한다.

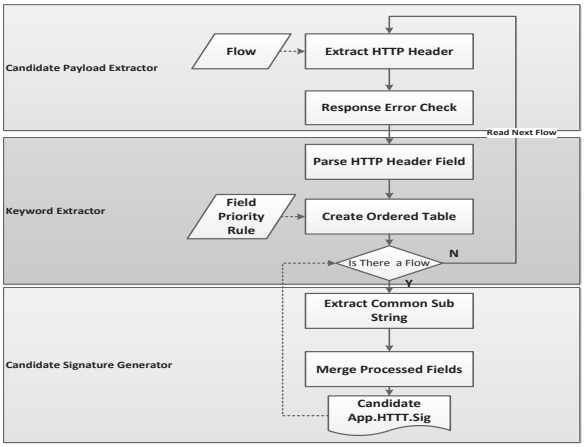


**그림 2 HTTP 프로토콜 필터의 특징 분석 파트**

실험을 위해 프로토콜 필터 기반 시그니쳐 자동 생성 시스템을 학내 망의 1 일 트래픽에 적용하였다. 표 2 는 자동 생성된 두 응용의 시그니쳐의 예로 추출된 시그니쳐를 검증 모듈인 SGV 모듈을 이용하여 AGT 모듈에 의해 생성된 두 응용의 정답지(106MB, 226MB)에 적용결과 nateonmain 의 경우 플로우 기준 60.8%의 분석율을 보였고 utorrent 의 경우 82.75%의 분석율을 보였다. 미 분석 트래픽 발생의 원인은 두 응용 모두 페이로드의 암호화 및 공통 스트링의 부제로 인해 시그니쳐가 생성되지 않아 분석되지 않은 것으로 나타났다. 또한 utorrent 의 경우 1 개의 플로우가 duri.ahn 응용으로 오 분류 되었다. 오 분류의 원인은 V3 업데이트를 위한 파일을 BitTorrent 프로토콜을 이용하여 전송하기 때문인 것으로 나타났다.

| 응용이름 | 시그니쳐 |
|---|---|
| **nateonmain** | ^NCPT 1*0<br>^RCON 1 dpl.nate.com 5004 |
| **nateonmain.http** | ^GET /exndr.jsp*<br>^GET /popup/ad_login/rollingad.html*<br>*User-Agent: NateOn/4.1.4.0 (2010)<br>*Host: nateonevent.nate.com |
| **utorrent** | ^!! BitTorrent protocol * WHB<br>^Go away, we're not home<br>*12:complete_agoi4e1:md11:upload_onlyi3e1<br>2:u*<br>*-UT2210-sb* |
| **utorrent.http** | ^GET /announce?info_hash=*<br>Host:t.com:2710<br>User-Agent:uTorrent/2210(25203) |

**표 2 자동생성 시그니쳐**

## Ⅳ. 결론

본 논문에서는 프로토콜 필터 기반 시그니쳐 자동 생성 시스템을 제안 하였다. 관련연구에서 제시한 페이로드 시그니쳐 자동 생성 시스템의 단점을 프로토콜 필터를 통해 보완하여 HTTP 프로토콜 필터 적용을 통해 프로토콜 별 시그니쳐를 생성 할 수 있는 시스템을 제시하였다. 또한 추출된 시그니쳐를 이용하여 정답지 트래픽 분석을 통해 시스템의 타당성을 보였다. 향후 연구로는 필터를 더욱 세분화 하여 정확도가 높은 프로토콜 별 시그니쳐를 생성하기 위한 시그니쳐 자동 생성 시스템의 연구를 진행 할 계획이다.

## 참 고 문 헌

[1] Jun-Sang Park, Jin-Wan Park, Sung-Ho Yoon, Hyun-Shin Lee and Myung-Sup Kim, “ Development of Signature Generation and Update System for Application-level Traffic Classification,“ KIPS, Feb. 2010, pp. 99-108

[2] M. Ye, K. Xu, J. Wu, and H. Po. Autosig-automatically generating signatures for applications. In CIT (2), pages 104- 109. IEEE Computer Society, 2009.

[3] Ji-hye Kim, Sung-Ho Yoon and Myung-Sup Kim, "Research on Traffic Taxonomy for Internet Traffic Classification," Proc. of the Asia-Pacific Network Operations and Management Symposium (APNOMS) 2011, Taipei, Taiwan, Sep. 21-23, 2011.

[4] Sang-Woo Lee, Myung-Sup Kim, "A Study on the Smart-Phone Application Signature extraction using HTTP User-Agent Field", KICS, June. 20-22, 2011, pp.614