응용 트래픽 분류에 정교한 페이로드 시그니쳐 구조

Payload Signature Structure for Elaborate Application Traffic Classification

구영훈, 이성호, 김성민, 이수강, 김명섭

고려대학교 컴퓨터정보학과

{gyh0808, gaek5, gogumiking, sukanglee, tmskim}@korea.ac.kr

요 약

오늘날 네트워크 고속화와 함께 인터넷 연결이 가능한 다양한 디바이스의 보편화로 인해 네트워크 기능을 사용하는 응용이 급속도로 생성되고 있다. 이에 따라 한정된 네트워크 자원을 효율적으로 사용하고 안정적인 서비스를 제공할 수 있도록 다양한 종류의 응용에 대한 정확한 트래픽 분류가 필수적이다. 다양한 페이로드 시그니쳐 추출 방법론 가운데, 대부분의 페이로드 시그니쳐 구조는 단순히 트래픽의 페이로드 내 출현 빈도가 높은 공통 부분 문자열이다. 따라서, 특정 응용프로그램에 고유하지 않고 다른 응용프로그램과 중복되는 페이로드 시그니쳐가 추출될 문제가 있다. 이에 본 논문에서는 다른 응용프로그램과 중복되지 않고 특정 응용프로그램에 특화된 보다 정교한 시그니쳐의 구조를 제안한다. 본 시그니쳐의 구조는 3 단계로 공통 부분 문자열인 Content 시그니쳐, 동일한 Packet 내 Content 시그니쳐의 집합인 Packet 시그니쳐, 동일한 Flow 내 Packet 시그니쳐의 집합인 Flow 시그니쳐로 구성된다. 본 시그니쳐 구조와 기존의 페이로드 시그니쳐 구조를 실제 응용 트래픽 분류에 적용하여 그 실효성을 증명한다.

Keyword: Signature structure, Content signature, Packet signature, Flow signature, Completeness, False positive

1. 서론

오늘날 네트워크 고속화와 더불어 인터넷 연결이 가능한 스마트 기기들의 보편화에 따라 인터넷에 기반한 응용프로그램의 사용이 급격하게 증가하고 있다. 이에 따라 한정된 네트워크 자원을 효율적으로 사용하고, 사용자에게 안정적인 서비스를 제공하기 위한 많은 연구가 수행되고 있다. 이를 위해서는 다양한 종류의 응용 레벨 트래픽을 정확하게 분류할 수 있는 방법이 필요하다.

트래픽의 분류를 위한 다양한 방법론이 존재하지만 분석률과 정확도 측면에서 가장 성능이 높은 방법론은 페이로드 시그니쳐 기반 분석 방법이다[1,2,7,8]. 이는 페이로드를 추출한 응용의 트래픽에 대해서는 정확한 분석이 가능하기 때문이다. 하지만 오늘날 급증하는 응용과 함께 확인되지 않은 트래픽에 대하여 적용하였을 시 시그니쳐의 중복성에 의해 잘못 분류할 가능성이 존재한다.

기존에 연구된 대부분의 페이로드 시그니쳐의 구조는 단순히 페이로드 내의 공통 부분 문자열이다[3,5,10]. 따라서 특정 응용프로그램에 고유하지 않고 다른 응용프로그램과 중복되는 페이로드 시그니쳐가 추출될 가능성이 있다. 이는 네트워크 관리에 있어 신뢰도를 떨어뜨리며 네트워크 정책이나 고장 진단, 용량 계획 등을 정확히 수행할 수 없다. 특히 네트워크 보안 분야에서는 악성 트래픽을 오탐(False Positive)하거나 미탐(False Negative)할 경우 엄청난 손실을 초래할 수 있다.

급격하게 증가하는 응용 트래픽에서 네트워크 관리자가 페이로드를 일일히 확인하여 검증된 시그니쳐를 수작업으로 추출하는 과정은 많은 시간과 인력을 낭비한다. 이를 해결하기 위한 다양한 자동 페이로드 시그니쳐 추출 방법 연구가 진행되고 있지만, 자동으로 추출된 페이로드 시그니쳐의 경우 각 응용 및 서비스를 오탐이나 미탐이 없이 그 응용만을 정확하게 탐지할 수 있는 시그니쳐인지 보장할 수 없다. 따라서 좀 더 특정 응용 프로그램에 고유하고 다른 응용프로그램의 시그니쳐와 확실히 구분되는 정교한 페이로드 시그니쳐 구조가 필요하다.

본 논문에서는 특정 응용프로그램에 특화된 보다 정교한 페이로드 시그니쳐 구조를 제안한다. 본 시그니쳐의 구조는 총 3 단계로 페이로드 내 공통 부분 문자열인 Content 시그니쳐, 동일한 패킷 내 Content 시그니쳐의 집합인 Packet 시그니쳐, 동일한 Flow 내 Packet 시그니쳐의 집합인 Flow 시그니쳐로

이 논문은 2015 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2015R1D1A3A01018057).

구성된다. 본 시그니쳐 구조와 선행 연구의 시그니쳐 구조를 실제 응용 트래픽 분류에 적용하여 비교한 결과 응용프로그램의 분석률은 유지하면서 오탐률을 감소시켜 정확한 분류가 이루어질 수 있도록 하였다.

본 논문은 다음과 같은 순서로 기술한다. 본 장의 서론에 이어 2 장에서는 관련 연구에 대해 살펴보고, 3 장에서는 선행 연구의 한계를 설명한다. 4 장에서는 제안하는 페이로드 시그니쳐의 정의와 구조에 대해 설명한다. 5 장에서는 실제 응용 트래픽 분류에 제안한 페이로드 시그니쳐 구조를 적용하여 그 타당성을 증명하기 위한 실험 결과를 기술한다. 마지막으로 6 장에서는 결론 및 향후연구를 기술한다.

2. 관련 연구

페이로드 시그니쳐 분석 방법은 Packet 의 페이로드를 확인하여 다른 응용 프로그램과 구분 지을 수 있는 특징을 추출하고 응용을 식별할 수 있는 시그니쳐로 정의한 후 분석할 트래픽 Packet 의 페이로드 내에 응용의 시그니쳐 포함 여부를 판단하여 응용을 분석하는 방법이다. 페이로드 시그니쳐 추출 방법론은 다양한 방법으로 연구되고 있다.

Kim 의 연구[3]에서는 웜 시그니쳐 생성을 위한 방법으로 COPP(Content-based Payload Partitioning)를 사용하여 페이로드 내 연속적으로 발생하는 공통 부분 문자열을 breakmark 로 사용하였다. Cheng 의 연구[10]에서는 페이로드를 3 종류의 비트(1bit, 4bit, 8bit) 단위로 자른 후 같은 오프셋에서 두 비트 시퀀스 간의 공통 비트 시퀀스를 찾아 시그니쳐로 활용하였다. Mingjiang 의 연구[11]에서는 Shingle 이라는 단위의 서브 스트링을 정의하고 Shingle 의 발생 빈도수 및 임계값을 이용하여 최종 공통 문자열을 페이로드 시그니쳐로 사용하였다. Park 의 연구[5]에서는 바이오 인포매틱스의 유전자 분석 알고리즘 기법의 하나인 LCS (Longest Common String) 알고리즘을 응용 트래픽 시그니쳐 추출 목적에 맞게 변형한 LASER(LCS-based Application Signature ExtRaction) 알고리즘을 사용하여 비교할 두 페이로드 내에서의 최장 길이 공통 문자열을 시그니쳐로 활용하였다. Newsome 의 연구[4]와 Feng 의 연구[6]에서는 Smith-Waterman 알고리즘을 이용하여 두 페이로드 내의 공통적인 부분 문자열의 집합을 시그니쳐로 사용하였다.

앞서 언급한 연구들의 페이로드 시그니쳐 구조는 단순히 페이로드 내 공통 부분 문자열이므로 다른 응용과 중복이 되는 시그니쳐가 추출될 문제가 있다. 이에 정확한 트래픽 분류를 위하여 특정 응용에 더 정교한 페이로드 시그니쳐 구조가 필요하다.

3. 선행 연구의 페이로드 시그니쳐 구조의 한계

본 장에서는 선행 연구의 페이로드 시그니쳐 구조에 대한 한계를 설명한다. 위에서 언급한 논문들의 페이로드 시그니쳐 구조는 Smith-Waterman 알고리즘을 제외하고 단순히 여러 페이로드에서 공통적으로 발견되는 부분 문자열이다. 이와 같은 페이로드 시그니쳐 구조는 특정 응용에 대한 정답지 트래픽을 수집하여 추출한 시그니쳐라 할지라도 이 시그니쳐가 목적하고 있는 특정 응용에 대해 고유한지 다른 응용의 트래픽에서는 전혀 발견이 되지 않는지 보장할 수 없다.

표 1 과 표 2 는 추출된 시그니쳐 중 다른 응용과 중복이 될 가능성이 큰 시그니쳐의 예이다.

Example
HTTP /1.1 Accept Referrer GET / 201 User-Agent Content-Type

표 1. 특정 프로토콜의 메타데이터로 쓰이는 문자열

표 1 과 같이 특정 프로토콜에 메타 데이터로 쓰이는 문자열이 시그니쳐로 추출될 경우 이 프로토콜을 쓰는 응용에 대해서는 모두 같은 시그니쳐가 추출이 될 수 있다. 예를 들어 여러 Packet 에서 발견되는 HTTP Request, Response 의 명령어나 HTTP header Field 의 경우 페이로드 시그니쳐로 추출될 가능성이 크고, HTTP 프로토콜은 다양한 응용에서 사용되는 프로토콜이므로 특정 응용의 식별에 큰 효용 가치가 없는 시그니쳐이다.

또한 사전적 의미를 가지는 문자열이 특정 응용의 페이로드 시그니쳐로 추출되는 경우 이는 다른 응용의 트래픽에서도 발견될 가능성이 매우 크다. 표 2 는 사전적 의미를 가지는 단어가 특정 응용의 페이로드 시그니쳐로 추출되는 경우이다.

Example
Album Image Data : Server : Music

표 2. 사전적 의미를 가지는 문자열

예를 들어 Album 과 같은 문자열은 웹 캠 응용의 트래픽에서도 나올 수 있고 음악 관련 응용의 트래픽에서도 나올 수 있다. Music 과 같은 문자열의 경우에도 같은 목적의 음악 관련 응용이라 할지라도 다양한 음악 서비스를 제공하는 응용 중 정확히 어

떤 음악 응용 프로그램인지 구분할 수 없다.

그림 1 은 Smith-Waterman 알고리즘을 이용하여 추출한 페이로드 시그니쳐의 구조로 단순히 공통 부분 문자열이 아닌 두 페이로드 내의 공통 부분 문자열의 집합이다. 이와 같은 페이로드 시그니쳐 구조의 경우 앞서 언급한 논문들의 페이로드 시그니쳐 구조보다는 응용에 대한 고유성이 높아질 수 있다. 두 개의 Packet 페이로드를 입력으로 하여 공통 부분 문자열의 집합을 찾기 때문에 이 시그니쳐는 한 Packet 단위의 시그니쳐라 할 수 있다. 이는 앞서 언급한 공통 부분 문자열 시그니쳐보다 정교하다.

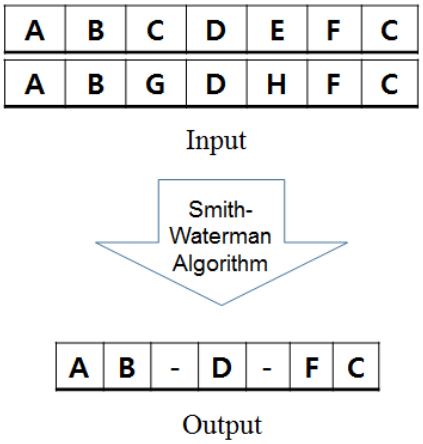

그림 1. Smith-Waterman 알고리즘

그러나 Smith-Waterman 알고리즘의 경우 시간복잡도와 계산 복잡도가 크며 입력으로 항상 2 개의 packet 페이로드를 사용하여 하나의 공통 문자열 집합을 생성하기 때문에 모든 Packet 내의 공통 부분 문자열 집합을 찾기 위해서는 Packet 개수의 제곱만큼 Smith-Waterman 알고리즘 과정을 수행해야 한다. 결과물로 나온 공통 부분 문자열 집합 중 시그니쳐로 추출할 공통 부분 문자열 집합을 선정하기 위하여 일정 빈도수나 임계값을 계산하고 이를 넘는 공통 부분 문자열 집합을 탐색하기 위해서는 더 많은 계산 과정이 필요하다. 또한 그림 1 에서 LASER 알고리즘과 같은 선행 연구의 방법에서는 시그니쳐로 추출될 수 있었던 공통 부분 문자열 AB, D, FC 는 Smith-Waterman 알고리즘을 사용시 추출이 되지 않아 AB, D, FC 가 분석할 수 있었던 트래픽을 분류할 수 없게 되어 분석률이 감소할 수 밖에 없다. 그림 2 에서는 AB 와 D 와 FC 가 동시에 존재하는 Packet 만을 분석하는 Smith-Waterman 알고리즘 분석률의 한계를 도식화한 것이다.

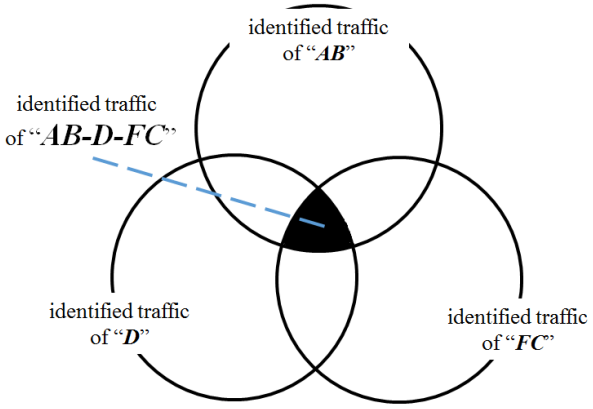

그림 2. Smith-Waterman 알고리즘의 분석률의 한계

다음 장에서는 다른 응용과 중복되지 않고 특정 응용에 정교한 페이로드 시그니쳐의 구조를 설명한다.

4. 제안하는 페이로드 시그니쳐의 구조

본 논문에서는 3 단계로 구성된 페이로드 시그니쳐 구조를 제안한다. 먼저 1 차적으로 페이로드 내 일정 빈도 수를 만족하는 공통 부분 문자열을 시그니쳐로 추출하고 이를 Content 시그니쳐라 명명한다. 이는 선행연구의 대부분의 페이로드 시그니쳐의 구조와 같다. 2 차적으로 동일한 Packet 내에서 추출되는 Content 시그니쳐의 집합 중 일정 빈도수를 만족하는 집합을 시그니쳐로 추출하고 이를 Packet 시그니쳐라 명명한다. Packet 시그니쳐는 하나의 Packet 에서 여러 개의 Content 시그니쳐가 매칭되어야 하기 때문에 오탐률이 적어 정확성이 향상된다. 따라서 Content 시그니쳐보다 응용에 정교한 시그니쳐이다. 3 차적으로 동일한 Flow 내에서 추출되는 Packet 시그니쳐의 집합 중 일정 빈도 수를 만족하는 집합을 Flow 시그니쳐라 명명한다. Flow 시그니쳐는 하나의 Flow 내에서 여러 개의 Packet 시그니쳐가 매칭되어야 하기 때문에 Packet 시그니쳐보다 더 응용에 정교하고 오탐률이 적다.

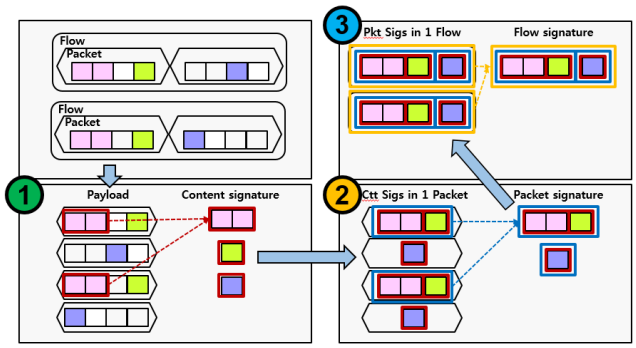

그림 3. 제안하는 시그니쳐 구조의 추출 과정

수식 1 은 제안하는 시그니쳐의 구조이다. C 는 Content 시그니쳐, P 는 Packet 시그니쳐, F 는 Flow

시그니쳐를 의미한다.

C={c | c is single substring in a payload}

P={p | p is a subset of C, p appears in a packet}

F={f | f is subset of P, f appears in a flow}

수식 1. 제안하는 시그니쳐의 구조

한 응용에 대한 공통 부분 문자열 중 일정 빈도수가 넘는 9 개의 Content 시그니쳐를 추출하였을 때 $C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9\}$와 같이 표현할 수 있다. 이 중 동일한 Packet 내에 존재하는 Content 시그니쳐의 집합 중 일정 빈도 수가 넘는 Packet 시그니쳐가 4 개가 추출되었을 때 $P = \{p_1, p_2, p_3, p_4\}$이며 이는 $P = \{p_1, p_2, p_3, p_4\} = \{\{c_1, c_3, c_9\}, \{c_2, c_5, c_9\}, \{c_8, c_4\}, \{c_6\}\}$과 같다. 또한 Flow 시그니쳐가 2 개 추출이 되었을 때는 $F = \{f_1, f_2\}$라 표현하며 $F = \{f_1, f_2\} = \{\{p_1, p_3\}, \{p_4\}\} = \{\{\{c_1, c_3, c_9\}, \{c_8, c_4\}\}, \{\{c_6\}\}\}$과 같은 의미이다.

표 3, 표 4, 표 5 는 제안한 시그니쳐의 구조로 추출한 Content 시그니쳐, Packet 시그니쳐, Flow 시그니쳐의 예시이다.

$Signature_{id}$	Content Signature
1	<GET />
2	<HTTP /1.1>
3	<MUSIC>
4	<CACHE-CONTROL>
5	<NO>
6	<FACEBOOK>
7	<MUSIC>
8	<ALBUM>
9	<32>

표 3. Content 시그니쳐의 예시

$Signature_{id}$	Packet Signature
1	< <GET /> <MUSIC> <32> >
2	< <HTTP /1.1> <NO> <32> >
3	< <ALBUM> <CASHE-CONTROL> >
4	< <FACEBOOK> >

표 4. Packet 시그니쳐의 예시

$Signature_{id}$	Flow Signature
1	< < <GET /> <MUSIC> <32> > < <ALBUM> <CASHE-CONTROL> > >
2	<<<FACEBOOK>>>

표 5. Flow 시그니쳐의 예시

표 3 의 Content 시그니쳐 6 과, Packet 시그니쳐 4 와 표 5 의 Flow 시그니쳐 2 는 모두 “FACEBOOK” 이라는 문자열을 나타내는 시그니쳐이다. 응용에 대한 분석률은 모두 같을지라도 Content 시그니쳐 6 보다 Packet 시그니쳐 4 가, Packet 시그니쳐 4 보다 Flow 시그니쳐 2 가 더 응용에 정교하다고 할 수 있다.

한편, Packet 시그니쳐를 사용하거나 Flow 시그니쳐를 사용할 시 그림 2 에서 표현한 분석률의 한계가 존재할 수 있다. 하나의 Packet 시그니쳐가 매칭되기 위해서는 그 Packet 시그니쳐를 생성하기 위해 사용된 Content 시그니쳐도 모두 매칭되어야 하기 때문이다. 일반적으로 분석률은 수식 2 와 같이 계산할 수 있다. 트래픽의 양으로는 Flow 의 수 혹은 Packet 의 수나 용량(Bytes)를 사용할 수 있다.

$$\text{분석률} = \frac{\text{시그니쳐가 매칭된 트래픽의 양}}{\text{전체 트래픽의 양}} \times 100$$

수식 2. 일반적인 분석률의 계산식

그러나 Packet 시그니쳐 혹은 Flow 시그니쳐의 분석률은 수식 2 와는 다른 방법으로 계산되어야 한다. 상위 단계 시그니쳐는 하위 단계 시그니쳐의 집합으로 생성되었기 때문에 하위 단계 시그니쳐가 분석한 트래픽은 상위 단계의 시그니쳐도 분석을 할 수 있다 해도 무방하다. 예를 들어, 3 개의 Content 시그니쳐 A, B, C 가 하나의 Packet 시그니쳐를 생성하는데 사용되었다면 이 Packet 시그니쳐의 분석률은 A, B, C 가 매칭된 트래픽의 양을 단순히 더하는 것이 아닌 분석한 트래픽들의 합집합의 양을 전체 트래픽의 양으로 나눈 것의 백분율로 계산할 수 있다. A 와 B 와 C 가 각각 분석한 트래픽이 서로 중복될 수 있기 때문이다.

$$\text{분석률} = \frac{A \cup B \cup C\text{의 양}}{\text{전체 트래픽의 양}} \times 100$$

수식 3. Packet 시그니쳐의 계산식의 예

따라서, 응용 트래픽 분석에 있어 상위 단계의 시그니쳐를 사용할 시 하위 단계 시그니쳐의 분석률을 그대로 유지하면서 오탐률은 낮추어 특정 응용을 더 정확히 분석할 수 있다.

상위 단계의 시그니쳐를 다수 보유할 시 대용량의 트래픽을 효율적으로 분석할 수 있다. 대용량의 트래픽을 발생하는 네트워크의 효율적인 분석을 위해서는 Flow 단위의 트래픽 수집과 분석이 필요하다[9]. 상위 단계의 시그니쳐일수록 특정 응용에 정교하므로 생성되기 위한 조건이 까다롭다. 동일한 단위(Packet 또는 Flow) 내에 하위 단계 시그니쳐들이 모두 매칭되어야 하며 이로써 만들어진 하위 단계 시그니쳐들의 집합이 일정 이상의 빈도 수를 만족해야 시그니쳐로 생성되기 때문이다. 따라서, 상위 단계의 시그니쳐 구조를 사용할 시 시그니쳐의 개수를 줄이면서 특정 응용에 더 정교한 시그니쳐가 생성된다.

5. 실험 및 결과

본 장에서는 제안한 페이로드 시그니쳐의 구조와 선행 연구의 페이로드 시그니쳐 구조를 특정 응용 트래픽에 적용하여 그 실효성을 증명한다.

실험 환경은 다음과 같다. 먼저 수집된 트래픽을 TMA(Traffic Measurement Agent)를 이용하여 2 가지 응용에 대한 정답지 트래픽을 생성한다. 본 실험에서는 2 가지 응용을 Naver 와 Yahoo 로 선정하였다. 2 가지 응용의 정답지 트래픽을 바탕으로 각각 응용들의 Content 시그니쳐와 Packet 시그니쳐를 추출한다. 추출한 2 가지 응용의 각 시그니쳐를 2 가지 응용에 대한 정답지 트래픽에 반대로 적용하여 표 6 의 (2)와 (3)을 구한다. (1)과 (4)는 2 가지 응용의 시그니쳐로 각 응용에 해당하는 정답지 트래픽에 적용한 분석률이다. 최종적으로 (1), (2), (3), (4)의 값을 가지고 Content 시그니쳐와 Packet 시그니쳐의 분석률과 오탐률을 비교 분석한다.

	Naver.Sig	**Yahoo.Sig**
Naver.GTT	(1) TP_{Naver} $= TN_{Yahoo}$	(2) FN_{Naver} $= FP_{Yahoo}$
Yahoo.GTT	(3) FP_{Naver} $= FN_{Yahoo}$	(4) TN_{Naver} $= TP_{Yahoo}$

표 6. 실험 환경

표 6 에서 G.T.T 는 정답지 트래픽을 의미하고, Sig 는 시그니쳐를 의미한다. (1)은 Naver 의 시그니쳐를 Naver 의 정답지 트래픽에 적용한 분석률을 나타내며 이는 Naver 응용 분석에 대한 TP(True Positive)이다. (2)는 Yahoo 의 시그니쳐를 Naver 의 정답지 트래픽에 적용한 분석률을 나타내며 Yahoo 응용 분석에 대한 FP(False Positive), 즉 오탐률이다. (3)은 Naver 의 시그니쳐를 Yahoo 의 정답지 트래픽에 적용한 분석률을 나타내며 Naver 응용 분석에 대한 FP 이고 (4)는 Yahoo 의 시그니쳐로 Yahoo 의 정답지 트래픽을 분석한 분석률을 나타내며 Yahoo 응용에 대한 TP 이다.

	Content Signature		Packet Signature	
	TP	FP	TP	FP
Naver	5,567 /5,739 **97%**	403 /3,847 **10%**	5,459 /5,739 **95%**	205 /3,847 **5%**
Yahoo	3,774 /3,847 **98%**	635 /5,739 **11%**	3,740 /3,847 **97%**	309 /5,739 **5%**

표 7. 실험 결과

표 7 은 실험 결과로 모든 값은 트래픽의 Flow 단위이다. 본 실험 결과에서 Naver 와 Yahoo 의 응용에 대해 모두 Content 시그니쳐의 오탐률보다 Packet 시그니쳐의 오탐률이 더 낮았으며 Content 시그니쳐에 비해 Packet 시그니쳐의 분석률은 변화가 적었다.

6. 결론 및 향후 연구

본 논문에서는 다른 응용프로그램과 중복되지 않도록 특정 응용프로그램에 정교한 페이로드 시그니쳐의 구조를 제안하였다. 이를 통해 응용별 시그니쳐 추출시 시그니쳐 중복을 방지하고 응용 트래픽 분석에서 오탐률을 감소시킴으로 인해 정확도를 향상시킬 수 있다. 따라서 분석 결과에 신뢰성을 부여하여 효율적인 네트워크 관리가 가능하다. 실험을 통해 제안한 시그니쳐 구조와 선행 연구의 시그니쳐 구조의 TP(True Positive) 및 FP(False Positive)의 수치를 비교 분석한 결과, TP(True Positive)는 유지시키고 FP(False Positive)의 수치는 감소시킴으로써 본 페이로드 시그니쳐 구조의 실효성을 증명하였다.

향후 연구로써는 제안한 페이로드 시그니쳐의 구조를 실시간 트래픽 분류에 활용할 수 있도록 자동 페이로드 생성 시스템에 적용하고 이를 관리 및 검증할 수 있는 방안에 대한 연구를 통해 본 시그니쳐의 실용성을 확보하고자 한다. 또한 본 페이로드 시그니쳐 구조를 사용하여 추출한 시그니쳐들을 최적화할 수 있는 적절한 임계값 산출 방법을 연구하고자 한다.

참고 문헌

[1] 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, pp. 5869-5875, 2008.

[2] Liu, Hui Feng, Wenfeng Huang, Yongfeng Li, Xing “Accurate Traffic Classification”, Networking, Architecture, and Storage, NAS 2007. International Conference

[3] H.-A. Kim and B. Karp, “Autograph: Toward automated, distributed worm signature detection,” in *USENIX Security Symp.*, vol. 286, 2004.

[4] J. Newsome, B. Karp, and D. Song, “Polygraph: Automatically generating signatures for polymorphic worms,” *IEEE Symp. Security and Privacy*, pp. 226-241, 2005.

[5] B.-C. Park, Y. J. Won, M.-S. Kim, and J. W. Hong, “Towards automated application signature generation for traffic identification,” *IEEE Network Operations and Management Symp. (NOMS 2008)*, pp. 160-167, 2008.

[6] X. Feng, X. Huang, X. Tian, and Y. Ma, “Automatic traffic signature extraction based on Smith-waterman algorithm for traffic classification,” *IEEE Int. Conf. Broadband Netw. Multimedia Technol. (IC-BNMT)*, pp. 154-158, 2010.

[7] 박준상, 윤성호, 박진완, 이현신, 이상우, 김명섭, "페이로드 시그니쳐 기반 응용 레벨 트래픽 분류 시스템 성능 향상에 관한 연구," KNOM

Review, Vol. 12, No. 2, Dec. 2009, pp. 12-21.
[8] S.-H. Lee, J.-S. Park, S.-H. Yoon, and M.-S. Kim , "High performance payload signature-based Internet traffic classification system ," *Proc. of the Asia-Pacific Network Operations and Management Symposium (APNOMS) 2015, Busan, Korea*, Aug. 19-21, 2015, pp.491-494.
[9] 윤성호, 박준상, 김명섭, "멀티 코어 환경에서 실시간 트래픽 분석 시스템 처리속도 향상", 한국통신학회 논문지 '12-05 Vol.37B No.5, pp. 348-356, May 2012.
[10] C. MU, X.-h. HUANG, X. TIAN, Y. MA, and J.-l. Qi, "Automatic traffic signature extraction based on fixed bit offset algorithm for traffic classification," *The J. China Universities of Posts and Telecommun.*, vol. 18, pp. 79-85, 2011.
[11] M. Ye, K. Xu, J. Wu, and H. Po, "AutoSig-Automatically Generating Signatures for Applications", *Proc. Of IEEE 9th International Conference on Computer and Information Technology (ICTI)*, Xiamen, China, October 11 – 14 , 2009 .