트리 구조 기반 정교한 페이로드 시그니쳐 자동 생성 연구

심규석, 구영훈, 김명섭

고려대학교

{kusuk007, gyh0808, tmskim}@korea.ac.kr,

요 약

네트워크 분야의 범위가 넓어지고, 네트워크 자원의 활용도가 증대되면서 네트워크 관리를 위한 트래픽 분석에서 현재 연구단계의 한계가 증대되고 있다. 이러한 한계를 극복하기 위한 다양한 연구분야 중 하나는 시그니쳐 구조에 대한 연구이다. 시그니쳐 구조 연구는 시그니쳐 생성뿐만 아니라 시그니쳐를 이용한 트래픽 분류 과정에서도 중요하게 여겨지는 연구다. 본 논문에서는 트리 구조로 변화된 페이로드 시그니쳐 구조와 트리 구조 페이로드 시그니쳐로 분류되는 트래픽 분류 시스템을 제안한다. 제안된 방법론은 페이로드 시그니쳐를 단순화 할 뿐만 아니라 분류 과정에서도 속도 향상에 대한 기대 효과를 나타낼 수 있다.

1. 서론

오늘날 네트워크는 모든 분야에서 활용되고있다. 따라서 네트워크 자원의 활용도는 증대되면서 네트워크 관리 대상 트래픽의 양은 기하급수적으로 증대되고 있다. 따라서 네트워크 관리 분야는 모든 트래픽의 출처를 파악하고, 빠른 속도로 트래픽 분류해야 되는 과제를 가지고 있다. 이러한 과제를 해결하기 위해 다양한 연구가 진행되고 있다.

다양한 연구 중 하나는 트래픽을 분류할 수 있는 키에 해당하는 시그니쳐를 자동으로 추출하고, 추출된 시그니쳐를 바로 적용할 수 있는 구조를 연구하는 것이다. 그러나, 기존 트래픽 분류 방법론을 유지함과 동시에 시그니쳐 자동 추출 시스템에서 추출된 시그니쳐를 적용한다면 무차별적으로 추출된 모든 시그니쳐를 적용함으로써 발생하는 부작용인 트래픽 분류 속도 저하를 감당해야한다.

따라서 본 연구에서는 기존 무질서하게 자동으로 추출되는 페이로드 시그니쳐 구조에서 트리구조로된 페이로드 시그니쳐 구조에 대해 제안하고, 트리구조 페이로드 시그니쳐를 이용한 트래픽 분류 시스템에 대해 제안한다. 제안하는 페이로드 시그니쳐 구조는 트리구조로 이루어져 있어서 시그니쳐를 단순화 할 뿐만 아니라 트래픽 분류 시스템에 적용할 때 빠른 시간내에 트래픽을 분류할 수 있는 장점이 있다.

본 논문은 본 장 서론에 이어, 2 장에서 페이로드 시그니쳐 구조에 대해 제안하고, 마지막 3 장에서 결론 및 향후연구에 대해 언급한 후 본 논문을 마친다.

2. 제안하는 페이로드 시그니쳐 구조

기존 시그니쳐 자동 생성 시스템을 통해 추출된 시그니쳐는 총 3 가지 타입으로 나누어진다. 페이로드에서 공통으로 발견되는 연속된 문자열을 의미하는 콘텐츠 시그니쳐, 같은 패킷 내에 공통으로 발견되는 콘텐츠 시그니쳐의 집합을 의미하는 패킷 시그니쳐, 마지막으로 같은 플로우 내에 공통으로 발견되는 패킷 시그니쳐의 집합을 의미하는 플로우 시그니쳐이다.

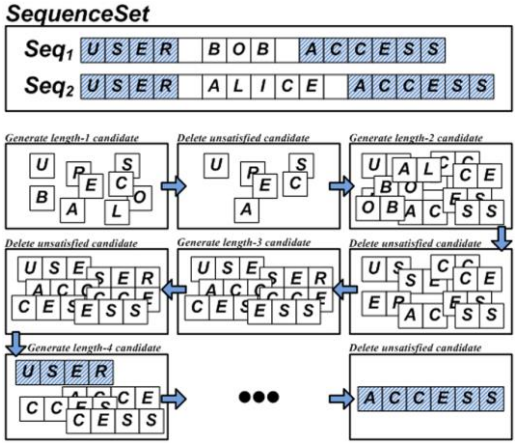

그림 1. 페이로드 콘텐츠 시그니쳐 자동 생성 과정

시그니쳐 자동 생성은 패킷 페이로드 내용을 기반으로 순차 패턴 알고리즘을 통해 각 트래픽 트레이스에서 공통으로 발생하는 패턴을 추출한다[1,2]. 그림 1 은 다음의 과정을 나타낸다. 패킷 페이로드를 추출하여 시퀀스를 생성하고, 생성된 시퀀스들의 집합에서 길이 1 의 콘텐츠 시그니쳐 후보를 생성한다. 길이 1 의 콘텐츠 시그니쳐 후보들의 미리 정의한 최소 지지도 검사를 통해 만족하지 않는 후보자들은 삭제된다. 남은 길이 1 의 콘텐츠 시그니쳐 후보들을 조합하여 길이 2 의 콘텐츠 시그니쳐 후보자를 생성하고, 위의 과정을 더 이상 길이가 증가되지 않을 때까지 반복하여 최종 공통으로 발생하는 연속된 문자열을 추출함으로써 콘텐츠 시그니쳐들을 추

이 논문은 2015 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No.2015R1D1A3A01018057)과 2016 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 과학기술인문융합연구사업임(No.NRF-2016M3C1B6929228).

출한다.

콘텐츠 시그니쳐 추출과 마찬가지로 패킷 시그니쳐, 플로우 시그니쳐도 같은 방법으로 추출한다. 콘텐츠 시그니쳐 추출을 위해 길이 1 을 문자하나로 정의했다면, 패킷 시그니쳐 추출을 위해 길이 1 을 콘텐츠 시그니쳐 하나도 정의한다. 플로우 시그니쳐에서는 길이 1 을 하나의 패킷 시그니쳐로 정의한다.

위의 방법으로 추출된 시그니쳐의 구조는 다음 그림 2 에서 패킷 시그니쳐의 예와 같이 비슷한 내용이 포함되어 있는 시그니쳐가 존재한다. 이러한 시그니쳐들은 분석률, 정확함 그리고 정교함을 고려하기 위해 삭제 및 통합하기 어려운 조건의 시그니쳐들이다. 그러나 이러한 구조는 추출된 시그니쳐의 개수를 증가시켜 추출된 시그니쳐를 분석하기 어려울 뿐만 아니라 트래픽 분류 과정에서도 비슷한 시그니쳐들을 지속적으로 매칭해야 하기 때문에 소비되는 시간이 증가한다.

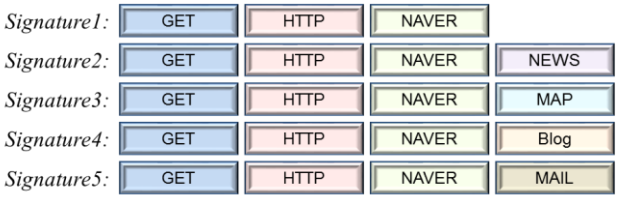

그림 2. 기존 페이로드 시그니쳐 구조

따라서 본 시그니쳐 구조의 한계를 극복하기 위해 본 논문은 트리화 된 페이로드 시그니쳐 구조를 제안한다. 트리화 된 페이로드 시그니쳐 구조는 시그니쳐의 개수를 대폭적으로 줄일 수 있고, 트래픽 분류과정에서 시간 단축을 할 수 있는 장점이 있다. 기존 HTTP 트래픽을 대상으로 연구되고 있지만, 본 방법은 HTTP 트래픽 뿐만 아니라 시그니쳐 자동생성 모듈에서 추출된 모든 트래픽을 대상이 가능하다[3].

트리화 된 페이로드 시그니쳐 구조는 시그니쳐 추출 과정에서 생성된다. 시그니쳐 생성 시 콘텐츠 시그니쳐는 기존 방법과 동일하게 생성되지만, 페이로드 시그니쳐는 같은 패킷 내에서 공통으로 발생하는 콘텐츠 시그니쳐의 집합이기 때문에 그림 2 과 같이 공통적으로 발생하지만 다른 시그니쳐가 추출될 때 해당 시그니쳐 구조를 트리화하여 추출한다. 트리 화된 페이로드 시그니쳐의 구조는 다음 그림 3 와 같다. 다음 그림 3 과 같이 트리화 된 페이로드 시그니쳐 구조는 기존 그림 2 에서의 패킷 시그니쳐 5 개의 시그니쳐를 하나의 트리 패킷 시그니쳐로 변화할 수 있다.

본 장에서는 트리화 된 페이로드 시그니쳐 구조뿐만 아니라 트리화 된 페이로드 시그니쳐를 이용하여 트래픽을 분류할 수 있는 시스템의 구조에 대해 제안한다. 제안하는 트래픽 분류 시스템은 기존 방법인 각 시그니쳐를 하나씩 매칭하는 방법이 아닌, 트리 구조의 페이로드 시그니쳐 하나에서 Root 노드부터 매칭을 시작한다. Root 노드로 매칭이 된 플로우들을 선별하고, 다시 Root 노드의 바로 밑 자식노드를 매칭하는 방식으로 마지막 Leaf 노드까지 분석이 완료되었을 때, 트래픽 분류를 완료한다.

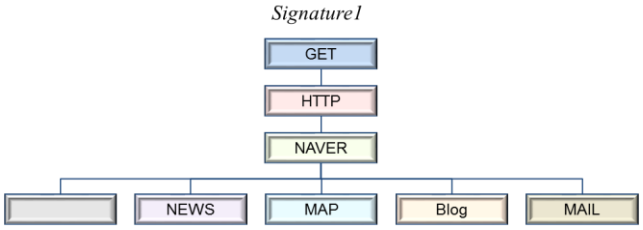

그림 3. 트리화 된 페이로드 시그니쳐 구조

이러한 시그니쳐와 트래픽의 매칭 방법은 기존 방법보다 더 세부적으로 응용 및 서비스를 분류할 수 있고, 비슷한 시그니쳐들을 기존 방식대로 처음부터 새롭게 분석하는 것이 아니기 때문에 분석 속도 측면 향상에도 많은 기여를 할 수 있다.

3. 결론 및 향후연구

본 논문에서는 시그니쳐 자동 생성 시스템에서 추출되는 페이로드 시그니쳐의 구조에 대해 제안했다. 기존 페이로드 시그니쳐 구조는 비슷한 문자열을 포함한 다른 시그니쳐로 추출되어 시그니쳐의 개수가 증가되고, 시그니쳐의 개수는 트래픽 분류 시스템에서 과부화의 원인이 되기 때문에 문제가 있었다. 그러나, 본 논문에서 이러한 시그니쳐 구조를 트리 구조로 변화함으로써 시그니쳐의 개수를 감소시키고, 그에 따른 트래픽 분류 속도도 감소시킬 수 있는 효과를 기대할 수 있다.

향후 본 논문에서 제안한 트리 시그니쳐 구조와 기존 시그니쳐 구조의 차이점에 대해 비교실험을 진행할 예정이다. 또한, 트리 구조로 자동으로 추출되는 페이로드 시그니쳐를 이용하여 기존 방법보다 더 효율적으로 트래픽을 분류할 수 있는 트래픽 분류 시스템 개발할 예정이다.

4. 참고 문헌

[1] 심규석, 구영훈, 이성호, Baraka D. Sija, 김명섭, " 최신 네트워크 응용 분류를 위한 자동화 페이로드 시그니쳐 업데이트 시스템", 통신학회 논문지 Vol.42 No.01, Jan. 2017, pp. 1-10

[2] 심규석, 김종현, 김성민, 김명섭, "급변하는 네트워크 트래픽을 효율적으로 분류하기 위한 응용 시그니쳐 자동 생성 시스템에 대한 연구", 2016 년도 정보과학회 동계종합학술발표회, 휘닉스파크, 강원, Dec. 21-23, 2016, pp.1021-1023

[3] 최지혁, 박준상, 김명섭, "시그니쳐 계층 구조에 기반한 HTTP 트래픽 분석 시스템의 처리 속도 향상", 통신학회 논문지 Vol.39B No.04, Apr. 2014, pp.191-199.