헤더 시그니쳐 기반 트래픽 분석 시스템 구조에 관한 연구

윤성호, 김명섭
고려대학교

{sungho_yoon,tmskim}@korea.ac.kr

A Study on the Architecture of Header Signature based Traffic Identification System

Sung-Ho Yoon, Myung-Sup Kim
Korea Univ.

요 약

정확한 트래픽 분석은 효과적인 네트워크 관리를 위해 반드시 선행되어야 한다. 이를 위해 다양한 분석 방법론이 제안되었지만, 실제 네트워크 관리에 활용하기에는 많은 한계점들이 존재한다. 헤더 시그니쳐는 특정 응용(서비스)을 지속적으로 서비스하는 서버의 헤더 정보를 사용하기 때문에 기존 분석 방법론의 한계점을 효과적으로 해결할 수 있다. 본 논문에서는 정확한 시그니쳐를 생성하고 급격히 증가하는 시그니쳐를 효과적으로 관리할 수 있는 헤더 시그니쳐 기반 트래픽 분석 시스템의 구조에 대해 설명한다.

Ⅰ. 서 론

인터넷의 보급과 다양한 서비스가 제공됨에 따라 인터넷 트래픽은 폭발적으로 증가하고 있다. 안정적인 서비스 제공과 망 관리 비용 최소화를 위해 다양한 네트워크 관리 정책의 수립이 요구되는 시점이다[1]. 따라서, 네트워크 트래픽의 응용을 분석하는 트래픽 분석은 다양한 네트워크 관리 정책을 적용하기 위해 반드시 필요한 선행 기술이다.

트래픽 분석을 위해 다양한 방법론들이 제안되었지만, 실제 네트워크 관리에 활용하기에는 많은 한계점들이 존재한다. 실제 네트워크 관리 장비에서 많이 사용되는 페이로드 기반 분석 방법은 특정 응용 트래픽에서 관찰되는 문자열을 시그니쳐로 생성하고, 해당 시그니쳐의 유무를 통해 트래픽을 분석한다. 하지만, 시그니쳐 생성 및 관리의 어려움, 높은 계산 복잡도, 사생활 침해, 실시간 제어의 어려움 등과 같은 많은 한계점 및 문제점을 가지고 있다. 이와 대조적으로 헤더 시그니쳐는 특정 응용(서비스)을 지속적으로 서비스하는 서버의 헤더 정보(IP, Port, Protocol)을 사용한다. 단순히 헤더 정보만을 비교하여 트래픽을 분석하기 때문에 기존 페이로드 시그니쳐의 한계점을 극복할 수 있다[2].

본 논문에서는 헤더 시그니쳐 기반 트래픽 분석을 위한 네트워크 구성과 분석 시스템 구조를 제안한다. 정확한 시그니쳐를 생성하기 위해 복수 생성 네트워크에서 생성한 시그니쳐를 중앙에 위치한 시그니쳐 DB 에서 통합 관리하며, 시그니쳐가 필요한 분석 네트워크에 배포한다. 또한, 급격히 증가하는 시그니쳐 중 의미 있는 시그니쳐를 유지하기 위해 분석 이력을 기반으로 갱신된 weight 값을 사용으로 시그니쳐를 관리한다.

본 논문은 다음과 같은 순서로 기술한다. 2 장에서는 헤더 기반 분석을 위한 네트워크 구성에 대해 설명하고, 3 장에서는 분석 시스템의 구조에 대해 설명한다. 마지막으로 4 장에서는 결론과 향후 연구를 언급한다.

Ⅱ. 네트워크 구성

본 장에서는 헤더 기반 분석을 위한 네트워크 구성에 대해 설명한다.

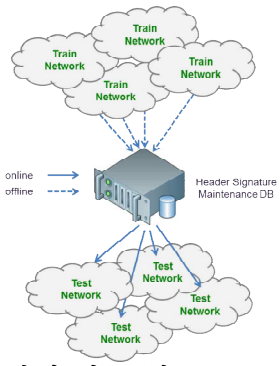

그림 1. 분석 시스템 네트워크 구성

그림 1 은 분석 시스템의 네트워크 구성를 나타낸다. 헤더 시그니쳐는 복수 네트워크에서 생성한 시그니쳐를 중앙에 위치한 시그니쳐 DB 에서 통합 관리한다. 또한, 생성된 시그니쳐는 분석 네트워크의 분석 시스템으로

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

배포된다. 시그니쳐 생성은 생성 네트워크의 다양한 상황을 고려하여 생성이 가능한 시점에 한해 오프라인으로 수행한다. 반면에, 시그니쳐 배포는 최신의 시그니쳐를 분석 네트워크에 적용해야 함으로 실시간으로 수행한다.

헤더 시그니쳐의 특성상, 특정 서버에서 제공되는 응용(서비스)이 변경될 수 있고, 생성 네트워크의 구성원, 지리적 위치 등과 같은 다양한 상황에 따라 특정 헤더 정보를 가지는 트래픽이 발생 될 수 있기 때문에 복수 네트워크에서 시그니쳐를 생성한다.

Ⅲ. 시스템 구조

본 장에서는 헤더 시그니쳐 기반 트래픽 분석 시스템의 구조에 대해 설명한다.

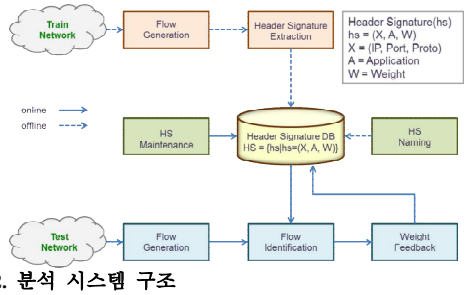

그림 2. 분석 시스템 구조

그림 2 는 본 논문에서 제안하는 시스템의 구조를 나타낸다. 분석 시스템은 크게 3 부분으로 구성된다. 생성 네트워크에서 시그니쳐를 생성하는 시그니쳐 생성부, 생성된 시그니쳐를 적용하여 트래픽을 분석하는 트래픽 분석부, 그리고 시그니쳐의 응용을 명명하고 관리하는 시그니쳐 관리부이다.

시그니쳐 생성부에서는 생성 네트워크에서 수집한 트래픽을 플로우 단위로 변경(Flow Generation)하고 시그니쳐를 생성(Header Signature Extraction)한다. 플로우는 5-tuple(SrcIP, SrcPort, DstIP, DstPort, Transport Layer Protocol)이 동일한 패킷의 집합을 의미한다. 헤더 시그니쳐는 X, A, W 로 구성된다. X 는 헤더정보(IP, Port, Protocol)를 의미하고, A 는 해당 시그니쳐로 분석된 트래픽에 기록 할 응용(서비스)을 의미한다. 그리고 W 는 시그니쳐 관리를 위해 사용하는 weight 를 의미한다. 생성된 시그니쳐는 헤더 시그니쳐 DB 에 저장된다.

트래픽 분석부는 분석 대상이 되는 분석 네트워크의 트래픽을 플로우 단위로 변경(Flow Generation)하고 시그니쳐 DB 에 저장된 시그니쳐와 헤더 정보 비교를 통해 트래픽을 분석(Flow Identification)한다. 즉, 동일한 헤더 정보(X)를 가지는 경우 해당 시그니쳐의 응용(A)을 트래픽에 기록한다. 또한, 시그니쳐 관리를 위해 분석에 사용된 시그니쳐의 사용 이력을 기반으로 해당 시그니쳐의 Weight(W)를 갱신(Weight Feedback)한다.

시그니쳐 관리부에서는 불필요한 시그니쳐를 삭제(HS Maintenance)하고 시그니쳐의 응용(A)를 명명(HS Naming)한다. 시그니쳐 DB 의 물리적 용량은 제한적이고, P2P 응용 트래픽의 Peer 와 같이 임시로 사용되는 HS 가 존재하기 때문에 시그니쳐의 weight(W)를 이용하여 분석에 사용될 가능성이 적은 시그니쳐를 삭제한다. 생성된 시그니쳐를 삭제하지 않고 영구적으로 유지할 경우 분석률은 상승하지만, DB 용량의 증가와 시그니쳐 개수 증가로 인한 분석 속도 저하의 문제점이 발생한다. Weight 는 분석 시 사용되는 시그니쳐의 사용 이력을 기반으로 갱신된다. 시그니쳐의 응용(A)은 다양한 분석 방법(페이로드, 통계, 에이전트 기반 등)을 통해 분석된 결과를 이용하여 응용을 명명한다.

좀 더 다양한 시그니쳐를 얻기 위해 시그니쳐의 생성과 응용(A)을 명명하는 부분을 분리하였다. 만약 응용을 알고 있는 트래픽만을 대상으로 시그니쳐를 생성할 경우, 시그니쳐의 추출 범위가 줄어든다. 따라서, 생성 네트워크의 모든 트래픽을 대상으로 시그니쳐의 헤더정보(X)를 생성하고 추후에 다양한 분석 방법의 결과를 사용하여 응용(A)을 명명함으로써 좀 더 많은 시그니쳐를 생성하여 트래픽 분석 성능을 향상시킨다.

Weight 는 해당 시그니쳐가 분석한 트래픽의 양(flow, packet, byte), 분석에 사용되지 않은 시간, 분석한 트래픽의 고유한 클라이언트의 수 등과 같은 다양한 측정 값들을 이용하여 설정하고 시그니쳐 관리에 사용한다. 간단한 관리 방법으로는 특정 시간 이상 분석에 사용되지 않은 시그니쳐를 삭제하는 방법이다[3]. 하지만, 이 방법은 응용 사용의 주기성 및 트래픽 특성을 고려하지 않기 때문에 최적의 시그니쳐를 유지하기 어렵다. 따라서 다양한 측정값들을 조합한 Weight 를 계산하는 관리 함수에 대한 연구가 필요하다.

Ⅳ. 결론 및 향후 연구

트래픽 분석은 효과적인 네트워크 관리를 위해 필수적인 선행 작업이다. 이를 위해 다양한 분석 방법론이 제안되었지만, 많은 한계점을 가지고 있다. 본 논문에서는 헤더 시그니쳐 기반 트래픽 분석을 네트워크 구성과 시스템 구조를 제안하였다.

복수 생성 네트워크에서 생성된 시그니쳐를 중앙 시그니쳐 DB 에서 통합 관리하고, 생성된 시그니쳐를 분석 네트워크에 배포하여 트래픽을 분석한다. 시그니쳐 생성과 응용 명명을 분리하여 시그니쳐 생성 범위를 확대하며, 급격히 증가하는 시그니쳐 중 의미 있는 시그니쳐를 관리하기 위해 분석 이력을 기반으로 갱신된 시그니쳐의 weight 을 사용한다.

향후 연구로는 다양한 측정값을 조합한 시그니쳐 weight 갱신 방법에 대한 연구와 실제 네트워크를 대상으로 해당 시스템을 구현하는 연구를 진행하겠다.

참 고 문 헌

[1] S. Sen, J. Wang, "Analyzing peer-to-peer traffic across large networks," in Proc. Internet Measurement Conference (IMC), pp. 137-150, 2002.

[2] Sung-Ho Yoon, Jin-Wan Park, Young-Seok Oh, Jun-Sang Park, and Myung-Sup Kim, "Internet Application Traffic Classification Using Fixed IP-port," Proc. of the Asia-Pacific Network Operations and Management Symposium (APNOMS) 2009, Jeju, Korea, Sep. 2009.

[3] M. Baldi, A. Baldini, N. Cascarano, and F. Risso, " Service-based traffic classification: Principles and validation" , Proc. of the IEEE 2009 Sarnoff Symposium, Princeton, NJ, USA, Mar. 2009.