인터넷 트래픽 분석을 위한 멀티 레벨 트래픽 분석 프레임워크 설계

윤성호, 박준상, 안현민, 김명섭

고려대학교 컴퓨터정보학과

{sungho_yoon, junsang_park, queen26, tmskim}@korea.ac.kr

요 약

인터넷 속도의 증가와 다양한 응용의 개발로 인해 인터넷 사용자와 이들이 발생시키는 인터넷 트래픽의 양이 급격히 증가하고 있다. 안정적인 인터넷 서비스를 제공하기 위해서는 정확한 트래픽 분석을 기반한 효과적인 네트워크 관리가 필요하다. 트래픽 분석을 위해 다양한 분석 방법론이 제안되었지만 모든 트래픽을 정확하게 분석하는 단일 방법론은 존재하지 않는다. 따라서 본 논문에서는 다양한 분석 방법을 결합하여 트래픽을 분석하는 멀티레벨 트래픽 분석 프레임워크를 제안한다. 본 프레임워크는 다양한 모델을 기반한 시그니쳐들을 동시에 적용함으로써 서로의 한계점을 보완하고 분석 결과의 정확도와 분석률을 향상시킬 수 있는 구조이다.

1. 서론

멀티미디어 응용의 개발과 인터넷 기술의 발전으로 인해 네트워크 트래픽은 지속적으로 증가하고 있다. 이러한 상황에서 인터넷 사용자가 안정적인 인터넷 서비스를 제공받기 위해서는 효율적인 네트워크 관리가 필요하다[1, 2]. 특히, 특정 트래픽을 차단하거나 조절하는 트래픽 제어 정책은 한정된 네트워크 대역폭을 모든 사용자가 공평하게 사용할 수 있게 해줄 뿐만 아니라 다양한 응용의 서비스 품질 (QoS)을 보장받을 수 있게 한다. 네트워크 정책은 트래픽의 원천을 분석하는 트래픽 분석 결과를 기반으로 수행된다. 트래픽 분석 방법론 또는 시스템의 최종목표는 분석하고자 하는 대상 네트워크에서 발생하는 트래픽을 모두 정확하게 빠른 속도로 분석하는 것이다.

트래픽 분석 방법론은 그 중요성에 따라 지속적으로 연구가 진행되어 현재에는 다양한 트래픽 분석 방법론이 존재한다. 하지만, 오늘날의 네트워크에는 다양한 응용 프로그램 및 서비스가 사용됨에 따라 복잡 다양한 트래픽이 발생하므로 하나의 단일 트래픽 분석 방법론으로는 네트워크 내의 모든 트래픽을 분석하기 어렵다. 따라서 최근에는 여러 가지 트래픽 분석 방법론을 통합한 멀티 레벨 트래픽 분석 방법론이 제안되고 있다. 멀티 레벨 트래픽 분석 방법론은 기존의 여러 트래픽 분석 방법론들 중 몇 가지를 통합하여 만든다. 다양한 트래픽 특징을 사용하는 방법론들을 통합하여 사용하기 때문에 각 방법론이 가지는 한계점을 보완하고 분석률과 정확도 측면의 분석 성능을 향상 시킬 수 있다.

본 논문에서는 다양한 시그니쳐 모델을 사용하는 멀티레벨 트래픽 분석 프레임워크를 제안한다. 시스템에 적용 가능한 시그니쳐 모델은 헤더, 페이로드, 통계, 행위 모델이다. 제안된 시그니쳐 모델은 본 연구진들이 기존에 제안한 시그니쳐 모델을 기반으로 정의되었으며, 시스템 확장성을 고려하여 향후 다양한 시그니쳐 모델을 추가할 수 있는 구조로 설계하였다. 본 프레임워크는 5 개의 세부 모듈로 구성되며, 트래픽 분석에 필수적인 시그니쳐 생성, 트래픽 분석, 결과 출력 모듈뿐만 아니라 분석기의 성능을 향상시키는 시그니쳐 관리와 네트워크 관리에 적용하기 위한 분석결과 활용 모듈을 포함한다.

본 논문은 다음과 같은 순서로 기술한다. 2 장에서는 기존에 제시된 트래픽 분석 관련 연구에 대해 설명하고, 3 장에서는 멀티레벨 트래픽 분석 프레임워크를 제시한다. 마지막으로 4 장에서는 결론과 향후 연구를 언급한다.

2. 관련 연구

인터넷 트래픽 분석은 그 중요성이 증가함에 따라 다양한 방법이 제시되고 있다. 가장 원시적인 포트 기반 분석은 Internet Assigned Number Authority (IANA)[3] 에 등록된 포트 정보를 사용하여 트래픽을 분석한다. 초기 인터넷에서는 포트 번호와 대응하는 서비스(HTTP(80), telnet(23), e-mail(25,110), FTP(20,21))가 트래픽 대부분을 차지하였기 때문에 이를 기준으로 신뢰성과 정확성이 높은 트래픽 분석 결과를 도출할 수 있었다. 하지만, 최근 사용되는 응용들은 방화벽 및 IPS 장비를 통과하기 위해 포트 번호를 임으로 변경하여 트래픽을 발생시키므로 더 이상 포트 번호가 특정 서비스, 프로토콜을 의미하지 않는다. 따라서 포트 기반 분류 방법론은 더 이상 정확한 트래픽 분석 결과를 보장하지 못한다[4].

페이로드 기반 분석은 패킷의 페이로드 내에서

이 논문은 2012 년 정부(교육과학기술부)의 재원으로 한국연구재단(2012R1A1A2007483) 및 2013 년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보.컴퓨팅기술개발사업(2010-0020728)의 지원을 받아 수행된 연구임.

응용마다 가지는 특정 스트링의 포함 유무를 통해 트래픽을 분석하는 방법이다[5]. 트래픽의 페이로드를 직접 검사하기 때문에 분석 성능(분석률, 정확도)은 매우 높지만, 시그니쳐 생성 및 관리의 어려움, 암호화 트래픽 분석 불가, 높은 계산 복잡도, 패킷 단편화 고려, 사생활 침해 등과 같은 많은 한계점을 가진다.

통계 기반 분석은 트래픽 내용을 보지 않고 패킷 및 윈도우 크기, 패킷 간 시간 간격 등과 같은 통계적 특징만을 이용하여 트래픽을 분석한다[6]. 이를 위해 패킷을 플로우라는 형태로 변환하여 사용한다. 플로우는 5-tuple(SrcIP, SrcPort, DstIP, DstPort, Transport Layer Protoco)이 동일한 패킷의 집합을 의미한다. 기존 페이로드 기반 분석 방법이 가지는 한계점을 해결할 수 있지만, 같은 엔진 기반의 응용, 또는 같은 응용 레벨 프로토콜을 사용하는 응용들의 경우 동일한 통계적 특징을 가지기 때문에 상세 응용 분석이 어렵다는 한계점을 가진다.

단일 분석 방법이 가지는 한계점을 보완하기 위해 다양한 분석 방법을 통합한 멀티레벨 분석 방법이 제안되었다[7, 8]. 제안된 논문에서 공통적으로 주장하고 있는 점은 여러 분석기를 통합할 경우 분석 성능(분석률, 정확도)이 향상되고, 서로의 한계점을 보완해 준다는 것이다.

분석기에 입력되는 트래픽의 형태에 따라 트래픽 분석에 사용할 수 있는 트래픽 특징이 다르다는 점을 이용하여 패킷 단위의 분석을 선행하고 분석되지 않은 트래픽에 한해 플로우 단위로 재 조합하여 분석을 수행한다. 플로우는 패킷들의 집합이기 때문에 상대적으로 다양한 트래픽 특징을 제공한다.

입력 데이터의 형태가 동일한 경우, 복잡도가 낮은 분석기를 우선 적용하는 직렬 수행 방법과 모든 분석기를 병렬 수행하는 방법이 있다. 직렬로 수행하는 방법은 앞서 수행된 분석기에서 트래픽이 분석된 경우 더 이상 분석을 진행하지 않기 때문에 적용하는 분석기의 순서가 분석 결과의 정확도에 영향을 미친다. 병렬로 수행하는 방법은 동일한 트래픽을 각각의 분석기가 다르게 분석할 경우 분석 결과를 통합하는 추가적인 과정이 요구되지만 분석 결과의 정확성은 향상 시킬 수 있는 장점이 있다.

기존에 제안된 멀티레벨 논문에서는 여러 분석 방법론들의 통합에만 초점을 맞추었다. 트래픽을 정확하게 분석하기 위해서는 분류체계, 정답지 트래픽 생성, 입력 데이터 정의, 시그니쳐 관리, 분석 결과 활용과 같은 다양한 관점의 고려가 필요하다.

3. 멀티레벨 분석 프레임워크

본 장에서는 인터넷 트래픽을 분석하기 위한 멀티레벨 분석 프레임워크에 대해 설명한다. 그림 1 은 본 논문에서 제안하는 멀티레벨 트래픽 분석 프레임워크(FORMULA: Framework for Multi-Level Application Traffic Identification)를 나타낸다. FORMULA 는 총 다섯 개의 세부 모듈들로 구성된다. 시그니쳐 생성부(Signature Constructor)는 분류체계를 정의하고 다양한 모델의 시그니쳐를 생성한다. 트래픽 분석부(Identifier)는 분석 대상 네트워크의 트래픽을 수집하여 시그니쳐 모델 별 세부 분석기를 통해 트래픽을 분석한다. 분석결과 출력부(Visualizer)는 분석된 트래픽을 시각화하고 분석기의 성능을 검증한다. 시그니쳐 관리부(Signature Maintainer)는 분석결과를 기반으로 시그니쳐를 관리한다. 마지막으로 분석결과 활용부(Utilizer)에서는 분석된 결과를 다양한 관점으로 추가 분석한다.

3.1 시그니쳐 생성부

트래픽 분석은 다른 응용과 구별되는 고유한 트래픽 특징인 시그니쳐의 존재 유무를 확인함으로써 수행된다. 시그니쳐를 생성하기 위해서는 분석 대상 응용을 선정하고 해당 응용이 발생하는 트래픽을 수집한다. 수집된 특정 응용의 트래픽은 다각적이고 계층적인 분류 체계 정보와 함께 시그니쳐 생성기의 입력으로 사용된다. 시그니쳐 생성기는 기 정의된 다양한 시그니쳐 모델을 기반으로 시그니쳐를 생성한다.

발생 원천을 알고 있는 정답지 트래픽 생성(Ground-truth Traffic Generator)은 분석 대상 응용의 트래픽을 수집하는 역할을 한다. 정답지 트래픽은 시그니쳐 생성뿐만 아니라 분석기 성능을 확인하기 위한 분석 결과의 검증 과정에서도 활용된다. 정답지를 생성하는 방법은 특정 응용만을 사용하여 트래픽을 수집하는 수작업 방법, 공개 프로토콜의 키워드를 페이로드 시그니쳐로 사용하는 DPI(Deep Packet Inspection) 방법, 트래픽을 발생시키는 종단 호스트에 소켓 정보를 수집하는 프로그램을 설치하는 에이전트 방법이 있다. 정답지 트래픽의 정확성과 수집의 용이성 측면에서 에이전트 방법이 가장 적절하다.

분류 체계 수립(Category Coordinator)은 분석 대상 응용의 다각적이고 계층적인 구조를 정의하는 것이다. 응용의 명확한 분류 체계는 분석 결과의 현황 파악을 용이하게 할 뿐만 아니라 분석 결과의 활용도 측면에서도 매우 효과적이다. 또한, 시그니 쳐 모델 및 방법론의 객관적 평가와 이들 간의 비교 평가를 가능하게 한다. 본 논문에서 제안하는 분류 체계는 하나 이상의 수평적인 분류 기준(서비스, 응용, 프로토콜, 기능)으로 구성되고, 각 분류 기준은 계층화된 분류 속성(대분류, 중분류, 소분류)을 가진다. 이로써 동일한 트래픽을 분류 기준 개수 측면의 다각적인 분석과 각 분류 기준 하위의 분류 속성에 따른 계층적인 분석이 가능하다. 본 분석 프레임워크에서 사용하는 분류 기준은 특정 목적을 위해 사용자에게 제공되는 서비스, 사용자가 사용하는 응용, 트래픽을 전송하기 위해 사용하는 프로토콜, 사용자가 발생하는 트래픽의 기능으로 정의한다.

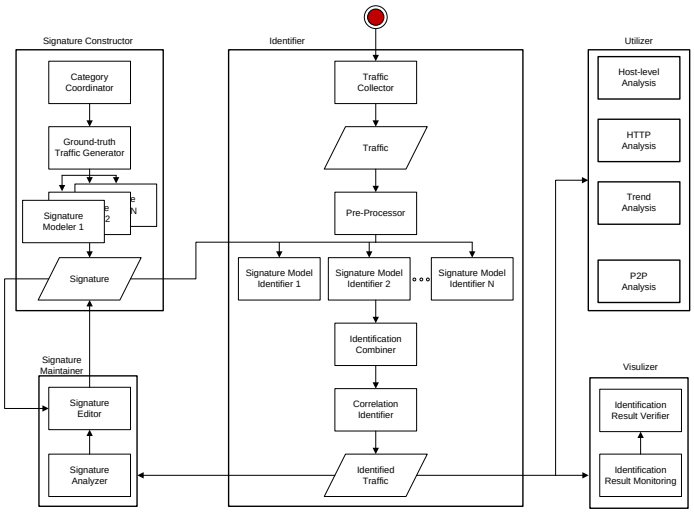

그림 1. 멀티레벨 트래픽 분석 프레임워크

본 논문에서 제안하는 분석 프레임워크는 다양한 형태의 시그니쳐 모델(Signature Modeler)을 지원한다. 응용이 발생하는 트래픽의 형태가 점점 복잡해지고, 네트워크 환경에 따라 수집 가능한 트래픽 특징이 상이하기 때문에 다양한 시그니쳐 모델이 적용되어야 한다. 본 시스템에서 적용된 시그니쳐 모델은 헤더, 페이로드, 통계, 행위이다. 헤더 시그니쳐 모델은 특정 응용 서버의 IP address 정보나 웹 서비스의 DNS 정보를 사용한다. 페이로드 시그니쳐 모델은 패킷의 IP 헤더 뒤에 위치한 페이로드의 일부분을 정규표현식으로 변환하여 사용한다. 통계 시그니쳐는 동일 세션에서 발생한 패킷의 순서와 크기 분포를 벡터로 변환하여 사용한다. 마지막으로 행위 시그니쳐는 특정 기능을 수행할 때 발생하는 고유한 패턴을 발생 순서와 시간 간격 정보와 함께 사용한다.

3.2 트래픽 분석부

분석기는 기 작성된 시그니쳐를 사용하여 분석 대상 네트워크에서 발생하는 트래픽을 분석한다. 효과적인 분석을 위해 수집된 트래픽을 분석에 적당한 형태로 재조합하고, 비정상 트래픽 분석과 운영체제 분석 등과 같은 분석 성능을 향상시키기 위한 전처리 과정을 거친다. 시그니쳐 모델에 특성화된 세부 분석기는 트래픽을 병렬 분석하며, 각 세부 분석기가 분석한 결과는 분석 결합자에 의해 통합된다. 통합된 분석 결과는 분석된 트래픽과 그렇지 못한 트래픽 간의 상관 관계를 이용하여 추가적인 분석을 수행한다.

트래픽 분석기는 다양한 방식으로 동작한다. 트래픽이 발생됨과 동시에 분석하는 실시간 방식, 일정 주기마다 수집된 트래픽을 분석하는 주기 방식, 저장 매체에 저장된 트래픽을 분석하는 오프라인 방식이 있다. 네트워크의 상황과 트래픽 분석의 목적에 따라 동작 방식을 선택할 수 있다.

트래픽 수집(Traffic Collector)은 분석 네트워크의 모든 패킷을 수집할 수 있는 지점, 즉 라우터나 스위치에서 포트미러닝 또는 탭핑을 통해 수집한다. 분석 네트워크에서 인터넷으로 연결된 경로가 여러 가지인 경우에는 모든 지점에서 수집하여야 정확한 분석이 가능하다. 수집된 패킷 단위의 트래픽은 플로우 단위로 재조합한다. 플로우 단위로 재조합된 트래픽은 질의(request)와 응답(response)과정의 트래픽이 한 단위로 구성되어 있기 때문에 더 많은 트래픽 정보를 제공하다. 따라서 다양한 시그니쳐 모델을 적용할 수 있게 한다. 또한, 메모리에 로드되는 트래픽의 양을 줄여 시스템의 부하를 줄일 수 있다. 본 시스템에서 사용하는 플로우는 양방향 플로우로써 5-tuple(source IP address/port, destination IP address/port, transport layer protocol)가 동일한 패킷들의 집합이다. 플로우는 시작 시간, 종료 시간, 패킷 개수, 바이트 크기, TCP 플래그 패킷의 개수, 데이터 패킷의 개수 등의 통계 정보뿐만 아니라 페이로드 정보까지 포함한다. 페이로드 정보는 플로우를 구성하는 초기 N 개의 데이터 패킷 페이로드를 의미하며 메모리 공간의 한계와 시그니쳐 모델의 적용 범위에 따라 N 의 값은 조절 가능하다. 본 시스템에서는 이를 flow_with_packet 이라 지칭하며, 플로우 생성 시간과 접근 시간을 최소화 하기 위해 해쉬 형태의 메모리 구조로 저장한다.

전처리 과정(Pre-Processor)은 분석 속도 향상과 정확도를 높이기 위해 분석기의 입력 데이터를 최적화 시키고 트래픽 발생 호스트의 부가적인 정보를 제공한다. 대표적인 전처리 과정으로써, 비정상 트래픽 판별, 호스트 운영체제 판별, 유무선 호스트 판별, NAT 공유기 판별 등이 있다. 본 논문에서는 비정상 트래픽 판별과 호스트 운영체제 판별에 대해 소개한다. 악의적인 목적을 가진 네트워크 공격이 활발해짐에 따라 비정상 트래픽의 양이 증가하고 있다. 이렇게 발생된 트래픽은 분석 프레임워크의 입력 데이터 양을 증가시켜 분석기의 부하를 증가시킨다. 본 시스템은 비정상으로 TCP 연결을 시작하거나 종료하는 트래픽을 비정상 트래픽으로 분

석하여 분석기의 입력에서 제외한다. 즉, 3-핸드셰이크(SYN-SYN/ACK-ACK)로 연결을 시작하지 않고나 4-핸드셰이크(FIN-ACK-FIN-ACK) 또는 3-핸드셰이크(FIN-FIN/ACK-ACK)로 종료하지 않는 경우 해당 트래픽을 비정상 트래픽으로 분석한다. 다양한 형태의 단말이 사용되면서 이를 운영하는 운영체제 또한 매우 다양해졌다. 응용은 운영체제 위에서 동작하기 때문에 운영체제 마다 지원하는 응용과 응용의 버전이 상이하다. 트래픽 분석에 있어 호스트 운영체제 정보는 적용해야 할 시그니쳐의 개수를 줄여 시스템 부하를 줄이고, 분석 결과의 정확성을 향상시킬 수 있다. 예를 들어 안드로이드 운영체제 사용자로 판명된 호스트에서 발생되는 트래픽은 안드로이드에서만 동작하는 응용의 시그니쳐를 적용시킨다. 이를 위해 트래픽을 호스트별로 구분하고 해당 트래픽의 헤더 정보를 기반으로 운영체제를 판별한다. 이때 사용하는 정보로는 TTL, Nop count, time stamp, SACK, window size scale, window size, MSS 등이 있다.

시그니쳐 모델 분석기(Signature Model Identifier)는 시그니쳐 모델에 따른 세부 분석기들로 구성된다. 시그니쳐 모델에 따라 트래픽을 분석하는 방법이 상이 하기 때문에 이를 사용하는 분석기 또한 개별적으로 동작되어야 한다. 예를 들어 헤더 시그니쳐 모델의 경우 헤더 정보(IP address, port number, L4 protocol)만을 사용하기 때문에 트래픽의 헤더 정보를 검색하는 기능만을 분석기에 적용하고 페이로드 시그니쳐 모델의 경우 플로우에 포함된 페이로드 정보를 검색하는 기능을 추가적으로 적용한다. 또한, 통계 시그니쳐 모델의 경우 플로우를 구성하는 패킷들의 방향과 크기를 비교하는 기능을 적용한다. 시그니쳐 모델에 따라 분석할 수 있는 분류 기준이 다르기 때문에 각 세부 분석기들은 병렬로 수행된다. 즉, 특정 플로우가 하나의 세부 분석기에 의해 분석이 된 경우에도 다른 모든 분석기에 적용한다. 이를 통해 다각적이 트래픽 분석을 가능하게 한다.

시그니쳐 모델 분석기의 개별 분석기를 통해 분석된 트래픽은 분석 결합자(Identification Combiner)에 의해 통합된다. 시그니쳐 모델 마다 분석 가능한 분류 기준이 다르기 때문에 세부 분석기의 결과를 통합하여 트래픽 당 모든 분류 기준을 명시하도록 한다. 만약 동일한 분류 기준에 서로 다른 분석기가 다른 결과로 분석하면, 즉 분석 결과의 충돌이 발생하면, 결합 알고리즘을 통해 하나의 결과를 선택한다. 본 시스템에서는 분석 결과의 빈도와 우선순위 기준으로 결합 알고리즘을 수행한다. 분석 결과 충돌이 발생하면 가장 많이 분석한 결과를 선택한다. 만약, 모든 세부 분석기의 결과가 다르면 사전에 정의한 분석기 우선 순위에 따라 최종 분석 결과를 선택한다.

상관관계 분석기(Correlation Identifier)는 분석 결합자에 의해 통합된 분석 결과를 입력 받아 분석된 트래픽과 그렇지 않은 트래픽 간의 상관관계를 파악하고 이를 통해 분석되지 않은 트래픽을 분석한다. 구체적으로 일부 분석된 결과를 입력 받아 트래픽에 나타나는 여러 정보 중 연관성을 가지고 있는 특정 속성 값을 기준으로 트래픽을 그룹화하고 분석된 트래픽이 존재하는 경우 해당 그룹에 속한 전체 트래픽을 분석된 트래픽과 동일한 결과로 분석한다. 본 시스템에서는 상관관계 분석을 위해 4 가지 알고리즘을 사용한다. 공통된 3-tuple(IP address, port, transport layer protocol)을 사용하는 트래픽을 분석하는 서버-클라이언트 방법, 특정 시간에 단일 호스트에서 발생한 트래픽을 분석하는 발생 시간 방법, 특정 두 호스트에서 발생한 트래픽을 분석하는 호스트-호스트 방법, 마지막으로 유사한 통계정보를 가지는 트래픽을 분석하는 통계방법이 있다.

3.3 분석결과 출력부

분석 결과 출력부는 분석기에 의해 분석된 트래픽의 분석 결과를 다양한 관점으로 시각화하고 분석 과정에서 측정된 다양한 정보를 제공한다. 이와 함께 분석된 트래픽의 정확도를 판별함으로써 분석에 사용된 시그니쳐의 타당성과 분석기의 성능을 검증한다. 본 모듈의 결과는 트래픽 분석 즉시 Web을 통해 시그니쳐 생성자, 네트워크 관리자, 분석기 운영자에게 보고된다.

분석 결과 측정(Identification Result Monitoring)은 분석 결과를 다양한 그래프와 표를 활용하여 한 눈에 확인할 수 있도록 한다. 결과를 측정하기 위해 사용한 척도는 Completeness 이다. 본 척도는 전체 트래픽 중 분석된 트래픽의 비율을 의미한다. 분석된 트래픽의 결과는 다각적이고 계층적인 분류 체계를 기반으로 분석되었기 때문에 이를 파악할 수 있는 트리 구조로 설계된다. 또한, 시간의 흐름에 따라 트래피 분석 결과의 추이를 확인하기 위하여 하루, 주, 월 단위의 분석 결과를 제공한다. 부가적으로 분석 프레임워크의 하드웨어적 성능을 파악하기 위하여 CPU 사용률, 메모리 사용량, 분석 시간 등 부가적인 정보를 제공한다.

분석 결과 검증(Identification Result Verifier)은 정답지 트래픽을 통해 분석 결과를 검증한다. 정답지 트래픽의 생성은 시그니쳐 생성부에서 설명한 에이전트 방법을 사용함으로써 분석 즉시 결과를 검증할 수 있다. 본 모듈에서 검증을 위해 사용한 척도는 분석된 트래픽 중 정확한 트래픽의 비율을 나타내는 Overall_accuracy 와 응용 또는 시그니쳐 별 성능을 측정하는 Precision 과 Recall, 그리고 이 두 척도를 하나의 수치로 표현한 F-Measure 가 있다. Precision 은 얼마나 정확히 분석하는지를 측정하고 Recall 은 얼마나 많이 트래픽을 분석하는지를 측정한다. 앞서 제시한 척도는 트래픽에 포함된 응용의 비율에 따라 편향된 결과를 나타낼 수 있다. 따라서 응용 단위 별 실제 분석 수량을 보여주는 Confusion_matrix 를 같이 제공한다.

3.4 시그니쳐 관리부

시그니쳐 관리부는 분석결과 출력부로부터 전송받은 분석 결과 검증 정보를 토대로 시그니쳐 분석 성능과 상태를 분석하고 이를 기반으로 분석기에 적용한 시그니쳐 목록을 개편한다.

시그니쳐 분석기(Signature Analyzer)는 시그니쳐 성능을 수량적인 수치로 표현하기 위하여 5 가지 관리 요소를 사용한다. 시그니쳐의 분석량을 의미하는 시그니쳐 분석도, 시그니쳐의 사용 횟수를 의미하는 시그니쳐 빈도, 단위 시간 동안 대상 시그니쳐에 의해 분석된 트래픽 중 정확하게 분석한 트래픽 비율을 의미하는 시그니쳐 정확도, 대상 시그니쳐와 해당 시그니쳐가 분석하는 응용의 마지막 사용 시간의 차를 의미하는 시그니쳐 유효도, 시그니쳐 간 Association Rule 에 기반한 상관관계를 의미하는 시그니쳐 상관관계도이다.

시그니쳐 편집기(Signature Editor)은 시그니쳐 분석기에 의해 분석된 시그니쳐 관리 요소에 의거하여 시그니쳐를 개편한다. 시그니쳐 분석도와 빈도를 기준으로 자주 사용하고 많은 트래픽을 분석하는 시그니쳐를 시그니쳐 목록의 앞쪽으로 이동시키고, 시그니쳐 정확도, 유효도, 상관관계도를 기준으로 분석 성능을 저해하는 시그니쳐를 목록에서 제거한다.

3.5 분석결과 활용부

분석기에 의해 분석된 결과는 단순히 분류체계에 따른 각 분류 기준 별 트래픽 양만을 나타내기 때문에 이를 네트워크 관리에 활용하기 위한 추가적인 작업이 필요하다. 분석결과 활용부에서는 네트워크 관리자, 서비스 제공자, 인터넷 사용자들의 요구에 부응하여 다양한 분석을 제공한다. 본 시스템에서는 트래픽을 발생시킨 호스트를 기준으로 트래픽 발생 현황을 분석하는 호스트 기반 분석, 인터넷 서비스 사용 실태를 분석하는 경향 분석, 인터넷 트래픽의 대부분을 차지하는 web 트래픽을 분석하는 HTTP 분석, 복잡한 형태의 트래픽을 발생하는 P2P 응용을 분석하는 P2P 분석을 제공한다. 앞서 기술된 4 가지 분석 이외에도 트래픽 활용을 위한 다양한 분석이 추가될 수 있다.

4. 결론 및 향후 연구

다양한 응용의 사용과 인터넷 기술의 발전으로 인해 다양한 분석 방법을 동시에 적용할 수 있는 멀티레벨 트래픽 분석기에 대한 요구가 증가하고 있다. 본 논문에서는 다양한 모델의 시그니쳐를 사용하여 트래픽을 분석하는 시스템 구조를 제안한다. 시그니쳐 생성부, 트래픽 분석부, 분석결과 분석부와 같은 기본적인 모듈뿐만 아니라 분석 결과의 정확성을 향상시키고 분석기 부하를 줄일 수 있는 시그니쳐 관리부, 분석된 결과를 네트워크 관리에 효과적으로 활용할 수 있는 분석결과 활용부를 제안한다.

향후 연구로는 본 논문에서 제안한 시스템 설계를 기반으로 실제 네트워크에서 운영되는 시스템을 구축할 계획이다. 또한, 본 시스템에 적용할 수 있는 다양한 시그니쳐 모델을 찾고 이를 적용하겠다.

5. 참고 문헌

[1] H. Kim, K. C. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee, "Internet traffic classification demystified: myths, caveats, and the best practices," in Proceedings of the 2008 ACM CoNEXT conference, 2008, p. 11.

[2] A. Dainotti, A. Pescape, and K. C. Claffy, "Issues and Future Directions in Traffic Classification," Ieee Network, vol. 26, pp. 35-40, Jan-Feb 2012.

[3] IANA port number list. Available: http://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xml

[4] A. W. Moore and K. Papagiannaki, "Toward the accurate identification of network applications," in Passive and Active Network Measurement, ed: Springer, 2005, pp. 41-54.

[5] Y. Wang, Y. Xiang, W. L. Zhou, and S. Z. Yu, "Generating regular expression signatures for network traffic classification in trusted network management," Journal of Network and Computer Applications, vol. 35, pp. 992-1000, May 2012.

[6] N. F. Huang, G. Y. Jai, H. C. Chao, Y. J. Tzang, and H. Y. Chang, "Application traffic classification at the early stage by characterizing application rounds," Information Sciences, vol. 232, pp. 130-142, May 2013.

[7] G. Szabó, I. Szabó, and D. Orincsay, "Accurate traffic classification," in World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a, 2007, pp. 1-8.

[8] A. Callado, J. Kelner, D. Sadok, C. Alberto Kamienski, and S. Fernandes, "Better network traffic identification through the independent combination of techniques," Journal of Network and Computer Applications, vol. 33, pp. 433-446, 2010.