KIPS Tr. Comp. and Comm. Sys.
Vol.3, No.2 pp.47~56 pISSN: 2287-5891

http://dx.doi.org/10.3745/KTCCS.2014.3.2.47

Study on Classification Scheme for Multilateral and Hierarchical Traffic Identification

Sung-Ho Yoon† · Hyun-Min An†† · Myung-Sup Kim†††

ABSTRACT

Internet traffic has rapidly increased due to the supplying wireless devices and the appearance of various applications and services. By increasing internet traffic rapidly, the need of Internet traffic classification becomes important for the effective use of network resource. However, the traffic classification scheme is not much studied comparing to the study for classification method. This paper proposes novel classification scheme for multilateral and hierarchical traffic identification. The proposed scheme can support multilateral identification with 4 classification criteria such as service, application, protocol, and function. In addition, the proposed scheme can support hierarchical analysis based on roll-up and drill-down operation. We prove the applicability and advantages of the proposed scheme by applying it to real campus network traffic.

Keywords : Traffic Classification, Traffic Identification, Classification Scheme

다각적이고 계층적인 트래픽 분석을 위한 트래픽 분류 체계에 관한 연구

윤 성 호† · 안 현 민†† · 김 명 섭†††

요 약

인터넷을 기반으로 하는 다양한 서비스 및 응용의 등장과 무선 디바이스의 보급은 인터넷 트래픽을 급격하게 증가시켰다. 인터넷 트래픽의 급격한 증가로 한정적인 네트워크 자원을 효율적으로 사용하기 위해 인터넷 트래픽 분석의 중요성이 증가하고 있다. 하지만 트래픽 분석 방법론에 비해 분석 결과를 체계적으로 관리하는 분류 체계에 대한 연구는 이루어지지 않고 있다. 본 논문에서는 다각적이고 계층적인 트래픽 분석을 위한 분류 체계를 제안한다. 제안하는 분류 체계는 서비스, 응용, 프로토콜, 기준의 4가지 분류 기준을 사용하여 다각적으로 분석이 가능하며, 분류 기준 별로 계층화된 속성을 가지고 있어 결과의 통합화 및 세분화가 가능하다. 논문에서는 제안한 분류 기준을 실제 학내 망에 적용하여 분석함으로 분류 체계의 장점과 활용성을 보인다.

키워드 : 트래픽 분류, 트래픽 분석, 분류 체계

1. 서 론

인터넷을 기반으로 하는 다양한 응용 및 서비스들의 등장과 무선 디바이스의 빠른 보급으로 인해 인터넷 트래픽은 급격하게 증가하고 있다. 응용 프로그램을 사용자의 PC에 설치하지 않고 인터넷을 통해 프로그램을 실제 PC에서 사용하는 것처럼 하는 서비스들과, 무선 스마트 디바이스와 PC의 동기화를 제공하는 클라우드 기반 응용들의 개발은 인터넷 트래픽양의 증가를 가속화시키고 있다. Cisco에서는 2015년 IP 트래픽이 2010년에 비해 약 4배 증가하는 242.4 엑사바이트에 이를 것으로 전망하였고, 그 중 비즈니스 IP 트래픽을 제외한 순수 인터넷 트래픽은 182.4 엑사바이트에 이를 것으로 전망하였다[1]. 트래픽의 양이 급격하게 증가함에 따라서, 예상되는 트래픽 대란과, 한정적인 네트워크 자원의 효율적인 사용을 위한 트래픽 분석 중요성이 증대되고 있다. 또한, 다양한 서비스 중에서 사용자의 기호 및 사회 인터넷 트렌드 분석도 중요해지고 있어, 트래픽 분석이 다양한 목적으로 사용되고 있다.

트래픽 분석의 결과는 네트워크 관리 분야에서 네트워크

※ 본 연구는 BK21플러스 사업의 지원을 받아 수행된 결과임.
† 준 회 원 : 고려대학교 컴퓨터정보학과 박사과정
†† 준 회 원 : 고려대학교 컴퓨터정보학과 석사과정
††† 종신회원 : 고려대학교 컴퓨터정보학과 교수
논문접수 : 2013년 9월 2일
수 정 일 : 1차 2013년 11월 15일, 2차 2013년 12월 30일
심사완료 : 2014년 1월 21일
* Corresponding Author : Myung-Sup Kim(tmskim@korea.ac.kr)

사용 현황 파악과 확장 계획 수립 등의 다양한 분야에서 활용될 수 있다. 대표적인 실제 활용 예는 QoS(Quality of Service) 정책 설정이다. 실시간 음성 및 영상 데이터 트래픽은 기존 일반 데이터와 달리 일정 수준 이상의 대역폭을 확보하여야 한다. 즉, 네트워크 관리자는 트래픽의 종류와 특성을 파악하고 우선순위와 대역폭 정책을 결정하여야 한다. 이를 위해서는 관리 네트워크에서 발생되는 트래픽을 다양한 관점으로 분석하는 것이 선행되어야 한다. 이외에도 트래픽 분석 결과는 종량제 과금, CRM, SLA, 보안 분석 등 다양한 곳에서 활용 가능하다.

이와 같은 트래픽 분석 중요성의 증가와 다양해지는 사용 목적에 따라 다양한 트래픽 분석 방법론 및 실시간 분석을 위한 성능 향상 방법론, 대용량 트래픽의 효과적 처리 방법론 등 다양한 주제로 트래픽 분석 방법론들이 연구되었다. 하지만, 트래픽의 분석 결과를 효과적이고 체계적으로 활용하고 관리하는 방법론에 대해서는 연구가 이루어지지 않고 있는 실정이다.

명확하지 않은 분류 체계는 다음과 같은 문제점들을 야기한다. 첫째, 트래픽 분석 방법론의 객관적 평가와 각 방법론 간의 비교 평가가 불가능하다. 심지어 동일한 분석 방법론일지라도 적용되는 분류 체계에 따라 평가가 달라질 수 있다. 둘째, 모호한 분류 체계로 분석된 결과는 명확한 현황 파악이 어려울 뿐만 아니라 실제 네트워크 관리에 활용하기 힘들다. 따라서 명확한 분류 체계에 대한 연구가 필요하다.

명확한 분류 체계의 확립은 앞에서 언급한 두 가지의 문제점을 해결할 뿐 아니라, 다양한 이점을 가진다. 첫째, 분석 방법론의 정교한 개발을 가능하게 한다. 명확한 분류 체계를 분석 방법론에 적용할 경우, 객관적인 결과를 얻을 수 있고, 더 정교한 분석 방법론의 개발이 가능해진다. 둘째, 다각적인 트래픽 분석을 가능하게 한다. 이를 통해 트래픽에 대한 이해를 높일 수 있고, 다양한 측면에서의 트래픽 분석이 가능해진다. 또한 사용자의 요구에 따라 다양한 측면의 결과 활용이 가능해진다. 셋째, 결과를 활용하는 사람과 관리하는 사람 모두에게 정확한 이해와 사용의 편리성을 제공할 수 있다.

본 논문에서는 분류 체계(Scheme)의 구조를 정의하고, 정의된 분류 체계 구조에 분류 기준(Dimension)과 요소(Attribute)들을 제안한다. 또한, 제안하는 분류 체계를 실제 분석 결과에 적용하여 그 활용성을 보인다. 본 논문에서 제안하는 분류 체계는 계층화된 분류 속성을 가지는 다양한 분류 기준을 사용하여 다각적 측면의 분석을 가능하게 하여 분석 결과의 활용도를 높인다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 분석 방법론 및 기존의 분류 사례에 대해 언급하고, 3장에서는 분류 체계의 구조에 대해 정의한다. 4장에서는 분류 체계 구조에 실제 분류 기준과 계층화된 속성들을 제안하고, 제안하는 기준에 대한 타당성 및 효과를 보이고, 5장에서는 실제 트래픽에 분류 체계를 적용하여 분석한 결과를 통해 제안하는 분류 체계의 활용도를 보인다. 마지막으로 6장에서는 결론 및 향후 연구를 제시한다.

2. 관련 연구

본 장에서는 기존 상용 네트워크 장비와 분석 방법론을 제안하는 논문들에서 사용하는 기존의 분류 체계들에 대해 간략히 살피고, 본 논문에서 사용한 분류 체계의 적용 실험에 사용되는 분석 방법론에 대해 살펴본다.

2.1 상용 트래픽 분석 장비의 분류 체계

선행 연구[2]에서는 네트워크 관리 분야에서 많이 사용되는 몇몇 상용 네트워크 트래픽 분석 장비(CheckPoint, Fortinet, Paloalto)의 분류 체계(Scheme)를 정리하고 분석하여 요구사항을 도출하였다. 정확한 분류 체계를 위해 분류 기준(Dimension)과 요소(Attribute)를 정의하였다.

CheckPoint[3]와 Fortinet[4]은 “응용(Application)”의 단일 기준을 사용하여 트래픽을 분류하고, 응용 형식으로 각 응용에 대한 그룹을 생성함으로 총 2계층으로 구성된 계층화된 속성(Layered-attribute)을 적용하여 트래픽을 분류한다. 그 외에 추가적인 정보(응용의 위험도, 트래픽 발생 형태, 개발 회사 등)에 대해 Tag기능을 사용하여 개별 응용에 대한 특징, 부가 설명을 명시한다. 이 방법은 사용자에게 Tag 기능을 사용하여 다양한 정보를 제공할 수 있지만, 정보에 대한 접근이 어렵다는 단점이 있다.

Paloalto[5]는 응용(Application)과 기능(Function)이 혼합된 단일 기준을 사용하여 트래픽을 분류한다. 응용 형식을 2계층으로 구성하고 3계층에 응용과 기능의 혼합 요소를 사용하였다. 응용과 기능의 혼합 요소는 예를 들어, NateOn이 하나의 요소가 되는 것이 아니라, NateOn이 제공하는 다른 기능들과 혼합되어 NateOn-Chat, NateOn-File Transfer 등의 여러 요소로 나누어져 사용되는 형식이다. 이 방법은 3계층으로 정의된 분류 속성을 사용하여 응용에 대한 체계적 접근을 제공하지만, 응용과 기능이 혼합된 기준을 사용하여 응용과 응용이 제공하는 기능의 관계 정립이 어렵다는 단점이 있다.

2.2 분석 방법론

트래픽 분석 방법론은 그 중요성이 증가함에 따라 지속적으로 연구가 진행되고 있다. 트래픽 분석 방법들은 트래픽 분석 시 사용하는 트래픽 특징을 기준으로 포트기반 분석[6, 7], 페이로드기반 분석[8, 9], 통계정보기반 분석[10-12], 상관관계기반 분석[13] 등으로 구분된다.

포트기반 분석은 Internet Assigned Number Authority (IANA)[6]에서 지정한 포트 정보를 이용한다. 포트 번호와 대응하는 서비스를 기준으로 분석하기 때문에 시스템 구현 시 오버헤드가 작다는 장점이 있다. 하지만, 최근 사용되는 응용들은 방화벽 및 IPS 장비를 통과하기 위해 포트 번호를 임의 또는 동적으로 설정하여 트래픽을 발생시키므로 더 이상 포트 번호가 특정 서비스, 프로토콜을 의미하지 않는다.

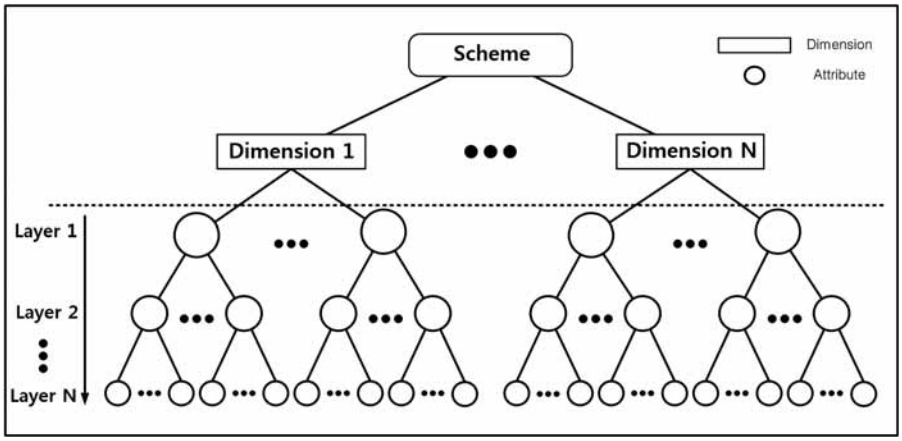


Fig. 1. Overall Structure of Classification Scheme

이러한 문제를 해결하기 위해 패킷의 페이로드 내에서 응용마다 가지는 특정 스트링(시그니쳐)의 포함 유무를 통해 트래픽을 분석하는 페이로드기반 분석 방법이 제안되었다. 트래픽의 내용을 확인하기 때문에 분석 성능(분석률, 정확도)이 매우 높지만, 시그니쳐 생성 및 관리, 암호화 트래픽, 높은 계산 복잡도, 패킷 단편화, 사생활 침해 등과 같은 많은 한계점을 가지고 있다[14].

트래픽 암호화 및 사생활 침해 문제를 해결하기 위해 트래픽 내용을 보지 않고 패킷 및 윈도우 크기, 패킷 간 시간 간격 등과 같은 통계적 특징만을 이용한 통계 기반 분석 방법이 제안되었다. 이 방법론은 패킷의 헤더 정보를 통해 통계 정보를 생성하므로 기존 트래픽 분류 방법론들의 한계점들을 보완할 수 있다. 하지만, 같은 엔진 기반의 응용이거나 같은 응용 레벨 프로토콜을 사용하는 경우 동일한 통계적 특징을 가지기 때문에 상세한 응용 별 분석이 어려운 한계점을 가진다.

최근에는 전통적인 트래픽 분석 방법의 한계점을 보완하기 위해 패킷 단위의 트래픽을 플로우 단위로 변경하고 이들의 상관관계를 분석하는 방법이 제안되었다. 플로우는 5-tuple (SrcIP, SrcPort, DstIP, DstPort, Transport Layer Protocol)이 동일한 패킷의 집합을 의미한다. 플로우의 크기, 기간 등과 같은 통계 정보와 플로우들 간의 연결 형태를 이용하여 트래픽을 분석한다. 패킷기반 분석 방법 보다 다양한 특징을 사용할 수 있기 때문에 다양한 분석이 가능하지만, 플로우 생성이 완료될 때까지 분석하지 못하며, 플로우의 통계 정보를 계산하는 오버헤드가 발생한다. 또한, 유사한 통계 정보를 가지는 응용 간 구별이 어려운 문제점을 가지고 있다.

3. 제안하는 분류 체계

본 장에서는 다각적으로 트래픽을 분석할 수 있는 트래픽 분류 체계의 기본 구조와, 효과적으로 트래픽을 분석/분류할 수 있는 분류 기준 및 속성에 대해 제안하고, 제안하는 분류 기준에 대한 타당성 및 적용 시 얻을 수 있는 효과에 대해 기술한다.

3.1 분류 체계 구조

본 장에서는 다양한 목적의 트래픽 분류에 적용가능하며, 트래픽 분류 결과를 다각적 측면과 계층적으로 활용할 수 있는 분류 체계를 제안한다.

제안하는 분류 체계는 그림 1과 같은 트리 구조로 이루어져 있다. 제안하는 분류 체계의 구조는 분류 체계라는 root(Scheme) 노드를 중심으로 구성된다. 분류 체계의 분류 기준(Dimension)을 기준으로 root 노드의 자식 노드 개수가 정해진다. 이는 동일한 트래픽을 분류 기준 개수 측면으로 분석함으로써 다각적인 분류가 가능하게 한다.

각 분류 기준들은 자식 노드로서 계층화된 분류 속성들을 가진다. 분류 속성에서 계층을 정의할 때, root 분류 체계와 그 자식노드인 분류 기준을 제외하고 계층의 수를 측정한다. 따라서 하나의 분류 기준(root의 자식 노드)의 하위 단이 분류 속성에서의 1계층으로 구성된다. 분류 계층은 사용 목적이나, 각 요소의 그룹화 방식에 따라 계층의 값과 정의가 달라질 수 있다.

계층화된 분류 속성을 사용하므로 분석된 결과를 세분화(Drill-Down)하거나 통합(Roll-up)하여 결과를 살펴볼 수 있어, 사용자가 원하는 결과를 다각적이고 계층적으로 제공할 수 있다. 각 분류 속성들과 분류 기준은 분류 결과에 따른 측정값(measured value)을 원자 값으로 가진다. 각 분류 속성들은 다양한 단위의 원자(Atomic) 값을 가질 수 있지만, 하나의 root 안에서의 원자 값의 단위는 통일되어야 한다. 예를 들어, 하나의 분류 체계가 byte 단위의 트래픽 측정 결과 값과 packet 단위의 트래픽 측정 결과 값을 동시에 지원할 수 없다. 또한, 각 분류 기준에 속하는 속성들에 아래와 같은 규칙을 적용하므로 분석 결과의 충돌 및 중복을 허용하지 않는다. 첫째, 계층 N의 분류 속성 노드들의 원자 값의 합은 계층 N-1의 분류 속성 노드들의 원자 값의 합과 동일해야 한다. 둘째, 계층 1의 분류 속성 노드들의 원자 값의 합은 분석 대상으로 입력된 데이터의 본 값과 동일해야 한다.

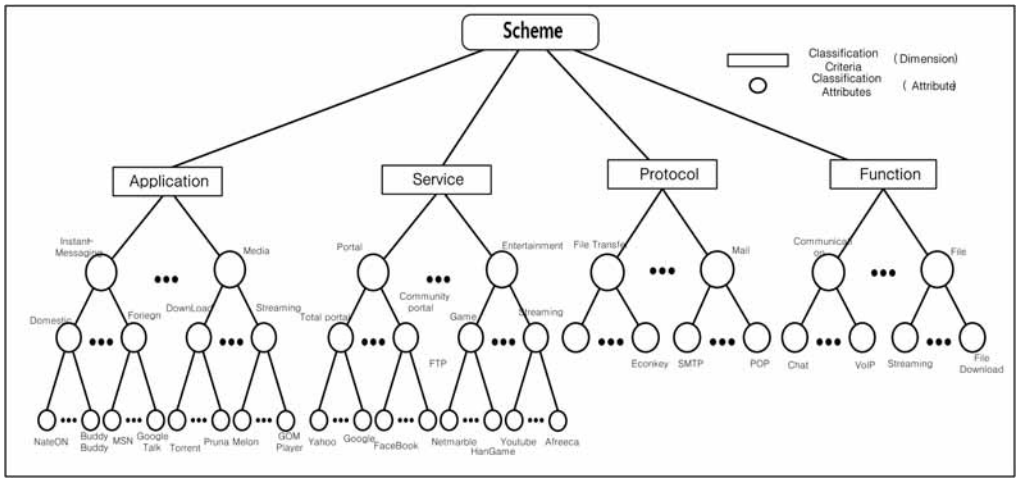


Fig. 2. Four Criteria of Classification Scheme

3.2 분류 기준의 제안 및 타당성 증명

본 장에서는 다양한 목적의 다각적 트래픽 분류에 효과적으로 사용될 수 있는 4가지 분류 기준을 제안하며 각 기준별 타당성을 보인다.

제안하는 분류 기준은 그림 2에서 표현한 것과 같이 응용(Application), 서비스(Service), 프로토콜(Protocol), 기능(Function)의 4가지의 다각적 분류 기준이다. 각 기준의 정의는 표 1과 같다.

Table 1. Definition of Four Classification Criteria

분류 기준	정의
서비스	특정 목적을 위해 사용자에게 제공되는 서비스
응용	사용자가 사용하는 응용 프로그램
프로토콜	트래픽을 전송하기 위해 사용하는 프로토콜
기능	사용자가 발생하는 트래픽의 목적

제안하는 서비스, 응용, 프로토콜, 기능의 4가지 분류 기준의 사용은 다양한 목적의 트래픽 분석에 대해 다각적 분석 결과를 제공하므로 활용도가 높고, 트래픽에 대해 정확한 정보를 제공할 수 있다. 또한 제안하는 각각의 기준별로 트래픽 분류가 이루어질 경우에는 다음과 같은 효과를 얻을 수 있다.

첫째, 서비스 기준으로 트래픽 분류를 수행할 경우, 다른 응용을 통해 동일한 서비스가 제공되어 분류가 모호해 지는 경우를 방지할 수 있다. 예를 들어, NateOn 메신저는 NateOn 자체 응용뿐 아니라 인터넷 브라우저를 통해 자체 응용에서 제공받는 동일한 서비스를 제공 받을 수 있다. 이런 경우 서비스 기준으로 트래픽을 분류하지 않고 응용 기준으로 트래픽을 분류한다면, 동일한 서비스를 사용하는 두 경우에 대해 Internet Explorer라는 분석 결과와 NateOn이라는 분석 결과 2가지로 분석될 수 있다. 또한 프로토콜 기준으로 트래픽을 분류한다면, NateOn 응용을 사용하여 접속하는 경우에는 NateOn 프로토콜로, 브라우저를 사용하여 접속하는 경우에는 HTTPS로 트래픽을 분류하게 될 것이다. 따라서 서비스 기준으로 트래픽을 분류할 경우, 응용 기준이나 프로토콜 기준으로 이루어지지 않았던 보다 정확하고 세밀한 분석을 수행할 수 있다는 장점이 있다.

둘째, 응용 기준으로 트래픽 분류를 수행할 경우, 서버가 존재하지 않는 P2P 통신 트래픽들을 응용으로 분류할 수 있다. P2P 통신의 경우 서비스를 제공하는 서버가 존재하지 않아 서비스 기준으로 트래픽을 분류할 수가 없게 된다. 따라서 P2P 통신의 경우에는 응용 기준으로 트래픽을 분류해야 한다.

셋째, 프로토콜 기준으로 트래픽 분류를 수행할 경우에는, 동일한 특정 프로토콜을 사용하여 동일 성격의 서비스나 응용의 추가적 개발 시에 추가 분석의 필요성을 줄일 수 있다. eDonkey 프로토콜은 당나귀, 프루나, eMule 등 다양한 P2P 서비스에 사용되는 프로토콜로써, 파일 공유 프로그램의 특성상 eDonkey 프로토콜을 사용하는 프로그램들은 업데이트 및 새로운 응용의 개발이 자주 일어난다. 이러한 경우, eDonkey 프로토콜에 대한 분석이 이루어지지 않는다면, 새로운 응용이 추가되거나 기존 응용이 업데이트 될 때마다 새로운 응용에 대한 추가적 분석이 계속적으로 요구될 것이다. 따라서 프로토콜 기준으로 트래픽을 분류할 경우, 특정 프로토콜을 사용하는 응용들의 등장에 대한 추가적 시그니쳐 생성 및 응용에 대한 분석 필요성을 줄일 수 있다는 장점이 있다.

넷째, 기능 기준으로 트래픽 분류를 수행할 경우에는, 동일 서비스나 응용 내의 특정 기능만을 분류할 수 있어, 특히 제어 측면에서 트래픽 분류 결과의 활용을 높일 수 있다. 예를 들어, NateOn 응용 안에서 파일 전송 기능을 수행할 경우 서비스나 응용 기준으로만 트래픽을 분류한다면, 다른 파일 전송 기능이 다른 기능들과 구분 없이 NateOn 응용이나 서비스로만 분류가 된다. 따라서, 트래픽 분류 결과를 제어 목적으로 활용할 경우에 NateOn 파일 전송 기능을 제어

하고자 할지라도, NateOn 서비스나 응용을 모두 제어하게 된다. 하지만, 기능 측면에서 트래픽 분류를 적용한다면 트래픽의 기능을 분석하므로 동일 서비스나 응용에서 동일한 프로토콜을 사용하는 기능들 중에서 원하는 기능만 분석/차단이 가능한 장점이 있다.

본 논문에서 제안하는 분류기준을 정리해 보면, 서비스는 응용을 제공하는 서버의 관점에서, 응용은 서비스를 제공받는 사용자의 관점에서, 그리고 프로토콜과 기능은 트래픽 발생 과정과 발생 목적 관점에서의 분류 결과를 나타낸다. 이처럼 4가지 분류 기준을 사용하여 분류를 수행할 경우, 위와 같은 장점들을 가지고 있지만, 4가지 분류 기준 모두를 적용하여 다각적인 분석을 할 경우에는 더욱 정확하고 세밀한 분석 결과를 제공할 수 있다는 장점이 있다.

다각적 분석은 하나의 트래픽에 대해 4가지의 정보를 모두 제공하는 분석으로써, 기존의 트래픽 분석에서는 할 수 없었던 여러 가지 정보를 제공할 수 있다. 예를 들어 사용자가 Internet Explorer 브라우저를 사용하여 Naver 포탈 사이트에서 블로그 글을 올리는 경우 발생하는 트래픽의 경우에는 기존 분석 결과로는 Naver로만 분석되거나, HTTP로만 분석되는 등, 사용자에게 충분한 정보를 제공하지 못했다. 기존 분석 결과가 전달하는 정보는 불충분할 뿐 아니라 정확하지 않은 분석이라고 할 수 있다. 하지만 본 논문에서 제안하는 4가지 분류 기준을 통해 다각적 분석을 수행하므로 표 2와 같은 결과를 제시하므로 정확하고 세밀한 분류 결과를 얻을 수 있다.

Table 2. Example of Four Dimension of Classification Criteria

분류 기준	분류 결과
서비스	Internet Explorer
응용	Naver
프로토콜	HTTP
기능	Posting

3.3 제안하는 분류 기준별 속성

본 장에서는 앞서 제안한 4가지 분류 기준별 계층화된 속성을 정의하여 제안한다. 앞서 언급한 것과 같이 계층화된 속성은 세분화되거나 통합된 결과를 사용자가 원하는 계층에서 사용할 수 있어 결과의 활용이 다양할 수 있다. 또한, 계층화되어 분류 결과가 관리되므로, 결과의 관리가 쉽다는 장점이 있다. 본 논문에서는 기준별로 표 3과 같은 계층화된 속성을 제안한다.

표 3에서 나타난 것과 같이 서비스와 응용 분류 기준은 3계층으로, 프로토콜과 기능 분류 기준은 각 2계층으로 구성되어 있다. 서비스 기준은 서비스의 이름으로 2계층을 구성하고, 서비스들의 그룹화를 통해 1계층을 구성하였다. 또한 2계층들의 서비스가 제공하는 세부 서비스들로 3계층을 구성하였다. 응용 기준의 경우에는 응용의 이름으로 3계층을 구성하고, 응용의 그룹화를 통해 응용의 목적과 종류로 각 1,2 계층을 구성하였다. 프로토콜 기준의 경우에는 프로토콜의 이름으로 2계층을 구성하고, 프로토콜들의 종류 별 그룹화를 통해 1계층을 구성하였다. 기능 기준의 경우에는 트래픽이 발생되는 목적으로 1계층을 구성하고 세부 행동으로 2계층을 구성하였다.

Table 3. Definition of Attributes for Each Criteria

분류 기준	Layer 1	Layer 2	Layer 3
서비스	Service Type	Service Name	Provided Service
	IM	NateOn	Chat
응용	Application Purpose	Application Type	Application Name
	Collaboration	UM	NateOn
프로토콜	Protocol Type	Protocol Name	-
	IM	NateOn	-
기능	User Purpose	User Action	-
	File Transfer	Download	-

표 3에서 나타난 기준별 속성에 따른 예시는 NateOn 응용을 통해 파일 전송을 하는 경우를 예로 든 것이다. 위의 경우를 서비스 기준으로 분류한다면, NateOn 서비스는 IM (Instant Messaging)으로 그룹화 되어 IM이라는 1계층 아래에 NateOn 서비스로 2계층에 속하게 된다. 또한, NateOn에서 파일 전송을 할 때 사용하는 Chat 서비스는 3계층에 속하게 된다. 응용 기준으로 분류할 때에는 NateOn 응용이 3계층에 속하게 되고, 그 목적과 종류에 따라 그룹화되어 1계층 Collaboration과 2계층 IM에 속하게 된다. 프로토콜 기준의 경우, NateOn 파일 전송 시 사용되는 NateOn 프로토콜이 2계층에 속하고, 그 프로토콜의 종류로 그룹화 되어 IM이라는 1계층 아래에 속하게 된다. 기능 기준의 경우에는 사용자가 사용하는 목적인 File transfer의 1계층으로 분류되고, 세부 행동으로 download로 2계층 분류가 된다.

위에서 제안한 분류 속성으로 각 분류 기준을 구성하지 않을 경우에는, 분류 결과의 관리 및 활용에 어려움이 있다.

서비스 기준을 2계층으로 구성하여 종류와 서비스만으로 구성할 경우에는, 하나의 서비스가 다양한 종류의 세부 서비스를 제공하는 경우에는 결과 사용의 부정확함 혹은 불편함을 가져올 수 있다. 하나의 서비스에서 제공하는 다양한 세부 서비스를 하위 계층으로 가지지 못하기 때문에 세밀한 정보의 전달을 하지 못하거나, 세밀한 정보 전달을 위해 2계층 요소를 세분화하는 결과를 가져오게 된다. 예를 들어, Naver에서 제공하는 메일 서비스와 뉴스 서비스의 경우, 3계층 속성을 사용하면 portal-Naver-mail과 portal-Naver-news로 정의할 수 있지만 2계층만을 사용하는 경우에는 portal-Naver의 간략한 정보만을 제공하거나 mail-Naver, news-Naver로 2계층 요소를 세분화하여 결과 정보를 전달

하게 된다. 이러한 결과 정보 전달은 결과의 이해 및 활용에 어려움을 겪게 된다.

또한, 기능 기준의 경우 사용자의 목적과 행동으로 계층화를 수행하지 않을 경우에는 제어 측면에서 활용할 때, 정확한 제어를 수행할 수 없다. 예를 들어, 트래픽 분류 결과를 통해 사용자가 메신저를 통한 파일 공유(file sharing)에서의 다운로드(download)는 허용하되, 업로드(upload)를 허용하지 않을 경우가 존재할 수 있다. 이럴 때, 기능 기준이 목적과 그에 따른 세부 행동으로 분류되지 않고, 단순히 목적으로만 트래픽의 기능이 구분된다면, 파일 공유의 업로드와 다운로드가 모두 제어되거나 혹은 제어할 수 없는 결과가 발생된다. 따라서 정확한 결과의 분석과 폭넓은 결과의 활용을 위해서 기능 기준의 2단계 계층화가 필요하다.

4. 적용 및 분석 결과

본 장은 위에서 제안한 분류 체계를 실제 트래픽 분석기와 분석 결과에 적용한 결과에 대해 언급한다.

4.1 분류 체계의 적용

본 장에서는 제안하는 분류 체계의 기준 및 속성에 실제적으로 분석에 사용되는 시그니쳐를 적용한 분류 체계 적용 결과에 대해 언급한다. 본 분석에서 사용한 시그니쳐는 페이로드 시그니쳐[8]와 헤더 시그니쳐[15] 그리고 통계 시그니쳐[12]를 사용하여 분석하였다. 사용된 시그니쳐는 현재 본 연구실에서 실제 분석기에 적용되고 있다. 각각의 시그니쳐가 분류 가능한 기준의 범위는 표 4와 같다.

Table 4. Range of Classification Criteria for Each Signature Type

시그니쳐 형태	분류 기준 종류
Payload Signature	서비스, 응용, 프로토콜, 기능
Header Signature	응용
DNS Header Signature	서비스
Statistic Signature	응용

헤더 시그니쳐와 통계 시그니쳐의 경우에는 프로세스별로 수집된 트래픽에서 시그니쳐를 추출하기 때문에 응용 기준으로 분류가 가능하고, DNS 헤더 시그니쳐의 경우 DNS 패킷을 사용하여 서버의 이름과 IP를 추출하기 때문에 서비스 측면에서 분석이 가능하다.

표 5는 현재 보유한 시그니쳐를 실제 제안하는 분류 체계의 속성에 적용하여 구성한 분류 체계이다. 분류 가능한 기준 별 요소들을 그룹화한 결과 다음 표 5와 같은 기준별 1계층 개수를 얻을 수 있었다. 각 분류 기준별 상위 5개에 대한 설명은 표 8, 11, 14, 16에서 설명한다.

Table 5. Number of Layer 1 Attribute

Layer	서비스	응용	프로토콜	기능
Layer 1	32	9	13	10

4.2 트래픽 분류 결과

본 장에서는 제안하는 트래픽 분류 체계를 실제 학내 망 트래픽에 적용하여 분석하고 실제 사용 가능함을 입증하였다. 실험에 사용된 트래픽은 학내 망에서 발생한 24시간 트래픽(2012년 3월 7일 00시부터 23시 59분까지)을 제안하는 분류 기준별로 분석하였다. 전체 트래픽의 양은 표 6과 같다.

Table 6. Traffic Trace

Flow	Packet	Byte
66,620K	3,417,367K	3,004,436MB

표 7은 트래픽의 전체 분석 결과이다. 각 항목은 전체 트래픽(100%)대비 분석된 트래픽을 나타내었다. Overall 분류 결과는 4가지 기준 중 하나라도 분석된 결과를 말한다.

Table 7. Traffic Classification Results

분류 기준	Flow	Packet	Byte
서비스	20.26%	30.89%	31.19%
응용	80.51%	96.61%	97.40%
프로토콜	43.04%	56.26%	54.68%
기능	8.41%	19.13%	20.32%
전체	86.61%	98.53%	99.09%

표 8은 서비스 기준으로 분류된 트래픽 중 1계층 분류에서 상위 5개의 트래픽을 정의한 표이다. 전체 분류된 서비스 트래픽의 양을 전체(100%)로 계산하였다. 표 8은 기준별 계층화를 통해 서비스 기준 분류 결과를 통합하여 본 것이고, 표 9와 10은 1계층 중 portal의 2계층과 portal 중 Naver의 3계층을 세분화하여 본 결과이다.

Table 8. Details of Layer 1 in Service Criteria

Layer 1	Flow	Packet	Byte
education	24.58%	7.57%	6.70%
portal	18.17%	7.57%	6.60%
commone_use	12.24%	8.21%	7.75%
community	6.95%	3.88%	3.33%
multimedia	5.04%	11.93%	12.90%
Etc	33.02%	60.83%	62.72%

Table 9. Details of Layer 2 of Portal in Service Criteria

	Flow	Packet	Byte
Daum	30.66%	46.42%	52.73%
Nate	29.75%	19.27%	16.09%
Naver	18.98%	16.52%	14.40%
Google	7.80%	3.63%	2.84%
Yagoo	1.15%	0.98%	0.96%
Etc.	11.66	13.19	12.99

Table 10. Details of Layer 3 of Naver in Service Criteria

	Flow	Packet	Byte
g	21.71%	34.36%	41.19%
search	31.55%	17.69%	12.01%
imgnews	7.33%	9.64%	11.08%
sstatic	8.40%	10.69%	9.93%
music	1.18%	5.95%	8.48%
map	7.59%	6.95%	7.18%
static	11.00%	6.95%	4.89%
blogimgs	5.80%	4.51%	2.91%
nmv	4.76%	2.61%	1.61%
kin	0.32%	0.49%	0.62%
apps	0.28%	0.11%	0.07%
tvguide	0.02%	0.02%	0.02%
blog	0.08%	0.03%	0.01%

표 11은 응용 기준에서 분류된 트래픽 중 1계층 분류에서 상위 5개의 트래픽을 정의한 표이다. 전체 분류된 서비스 트래픽의 양을 전체(100%)로 계산하여 결과를 나타내었다.

Table 11. Details of Layer 1 in Application Criteria

	Flow	Packet	Byte
utility _management	28.68%	5.30%	4.45%
collaboration	24.62%	50.46%	51.54%
file_sharing	22.24%	16.87%	16.86%
web	9.48%	6.44%	6.08%
entertainment	8.48%	17.46%	17.97%
etc	6.52%	3.47%	3.10%

다음의 표 12, 13은 1계층 file sharing의 2계층 결과와 p2p의 3계층 결과를 세분화하여 나타낸 것이다.

Table 12. Details of Layer 2 of file_sharing in Application Criteria

	Flow	Packet	Byte
p2p	93.32%	73.96%	69.67%
web_hard	6.68%	26.04%	30.33%

Table 13. Details of Layer 3 of p2p in Application Criteria

	Flow	Packet	Byte
bittorrent	91.99%	98.59%	99.21%
utorrent	7.99%	1.39%	0.77%
emul	0.0191%	0.01%	0.01%
soribada	0.0007%	0.00%	0.00%
mobile_fileguri	0.0004%	0.00%	0.01%
azureus	0.0000%	0.00%	0.00%

표 14는 프로토콜 기준에서 분류된 트래픽 중 1계층 분류에서 상위 5개의 트래픽을 정의한 표이다. 전체 분류된 서비스 트래픽의 양을 전체(100%)로 계산하였다.

Table 14. Details of Layer 1 in Protocol Criteria

	Flow	Packet	Byte
web	76.86%	55.09%	55.88%
p2p	13.94%	14.25%	13.85%
utility	4.14%	3.94%	1.49%
web_hard	1.87%	10.65%	12.84%
im	1.32%	1.59%	1.38%
etc	1.88%	0.00%	0.00%

표 15는 1계층 web의 2계층 결과를 세분화하여 나타낸 것이다.

Table 15. Details of Layer 2 of web in Protocol Criteria

	Flow	Packet	Byte
http	80.43%	98.98%	99.75%
dns	19.48%	0.99%	0.23%
https	0.09%	0.03%	0.02%

표 16은 기능 기준에서 분류된 트래픽 중 1계층 분류에서 상위 5개의 트래픽을 정의한 표이다. 전체 분류된 서비스 트래픽의 양을 전체(100%)로 계산하였다.

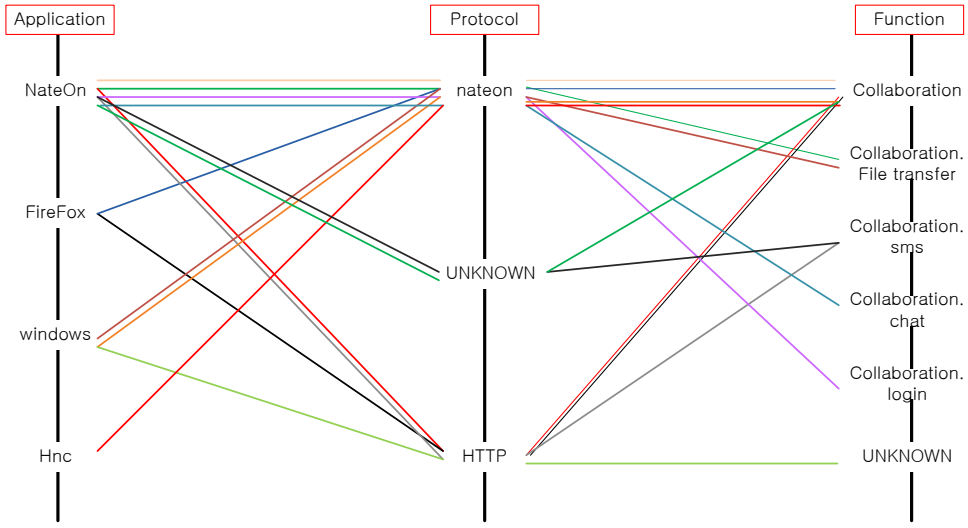


Fig. 3. Result of Multilateral and Hierarchical Traffic Identification in Nateon

Table 16. Details of Layer 1 in Function Criteria

	Flow	Packet	Byte
mail	24.36%	7.80%	5.81%
web	11.33%	5.56%	5.23%
internet_disk	6.52%	30.23%	33.56%
collaboration	2.70%	4.55%	4.19%
management	2.15%	1.45%	1.40%
etc	52.93%	50.41%	49.81%

표 17은 1계층 collaboration의 2계층 결과를 세분화하여 나타낸 것이다.

Table 17. Details of Layer 2 of collaboration in Function Criteria

	Flow	Packet	Byte
voip	40.69%	9.19%	7.38%
sns	26.78%	2.04%	1.67%
chat	21.13%	2.46%	1.36%
file_transfer	5.84%	80.93%	83.70%
im	3.13%	0.81%	0.84%
voice	1.77%	4.46%	4.98%
sms	0.66%	0.06%	0.03%
update	0.00%	0.00%	0.00%

그림 3은 NateOn 서비스 트래픽을 다각적인 측면에서 분석한 결과를 나타낸 것이다. 아래의 그림을 통해서 NateOn 서비스가 어떠한 응용을 통해 제공되고, 어떠한 프로토콜을 사용하여 통신하고, 트래픽의 기능은 무엇인지를 다각적으로 분석할 수 있다.

NateOn 서비스는 NateOn을 통해서 사용된 양이 가장 많았고, 대부분은 NateOn 자체 프로토콜을 사용하였고, collaboration 기능을 제공하였다. Collaboration은 기능의 1계층으로 분석된 트래픽은 대부분 collaboration의 chat 기능 트래픽으로 예상이 되지만, 실제 분석 결과 chat 기능 보다는 file transfer 기능을 사용할 때 더 많은 트래픽을 발생하는 것을 확인하였다. 특히, 파일 전송 시 발생하는 트래픽의 비율은 전체 트래픽 대비 76%를 기록 하였다. 이와 같이 기존의 단일 응용 기준 트래픽 분석에서는 단순히 Nateon으로만 분석되던 트래픽들이 다양한 관점에서 분석됨으로써, 해당 트래픽의 충분한 정보를 제공할 수 있게 된다. 이러한 분석 결과는 향후 다양한 분야에서 활용될 수 있다.

현재 보유 시그니쳐로는 collaboration 까지만 분석 가능하였으므로 추출된다. NateOn 서비스로 분류된 트래픽 중에서 가장 많은 Byte양을 차지한 트래픽은 NateOn응용과 NateOn 프로토콜을 통해 File transfer 기능을 사용한 트래픽으로 전체 NateOn 서비스 트래픽 양 중에서 76%를 차지하였다.

실험 결과, 실험을 통해 실제 트래픽 분류 결과에 제안하는 분류 체계를 적용할 수 있음을 입증하였다 또한, 제안하는 기준별 분류가 가능함과, 제안하는 계층 속성에 따라 결과 값의 통합화(Roll-up)와 세분화(Drill-down)가 가능함을 보이며 트래픽 분류 결과의 활용이 사용자의 필요에 따라 자유롭게 사용될 수 있음을 입증하였다. 또한, 하나의 트래픽에 대한 4가지 분류 결과의 다각적 분석 결과를 가능하게 함으로써 트래픽에 대한 이해를 높이고 분석 결과의 활용이 가능함을 증명하였다.

5. 결론 및 향후 연구

트래픽 분석의 필요성이 높아지고 분류 결과가 다양한 분야에서 적용됨에 따라 분류 결과를 효과적으로 활용할 수 있는 분류 체계의 필요성이 요구되었다.

본 논문에서는 다각적이고 계층적으로 트래픽 분류가 가능한 분류 체계의 구조와 기준, 기준별 계층화 속성에 대해 제안하였다. 또한, 실제 트래픽에 적용을 통하여 각 기준별 트래픽 분류가 가능함과 계층적으로 분류한 결과의 활용이 사용자의 다양한 필요와 요구에 따른 결과를 제공할 수 있음을 보였다.

본 논문에서 실험한 결과에서는 전체적으로는 byte 기준으로 약 99%의 분석률을 보였지만, 기능 기준의 분류 결과가 20%에 그치는 것으로 분석되었다. 이는 기능 기준으로 트래픽을 분류 할 수 있는 정교한 시그니쳐의 부재 때문으로 판단된다. 따라서, 분류 체계에 적용하여 분류할 수 있는 정교하고 정확한 시그니쳐에 대한 정의 및 추출 작업이 이루어져야 할 것이다.

다양한 분류 기준으로 분석된 분류 결과를 한눈에 파악할 수 있는 분류 결과 시각화에 대한 연구를 진행할 계획이다.

참 고 문 헌

[1] F. C. VNI, “Cisco Visual Networking Index: Global Mobile data Traffic Forecast Update 2009-2014,” Cisco Public Information, February, Vol.9, 2010.

[2] J.-h. Kim, S.-H. Yoon, and M.-S. Kim, “Research on traffic taxonomy for Internet traffic classification,” in Network Operations and Management Symposium (APNOMS), 2011 13th Asia-Pacific, 2011, pp.1-4.

[3] Check Point AppWiki. [Internet], http://appwiki.checkpoint.com/appwikisdb/public.htm

[4] FortiGuard Center App Control. [Internet], http://www.fortiguard.com/encyclopedia/applications/

[5] Paloalto Networks Applipedia. [Internet], https://applipedia.paloaltonetworks.com/

[6] IANA port number list. [Internet], http://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xml

[7] Z. Jian and A. Moore, “Traffic Trace Artifacts due to Monitoring Via Port Mirroring,” in End-to-End Monitoring Techniques and Services, 2007. E2EMON '07. Workshop on, 2007, pp.1-8.

[8] F. Risso, M. Baldi, O. Morandi, A. Baldini, and P. Monclus, “Lightweight, Payload-Based Traffic Classification: An Experimental Evaluation,” in Communications, 2008. ICC '08. IEEE International Conference on, 2008, pp.5869-5875.

[9] J.-S. Park, S.-H. Yoon, and M.-S. Kim, “Software Architecture for a Lightweight Payload Signature-Based Traffic Classification System,” in Traffic Monitoring and Analysis. Vol. 6613, J. Domingo-Pascual, Y. Shavitt, and S. Uhlig, Eds., ed: Springer Berlin Heidelberg, 2011, pp.136-149.

[10] A. W. Moore and D. Zuev, “Internet traffic classification using bayesian analysis techniques,” SIGMETRICS Perform. Eval. Rev., Vol.33, pp.50-60, 2005.

[11] K. Xu, Z.-L. Zhang, and S. Bhattacharyya, “Profiling internet backbone traffic: behavior models and applications,” in ACM SIGCOMM Computer Communication Review, 2005, pp.169-180.

[12] J.-W. Park, S.-H. Yoon, J.-S. Park, S.-W. Lee, and M.-S. Kim, “Statistic Signature based Application Traffic Classification,” KICS, Vol.34, pp.1234-1244, 2009.

[13] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, “BLINC: multilevel traffic classification in the dark,” in ACM SIGCOMM Computer Communication Review, 2005, pp.229-240.

[14] A. Callado, C. Kamienski, G. Szabo, B. Gero, J. Kelner, S. Fernandes, et al., “A Survey on Internet Traffic Identification,” Communications Surveys & Tutorials, IEEE, Vol.11, pp.37-52, 2009.

[15] S.-H. Yoon, J.-W. Park, J.-S. Park, Y.-S. Oh, and M.-S. Kim, “Internet Application Traffic Classification Using Fixed IP-Port,” in Management Enabling the Future Internet for Changing Business and New Computing Services, 2009, pp.21-30.

윤 성 호

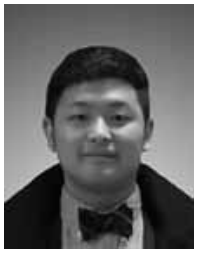
e-mail : sungho_yoon@korea.ac.kr

2009년 고려대학교 컴퓨터정보학과(학사)

2011년 고려대학교 컴퓨터정보학과(석사)

2010년~현 재 고려대학교 컴퓨터정보학과 박사과정

관심분야 : 네트워크 관리 및 보안, 트래픽 모니터링 및 분석



안 현 민

e-mail : queen26@korea.ac.kr

2012년 고려대학교 컴퓨터정보학과(학사)

2012년~현 재 고려대학교 컴퓨터정보학과 석사과정

관심분야 : 네트워크 관리 및 보안, 트래픽 모니터링 및 분석

김 명 섭

e-mail : tmskim@korea.ac.kr

1998년 포항공과대학교 전자계산학과(학사)

2000년 포항공과대학교 컴퓨터공학과(석사)

2004년 포항공과대학교 컴퓨터공학과(박사)

2004년~2006년 Post-Doc., Dept. of ECE, Univ. of Toronto, Canada

2006년~현 재 고려대학교 컴퓨터정보학과 교수

관심분야 : 네트워크 관리 및 보안, 트래픽 모니터링 및 분석, 멀티미디어 네트워크