DOI: 10.3745/KIPSTC.2010.17C.2.205

고정 IP-port 기반 응용 레벨 인터넷 트래픽 분석에 관한 연구

윤 성 호[†] · 박 준 상[†] · 박 진 완[†] · 이 상 우[††] · 김 명 섭[†††]

요 약

인터넷의 대중화로 인해 네트워크 트래픽은 나날이 증가되고 있다. 따라서 네트워크 자원의 효과적인 사용을 위한 응용 트래픽 분석의 중요성은 나날이 강조되고 있다. 본 논문에서는 고정 IP-port기반의 응용 트래픽 분석 방법론을 제안한다. 고정 IP-port는 오직 한 응용에서 고정적으로 사용하는 {IP address, port number, transport protocol}의 쌍으로써 각각의 응용을 분석해서 자동적으로 수집할 수 있다. 본 논문에서는 고정 IP-port를 사용하여 기존 연구에서 제안된 방법들 보다 매우 가볍고, 빠르며 정확한 실시간 트래픽 분석 시스템을 설계하였다. 또한, 기존의 연구에서 제안된 통일성 없는 검증 방법을 보완하여 객관적 검증 시스템을 설계하고 분석결과를 정확하게 검증하였다. 본 논문은 고정 IP-port를 추출하는 매우 효과적인 방법과 시스템 구조, 그리고 분석 결과의 객관적 검증 시스템을 제안한다. 그리고 실험과 검증 시스템을 통하여 고정 IP-port 기반 응용 레벨 인터넷 트래픽 분석 방법론의 타당성을 증명한다.

키워드 : 트래픽 모니터링, 트래픽 분석, 응용 프로그램 분석, 고정 IP-port

Fixed IP-port based Application-Level Internet Traffic Classification

Sung-Ho Yoon[†] · Jun-Sang Park[†] · Jin-Wan Park[†] · Sang-woo Lee[††] · Myung-Sup Kim[†††]

ABSTRACT

As network traffic is dramatically increasing due to the popularization of Internet, the need for application traffic classification becomes important for the effective use of network resources. In this paper, we present an application traffic classification method based on fixed IP-port information. A fixed IP-port is a {IP address, port number, transport protocol}triple dedicated to only one application, which is automatically collected from the behavior analysis of individual applications. We can classify the Internet traffic more accurately and quickly by simple packet header matching to the collected fixed IP-port information. Therefore, we can construct a lightweight, fast, and accurate real-time traffic classification system than other classification method. In this paper we propose a novel algorithm to extract the fixed IP-port information and the system architecture. Also we prove the feasibility and applicability of our proposed method by an acceptable experimental result.

Keywords : Traffic Monitoring and Analysis, Traffic Classification, Application Identification, Fixed IP-port

1. 서 론

최근 인터넷 사용자의 증가와 고속 네트워크의 보급으로 인해 네트워크 트래픽이 급증하였다. 이것은 단순히 WWW, FTP, e-mail 과 같은 정통적인 인터넷 서비스뿐만 아니라 멀티미디어 스트리밍, P2P(peer-to-peer) 파일 공유, 게임과 같은 다양한 멀티미디어 서비스의 대중화에 원인이 있다. 따라서 인터넷 트래픽이 급증함에 따라 효과적인 네트워크 관리를 위한 트래픽 모니터링 및 분석의 중요성이 커지고 있다[1, 2].

네트워크 트래픽 분석은 네트워크 링크로부터 패킷을 수집하고 분석 기준에 맞게 각 패킷을 분류하는 일련의 과정이다. 이러한 분석 결과는 네트워크 관리와 제어의 관점에서 효과적으로 사용된다. 관리자는 네트워크 트래픽을 각각의 기준 별로 차단 및 제어를 할 수 있다.

트래픽 분석은 여러 기준에 의해 수행된다. 대표적인 예로는 응용레벨 프로토콜, 전송계층 프로토콜, 트래픽 타입, 응용 등이 있다. 많은 머신 러닝 및 트래픽 행동 기반의 기

※ 이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단(KRF-2007-331-D00387)과 2009년 정부(교육과학기술부)의 재원으로 한국연구재단(2009-0090455)의 지원을 받아 수행된 연구임.
† 준 회 원 : 고려대학교 컴퓨터정보학과 석사과정
†† 준 회 원 : 고려대학교 컴퓨터정보학과 학사과정
††† 종신회원 : 고려대학교 컴퓨터정보학과 조교수
논문접수 : 2009년 8월 27일
수 정 일 : 1차 2009년 12월 14일
심사완료 : 2010년 1월 6일

존 논문[3, 8]에서는 트래픽의 타입에 따라 트래픽을 분석한다. 예를 들어, 통계적 특징과 행동유형이 유사한 msn, google talk 등과 같은 메신저들을 “CHAT” 이란 타입으로 통합 정의하여 분석한다[3]. 하지만, 이러한 트래픽 타입의 기준을 실제 네트워크에 적용하였을 때 네트워크 관리 측면에서 많은 이점을 얻지 못한다. 이에 반해 실제 네트워크 관리에서 트래픽을 응용 별로 제어, 차단 하는 것은 많은 이점을 가진다. 예를 들어 메신저 전체를 “CHAT”이란 타입으로 제어하기 보다는 msn이란 응용 단독으로 제어하는 것이 네트워크 관리 측면에서 더욱 효과적이다. 본 논문에서는 응용을 기준으로 트래픽을 분석한다.

이미 많은 기존 논문[1, 3-10]에서 트래픽을 분석하는 다양한 알고리즘을 제안하였다. 하지만, 높은 오버헤드와 비 정확성, 긴 시간의 사전 작업을 요구하기 때문에 실시간으로 발생되는 트래픽을 분석하는데 많은 어려움이 있었다. 또한 제안된 알고리즘을 객관적으로 평가 및 검증 할 수 있는 기준도 명확히 제시되지 않았다.

본 논문에서는 고정 IP-port 정보를 이용한 응용 트래픽 분석 방법론을 제안한다. 고정 IP-port는 오직 한 응용에서만 사용되는 {IP address, port number, transport protocol} 쌍을 의미한다. 이러한 고정 IP-port는 각 응용을 분석 함으로써 자동적으로 추출 할 수 있다.

고정 IP-port는 다음과 같은 2가지 특징을 가진다.

(1) 매우 정확하다. 고정 IP-port 추출 알고리즘은 실제 트래픽을 발생 시키는 종단 호스트에서 데이터를 직접 수집하기 때문에 매우 정확한 특징을 가지고 있다.
(2) 매우 빠르다. 고정 IP-port기반의 분석은 수집한 패킷 헤더와 단순히 비교 함으로써 인터넷 트래픽을 매우 효과적이고 빠르게 분석한다. 따라서 실시간 트래픽 분석을 가능하게 한다.

이러한 정확하고 빠른 고정 IP-port의 특징을 이용하여 기존 연구의 문제점인 큰 오버헤드와 비 정확성을 해결한다. 또한, 상태 전이 다이어그램을 이용한 추출 알고리즘을 사용하여 짧은 시간 안에 고정 IP-port를 추출하는 방법도 제시한다.

본 논문은 다음과 같이 구성되었다. 2 장에서는 관련 연구를 살펴보며3장에서는 고정 IP-port를 정의 하고, 각각의 응용에서 고정 IP-port를 효과적으로 추출하는 알고리즘을 4장에서 기술한다. 5장에서는 고정 IP-port 기반의 트래픽 분석 시스템을 소개한다. 6장에서는 실험 및 결과를 7장에서는 결론 및 향후 연구를 기술한다.

2. 관련 연구

트래픽 분석 방법들은 크게 4가지로 구분할 수 있는데, 이들은 포트 기반 분석 [3], 시그니처 기반 분석 [4], 머신 러닝기반의 분석 [5-7], 트래픽 상관관계 기반 분석 [1, 8]이다. 또한 검증 방법론[3, 8-10] 역시 많은 연구에서 다양하게 제안되었다.

2.1 포트 기반 트래픽 분석

잘 알려진 포트 기반 방법은 IANA에서 지정한 포트 정보를 이용한다. 따라서 동적 혹은 알려지지 않은 포트를 사용하는 응용에 대해서는 분석하지 못하는 단점이 있다. 최근 사용되는 응용들은 방화벽 및 IPS 장비를 통과하기 위하여 알려진 포트사용을 피하고 있다. 더 이상 포트 번호가 특정 프로토콜 또는 특정 응용을 의미하지 않는다.

2.2 시그니처 기반 트래픽 분석

시그니처 기반의 방법은 각 응용 트래픽 별로 그들만이 사용하는 다른 응용들과 구분되는 응용 내의 공통분모를 찾아내어 그것을 그 응용프로그램의 시그니처라고 명명하고, 인터넷 트래픽을 수집할 때 같은 시그니처를 가졌는지 아닌지를 확인하는 방법으로 응용을 분석하는 방법이다. 기존에 제안된 대부분의 시그니처 기반 분석은 트래픽의 페이로드를 기반으로 시그니처를 추출 하였지만, 포트번호나, 통계적 수치 등과 같이 특정 응용을 구분할 수 있는 것들은 모두 큰 범주 안에서 시그니처라 할 수 있다.

이 방법의 장점으로는 시그니처가 확인된 응용에 대해서는 정확한 분석이 가능하다는 것이다. 그러나 모든 응용 별로 시그니처를 수작업으로 찾아야 하고, 시그니처를 확인하기 힘든 응용들이 존재하며, 또 찾아진 시그니처가 응용프로그램의 변화(수정, 삭제)에 적절히 대처하지 못한다는 단점이 있다.

본 연구에서 제안하는 고정 IP-port 기반의 분석 방법도 시그니처 기반 분석 방법의 하나이다. 한 응응을 대표하는 {IP address, port number, transport protocol}를 찾아내어 시그니처로 사용하는 것이다. 앞서 설명한 기존 시그니처 분석 방법의 단점을 보완하기 위해 자동 추출 시스템을 사용한다. 따라서 새로운 응용에 대한 대처가 빠르다. 또한, 단순한 패킷 헤더 비교로 분석 오버헤드를 낮추었다.

2.3 머신 러닝 기반 트래픽 분석

머신 러닝방법은 응용 별 인터넷 트래픽의 특징이 될 수 있는 항목 (port, flow duration, inter-arrival time, packet size, byte size)들을 머신 러닝 알고리즘으로 학습시켜 분석하는 방법이다.

머신 러닝은 크게 분석 기법과 군집 기법으로 나뉜다. 분석 기법에서는 Bayesian Network, Decision Tree[7]가 있고, 군집 기법에서는 EM(Expectation Maximization)[6]가 있다.

이 방법의 장점은 머신 러닝의 고급 알고리즘을 이용함으로써 트래픽을 타입 별로 분류함에 있어 다른 분석 방법에 비해 높은 분석률을 제공한다는 것이다. 그러나 제한된 범위의 응용 트래픽에 한하여 트래픽 데이터를 수집하고 학습하였기 때문에 실제 네트워크 트래픽에 적용하였을 경우 분석의 정확성이 떨어지는 단점이 있다. 또한, 매번 새로운 응

용이 출현 하였을 때 머신 러닝 알고리즘을 다시 실행해야 하기 때문에 사전 준비 작업의 오버헤드가 크다.

2.4 상관관계 기반 트래픽 분석

트래픽 상관관계 기반 분석 방법은 인터넷 트래픽의 3 레벨 주소체계 {IP address, port number, transport protocol}와 트래픽 발생 시점, 발생 형태 등의 특성을 바탕으로 트래픽 플로우들 사이에 연관성을 가중치로 표현하고 가중치의 임계값을 적용하여 트래픽을 응용 별로 구분하는 방법이다.

이 방법의 장점으로는 트래픽의 분류에 있어 응용들이 가지는 특징을 분석에 활용하여 분석률을 높일 수 있다는 것이다. 그러나 응용 별 특징의 활용에 대한 명확한 알고리즘이 없이 trial-and-error의 방법으로 최적의 분석률을 보이는 임계값을 찾고 있기 때문에 실제 인터넷 트래픽에 적용하였을 경우 분석 결과에 대한 신뢰성을 보장하기 어렵다.

2.5 트래픽 분석 결과 검증

활발히 진행되고 있는 트래픽 분석 연구와 달리 분석 결과를 검증하는 연구는 아직 미흡한 실정이다. 각각의 연구에서 제안하는 무분별하고 불분명한 검증 기준과 방법은 여러 알고리즘들 간의 정확한 비교와 평가를 어렵게 한다. 따라서 정확하고 객관적인 검증 시스템 개발이 요구 되고 있다.

BLINC[8]에서는 기존의 제안된 트래픽 분석 알고리즘 중 성능(분석률과 정확도)이 우수한 시그니처 기반의 분석 방법의 결과를 검증 시스템의 정답지(Ground truth)데이터로 사용하였다. 이러한 검증 방법은 시그니처를 이용한 분석이 정확하다는 가정으로 이루어지기 때문에 검증의 결과를 100% 신뢰하기 어렵다.

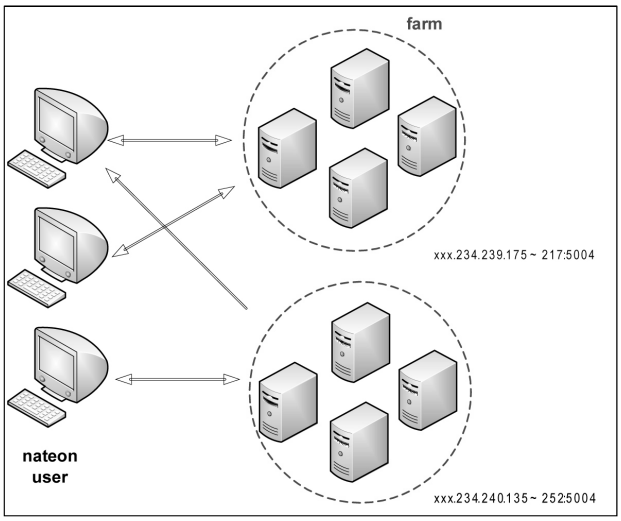
또한 최근에 발표된 논문 Szabó et al [9]에서는 종단 호스트에서 실제 발생되는 트래픽에 발생 응용의 ID를 심어 이를 기준으로 검증하는 방법을 제안한다. 하지만 이를 위해 많은 종단 호스트에 특정 드라이버를 설치해야 하고 발생 트래픽에 ID을 기록해야 하는 오버헤드를 가진다.

여러 알고리즘에 의해 분석 된 결과를 정확하고 객관적으로 평가 및 검증하기 위해서는 통일된 평가 요소와 검증 방법이 필요하다. 본 논문에서 제안하는 검증 시스템은 5.3장에서 좀더 자세하게 설명한다.

3. 고정 IP-port

고정 IP-port란 하나의 응용에서 고정적으로 사용하는 {IP address, port number, transport protocol} 쌍을 의미한다. 실제 이러한 정보는 인터넷 서비스를 제공하는 서버의 정보를 의미한다. 실제 많은 인터넷 응용들은 실제 서비스를 제공하기 전에 사용자 인증 과정을 거치게 되는데 이러한 과정을 로그인 서버가 담당한다. 따라서 로그인 서버의 {IP address, port number, transport protocol} 가 해당 응용의 고정 IP-port의 하나가 된다.

고정 IP-port을 확인하기 위해 간단한 실험을 하였다. 실험을 위하여 본교 네트워크에서 많은 트래픽을 발생하는 nateon 메신저 응용을 선정하여 7일간 응용을 사용할 때 발생되는 서버 IP-port를 조사 하였다.



(그림 1) 고정 IP-port의 예

(그림 1)은 네트워크 내 호스트들이 nateon 메신저 응용을 사용 했을 때, 단 한차례도 다른 응용과 공유하지 않은 서버의 IP-port가 존재함을 보여준다. (그림 1)에서 표시된 IP와 port는 오직 해당 응용을 사용하였을 때에만 나타난다. 또한, BLINC[8]에서 “farm”이라는 용어로 정의한 형태로 고정 IP-port가 존재 함을 확인할 수 있었다.

4. 고정 IP-port 추출 알고리즘

본 장에서는 고정 IP-port를 추출하는 알고리즘을 제안한다. 본 알고리즘은 분석 대상 응용의 트래픽 정보를 입력받아 해당 응용의 고정 IP-port를 출력한다.

본 알고리즘의 핵심은 해당 응용이 사용하는 모든 {IP address, port number, transport protocol} 쌍에서 고정 IP-port를 찾아내는 것이다. 정확한 고정 IP-port를 추출하는 알고리즘을 설계하기 위해, IP-port 를 정적, 임시, 동적과 같이 세가지 타입으로 정의하였다.

(가) 정적 IP-port는 한 응용이 단독으로 고정적으로 상용하는 IP-port를 의미한다. 이것의 예로는 로그인 서버의 IP-port가 있다. 즉, 본 논문에서 제안하는 고정 IP-port와 동일한 의미이다.

(나) 동적 IP-port는 한 응용에서 사용되긴 하지만, 매번 IP-port가 바뀌는 경우이다. 대표적인 예로는 HTTP를 사용하는 클라이언트 호스트의 IP-port이다.

(다) 임시 IP-port는 동적 IP-port과 그 의미는 같지만, 다른 여러 응용과 공유한다는 점에서 차이가 있다. 예로써는 P2P호스트의 IP-port이다. 이러한 IP-port들은 매우 짧은 시간 동안 소량의 트래픽을 사용하고 다른 응용과 공유한다.

본 알고리즘에서는 정적 IP-port(고정 IP-port)를 추출하는 것을 목표로 한다. 동적 IP-port의 경우에는 사용 빈도가 매우 낮을 뿐만 아니라 너무 많은 데이터를 유지해야 하기 때문에 오버헤드를 증가 시킨다. 임시 IP-port는 서로 다른 응용에서 공유하여 사용할 수 있기 때문에 고정 IP-port가 아니다.

4.1 고려사항

본 논문에서 제안하는 고정 IP-port 추출 알고리즘은 다음의 4가지 고려사항들을 충족시켜야 한다.

• 범위 (Coverage)

가능한 많은 응용에서 고정 IP-port를 찾아내어야 한다. 최소한 하나의 응용에서 하나 이상의 고정 IP-port를 찾아내어야 한다. 또한, 각각의 응용은 다양한 특징을 가지기 때문에 모든 응용에 적용가능 한 보편적인 고정 IP-port 추출 시스템을 만들어야 한다.

• 정확도 (Accuracy)

추출된 고정 IP-port를 이용하여 분석된 결과의 FP(false positive)와 FN(false negative)은 최소가 되어야 한다[11]. 고정 IP-port 기반의 분석 결과는 높은 분석률(Completeness)를 가지지 못하기 때문에 다른 분석 방법(상관관계)과 결합하여야 한다. 따라서 Multi-Level 분석 방법의 첫 단계로써 매우 높은 정확도를 가져야 한다.

• 오버헤드 (Overhead)

고정 IP-port 추출 알고리즘의 메모리사용과 프로세싱 오버헤드는 적어야 한다. 방대한 동적, 임시 IP-port들이 포함된 IP-port들 중에서 고정 IP-port를 찾아내기 위해서는 여러 임계값을 설정하여 낮은 오버헤드를 유지하도록 하여야 한다.

• 상호배제 (Mutual Exclusion)

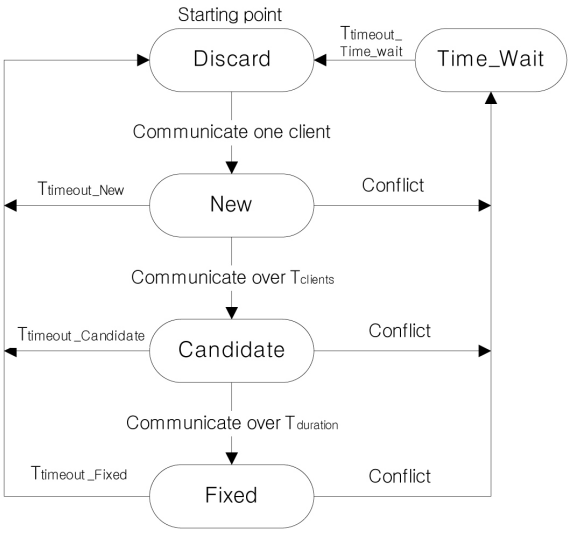
모든 고정 IP-port는 반드시 하나의 응용에 의해 사용되어야 한다. 둘 이상의 응용에서 공유하여 사용하는 IP-port는 고정 IP-port에서 제거되어야 한다.

4.2 고정 IP-port 추출 알고리즘

(그림 2)는 본 장에서 제안하는 알고리즘의 핵심인 고정 IP-port 상태 전이 다이어그램이다.

모든 IP-port들은 5가지 상태(Discard, New, Candidate, Fixed, Time_wait) 중 하나에 속한다. Fixed 상태의 IP-port들이 본 논문에서 제안하는 고정 IP-port이다. 앞에서 언급했듯이 본 추출 알고리즘은 각각의 대상 응용에 대해 독립적으로 실행한다.

최초 모든 IP-port는 Discard 상태에 속한다. 나머지 4 상태에 속하지 않은 모든 IP-port들이 이 상태에 속한다. 만약 발생시킨 응용을 알고 있는 IP-port가 있다면 해당



(그림 2) 고정 IP-port 상태 전이 다이어그램

IP-port는 New 상태로 이동한다. 또한 각 상태에서 적용하는 임계값(Tclienthost, Tduration)을 만족하는 IP-port들에 한해 Candidate, Fixed 상태로 차례로 이동하게 된다. Fixed 상태에 속한 IP-port들이 고정 IP-port로 결정 되며, 이것들은 트래픽 분석에 사용된다.

New, Candidate, Fixed 상태에서 임계값을 만족하지 못하여 무한히 한 상태에 남아있는 것을 방지하기 위하여 각 상태에 임계값($\mathbf{T}_{timelimit}$)을 사용 하였다. 또한 매번 IP-port들을 발생시킨 응용을 관찰하여 서로 다른 응용에 의해 공유 여부를 확인하여, 만약 충돌(Conflict)이 발생하면 Time_Wait 상태로 이동한다. Time_Wait 상태는 다른 응용과 충돌로 인해 Discard 상태로 이동한 IP-port가 무한히 반복되어 불필요한 오버헤드를 발생시키는 것을 방지하기 위해 존재한다. 충돌(Conflict)이란, 하나의 IP-port가 서로 다른 응용에 의해 공유되는 것을 의미한다.

본 논문에서는 원활한 상태 이동을 위하여 몇 가지 임계값을 고려하였다. 다음은 본 논문에서 제안하는 3가지 임계값에 대한 설명이다.

• $T_{clienthost}$

특정 고정 IP-port와 통신한 고유한 클라이언트 호스트의 개수이다. 많은 클라이언트 호스트와 통신하는 IP-port 일수록 고정 IP-port일 확률이 높다. 따라서 일정 임계값(Tclienthost) 이상의 클라이언트 호스트 개수를 가지는 IP-port만이 고정 IP-port로 결정된다. 임계값이 증가하면 입력 데이터의 양이 풍부한 고정 IP-port만 추출되어, 이를 이용한 트래픽 분석의 정확도(Accuracy)는 증가한다. 하지만, 임계값에 도달하지 못하는 고정 IP-port들이 추출되지 못해 분석률(Completeness)은 감소한다.

• $T_{duration}$

사용기간(마지막 사용 시각 - 최초 등록 시각)을 의미한

다. 만약 호스트 개수 임계값(Tclienthost)만을 이용하여 고정 IP-port를 결정한다면 P2P 응용을 사용하는 임시 IP-port를 고정 IP-port로 결정 하는 실수를 저지를 수 있다. 따라서 본 알고리즘은 이러한 임시 IP-port를 구별하기 위해 사용기간을 이용한다. P2P 응용에서 발생되는 트래픽은 짧은 사용시간의 특징을 가지므로, 사용시간 임계값(Tduration) 값을 이용하여 정확한 고정 IP-port를 추출한다.

• $T_{timelimit}$

추출 알고리즘에 등록된 모든 IP-port들은 임계값 Tclienthost와 Tduration 값을 넘지 못하면 계속 메모리에 존재하게 된다. 물리적인 메모리의 제한으로 인해 모든 IP-port들은 영원히 저장하지 못한다. 또한 무한히 많은 데이터들은 고정 IP-port 추출 시 높은 오버헤드를 발생하여 실시간 분석을 불가능하게 한다. 따라서, 임계값(Ttimelimit)을 이용하여 일정 시간, 상태의 변화가 없는 IP-port를 삭제한다.

이러한 임계값들은 우리의 알고리즘을 다양한 환경에서 유연하고 효과적으로 만든다. 고정 IP-port추출 알고리즘에서 가장 핵심적인 부분은 위에서 언급한 각 상태들의 이동이다. 제안한 알고리즘은 (그림 3)과 같은 코드에 의해 수행된다.

위 코드는 두 부분으로 나뉜다. 첫 번째 부분(line:3-7)은 새로운 IP-port 를 IPT(IP-port Table)에 등록시키거나 이미 등록되어 있다면 새로운 정보(충돌, 클라이언트의 수, 마지막 사용 시각)를 업데이트한다. 만약 새로운 IP-port의 정보가 기존의 응용과 다른 응용에 의해 발생되었다면, 충돌을 체크한다. 반면, 같은 응용에 의해 발생되었다면 마지막 시각을 업데이트한다.

두 번째 부분(line:8-23)은 IPT에 등록되어 있는 IP-port 레코드들을 앞서 설명한 고정IP-port 상태 전이 다이어그램에 따라 이동시킨다. 만약 현재 속한 상태의 임계값($T_{timelimit}$)보다 오래 기간 상태 변화가 없다면 Discard 상태로 이동한다. 또한 서로 다른 응용에 의해 충돌이 있으면 Time_Wait 상태로 이동한다. 앞에서 언급했듯이 Fixed 상태에 있는 IP-port들은 고정 IP-port 이다.

```
1:  procedure Fixed IP-port Extraction for an Application A
2:  IPT ← IP-port Table
3:  for each input flow record do
4:     search the corresponding IP-port record in IPT
5:     if found in IPT then update the status of the record;
6:     else add a new record with the state as New;
7:  end for
8:  for each flow record in IPT do
9:     Check the state of the record
10:    if state == New then
11:       if conflict then move to Time_Wait;
12:       else if over T_clienthost then move to Candidate;
13:       else if over T_timelimit then move to Discard;
14:    else if State == Candidate then
15:       if conflict then move to Time_Wait;
16:       else if over T_duration then move to Fixed;
17:       else if over T_timelimit then move to Discard;
18:    else if State == Fixed then
19:       if conflict then move to Time_Wait;
20:       else if over T_timelimit then move to Discard;
21:    else if State == Time_Wait then
22:       if over T_timelimit then  move to Discard;
23:  end for
24: end procedure
```

(그림 3) 고정 IP-port 추출 알고리즘 슈도코드

5. 고정 IP-port 기반 트래픽 분석 시스템

(그림 4)는 본교의 학내 네트워크에 설치한 고정 IP-port 기반 트래픽 분석 시스템 구조를 보여준다. 본 시스템은 3부분의 서브 시스템으로 나뉜다: 고정 IP-port 추출 시스템, 고정 IP-port 기반 응용 트래픽 분석 시스템, 트래픽 분석 결과 검증 시스템이다.

고정 IP-port 추출 시스템은 앞서 4장에서 설명한 고정 IP-port 추출 알고리즘에 기반한다. 고정 IP-port 기반 응용 트래픽 분석 시스템은 실시간으로 발생하는 네트워크 트래픽을 플로우 형태로 변환하고 추출된 고정 IP-port를 사용하여 응용 별로 분석한다. 트래픽 분석 결과 검증 시스템은 객관적 평가 요소를 사용하여 다양한 관점으로 분석된 결과를 검증한다.

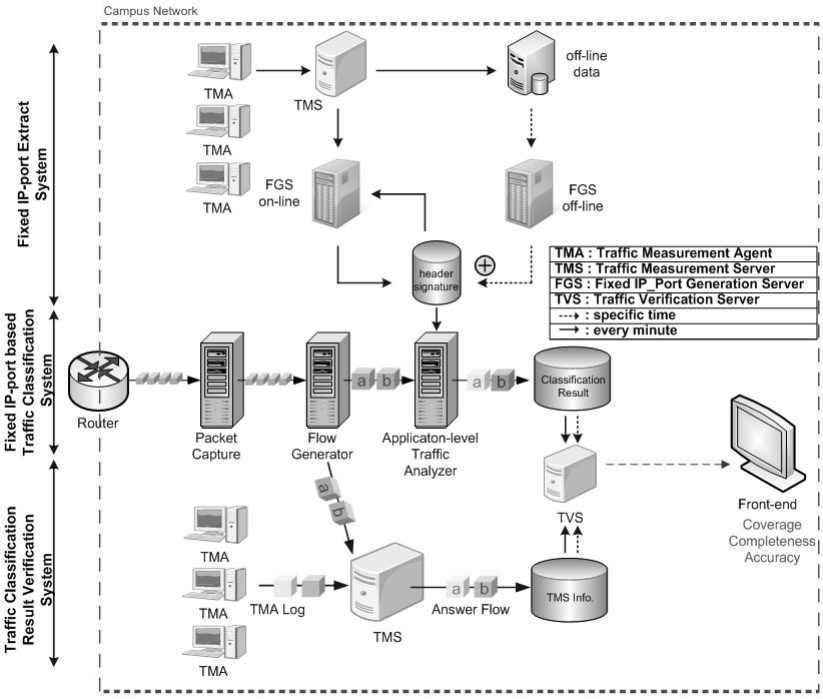
5.1 고정 IP-port 추출 시스템

고정 IP-port 추출 시스템은 입력으로 TMA(Traffic Measurement Agent) [11] 레코드와 플로우 데이터를 사용한다. TMA는 종단 호스트에 설치되는 에이전트로써 네트워크 소켓 정보(SrcIP, DstIP, SrcPort, DstPort, Protocol, 등)와 해당 소켓을 생성 시킨 프로세스의 이름과 설치된 경로 정보를 수집하여 주기적으로 서버로 보내준다. 2.5 장에서 설명한 [9]의 드라이버와 비슷한 기능을 하지만, 단순히 소켓 정보를 수집하므로 매우 가벼워 많은 종단 호스트에 설치 할 수 있다는 장점이 있다. 응용 프로세스에 의해 새로운 소켓이 생성될 때 마다 TMA는 소켓정보를 수집하여 FGS(고정 IP-port 추출 서버)와 TMS로 전송한다. 본 연구진은 학내 네트워크에 약 300대의 종단 호스트에 TMA를 설치하였다. 그 중 100대의 호스트는 고정 IP-port 추출을 위해 사용하였고 나머지는 트래픽 분석 결과 검증 시스템을 위해 사용하였다. 검증시스템은 5.3장에서 자세히 설명한다.

(그림 4)와 같이 고정 IP-port 추출 시스템은 두 부분으로 나뉘다. 실시간 추출 시스템과 off-line 추출 시스템이다. 실시간 추출 시스템은 급변하는 네트워크의 상황을 감안하여 실시간 추출을 하기 위해 매분 동작하는 시스템이다. 이에 반해, 장기간 많은 데이터를 유지하지 못하는 실시간 추출의 단점을 보완하기 위해 off-line으로 모아진 데이터에서 정기적으로 고정 IP-port를 추출하는 off-line 추출 시스템을 같이 동작시킨다.

5.2 고정 IP-port 기반 응용 트래픽 분석 시스템

고정 IP-port 기반 응용 트래픽 분석 시스템은 분석 대상



(그림 4) 고정IP-port 기반 분석 시스템 구조

이 되는 트래픽을 수집하는 부분과 분석하는 부분으로 나뉘다. 우선 패킷을 분석 대상 네트워크의 링크에서 수집(Packet Capture)한다. 수집된 패킷들을 이용하여 플로우를 만든다(Flow Generator). 본 시스템에서는 1분을 기준으로 플로우를 생성한다. 응용레벨 트래픽 분석기(Application-level Traffic Analyzer) 에서는 5 튜플(Source IP, Source port, Destination IP, Destination port, protocol)을 포함하는 플로우 헤더와 앞선 고정 IP-port 추출 시스템으로 생성된 고정 IP-port의 헤더를 비교함으로써 트래픽을 분석한다. 단순한 헤더 비교이므로 매우 빠른 특징이 있다. 따라서 이 모든 과정은 실시간으로 처리할 수 있다.

5.3 트래픽 분석 결과 검증 시스템

본 논문의 트래픽 분석 결과 검증 시스템은 다음과 같은 특징을 가진다.

(1) 매우 정확한 정답지(Ground truth)데이터를 사용한다.
(2) 객관적 평가요소를 사용한다.
(3) 실시간 검증과 off-line 검증이 가능하다.

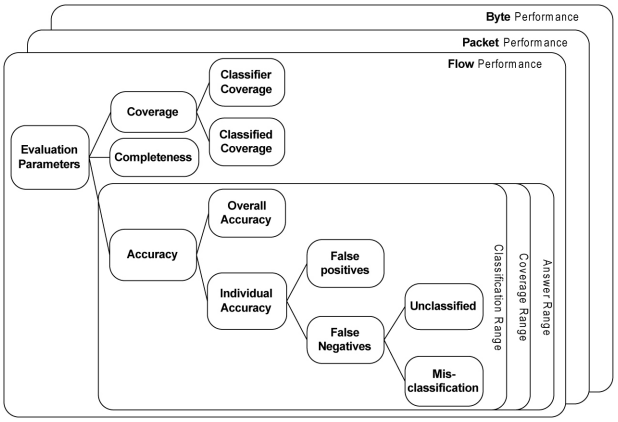
검증이란 분석 알고리즘에 의해 분석된 결과와 정답지 데이터를 비교하여 분석의 정확도(Accuracy)를 측정하는 것이다. 따라서, 2장에서 언급했듯이 검증에 있어 가장 핵심은 정확한 정답지를 생성하는 것이다.

본 논문에서는 5.1장에서 설명한 TMA를 이용한 검증을 제안한다. TMA는 단순한 소켓 정보의 수집을 통하여 앞선 연구의 단점인 오버헤드를 최소화하였다. 우리는 TMA를 추출과 검증 시스템, 두 시스템에서 사용하기 때문에 두 그룹으로 나누어 사용하였다. 한 그룹은 고정 IP-port를 추출하는 데에 사용하고, 다른 한 그룹은 분석된 결과를 검증하는 데에 사용하였다.

정확한 정답지 데이터의 중요성과 같은 맥락으로 객관적인 검증을 위한 평가 요소를 정의하는 것 또한 매우 중요하다. 본 논문에서는 분석 결과를 검증하기 위한 객관적 평가 요소를 제안한다. 제안된 평가 요소는 [10]에서 제안된 평가 요소를 기반으로 범위와 정확도 측면을 좀더 세분화 하여 정의하였다.

평가 요소는 범위(Coverage), 분석률(Complete-ness), 정확도(Accuracy), 와 같이 3가지 부분으로 구성된다. 첫 번째 평가 요소는 범위(Coverage)이다. 범위(Coverage)는 해당 알고리즘(분석기)이 분석 할 수 있는 응용의 개수인 Classifier Coverage와 실제 분석된 결과에 존재하는 응용의 개수인 Classified Coverage로 나뉜다.

두 번째 평가 요소인 분석률(Completeness)은 전체 트래픽 중에 해당 알고리즘이 분석한 트래픽의 비율을 나타낸다. 세 번째 평가 요소는 정확도(Accuracy)이다. 해당 알고리즘의 결과와 정답지 데이터를 비교하여 얼마나 알고리즘이 정확하게 분석하는지를 나타낸다. 정확도(Accuracy)는 다른 평가 요소와 달리 여러 검증 범위를 가진다. 검증 범위에 대한 내용은 평가 요소 설명 후 제시한다. 정확도(Accuracy)는 알고리즘 전반적인 Overall Accuracy와 각각의 응용에 적용되는 Individual Accuracy로 나뉜다. 특히 Individual Accuracy는 FP(False positive), FN(False Negative)으로 나타낸다. 응용 X의 FP란, 해당 알고리즘이 X가 아닌 응용을 X라 분



(그림 5) 객관적 검증 평가 요소

석한 것을 의미한다. 또한 응용 X의 FN이란, 해당 알고리즘이 X를 X가 아니라고 분석한 것을 의미한다. 특히, FN은 FN-Unclassified와 FN-Mis_Classification으로 나눌 수 있는데, 전자는 해당 알고리즘이 X를 분석하지 못한 것이고, 후자는 해당 알고리즘이 X를 다른 응용으로 분석한 것이다. 네트워크 관리의 트래픽 제어의 관점에서 본다면 FN-Unclassified보다 FN-Mis_Classification이 더 큰 위험성을 가진다. 즉, 잘못 분석하는 것은 아니 분석한 만 못하다는 의미이다.

본 검증 시스템의 특징은 다른 논문에서 제안하는 off-line 검증뿐만 아니라 실시간 검증을 가능하게 한다. off-line 검증은 알고리즘을 동일한 데이터를 대상으로 반복적으로 실험할 수 있기 때문에 알고리즘을 정교화하는데 유용한다. 또한, 여러 다양한 알고리즘을 비교 분석하는 것이 유용하다. 실시간 검증은 실제 네트워크에 개발된 알고리즘을 적용하였을 때 급변하는 네트워크 트래픽에 대해 현재 분석 시스템이 어떠한 상태인지를 확인하는 데에 있어 매우 유용하다.

6. 실험 및 결과

본 논문에서 제안한 고정 IP-port 기반의 분석을 위하여 학내 네트워크 트래픽을 대상으로 실험하였다. 실험은 Intel Dual-Core E2140 1.60GHz CPU와 3GB RAM 이 탑재된 범용 컴퓨터에서 수행하였다.

실험을 위해 학내 네트워크에서 5일 동안 트래픽을 수집하였다. <표 1>은 실험에 사용한 트래픽의 총량과 정답지(Ground truth)데이터의 량을 보인다.

〈표 1〉 실험에 사용한 트래픽 정보

	Flows(K)	Packets(M)	Bytes(G)
Day1	3,089 / 48,872	111 / 1,801	86 / 1,541
Day2	3,200 / 21,767	95 / 749	67 / 624
Day3	3,352 / 55,940	90 / 2,034	65 / 1,707
Day4	3,450 / 53,353	136 / 1,969	112 / 1,673
Day5	2,932 / 52,282	95 / 6,051	70 / 58,334

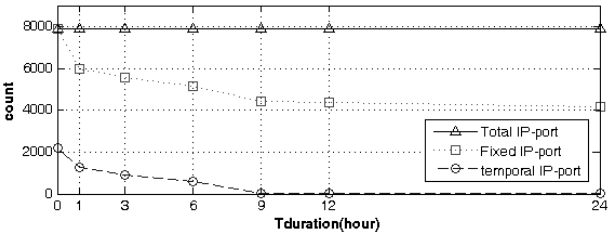
(Ground truth / Total)

6.1 임계값 설정

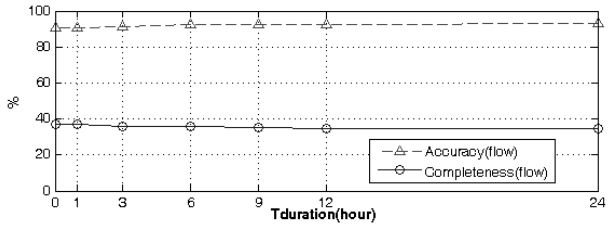
임계값($T_{duration}$)은 고정 IP-port 추출에 있어 불필요한 임시 IP-port를 제거하는 데에 목적이 있다. 임시 IP-port는 짧은 사용시간을 특징으로 가지기 때문에 적절한 임계값($T_{duration}$)을 설정하여 고정 IP-port와 구별 할 수 있다. 적절한 임계값($T_{duration}$)을 설정하기 위해 동일한 데이터를 대상으로 다양한 임계값($T_{duration}$)을 적용시켜 전체 IP-port의 개수와 고정 IP-port의 개수, 그리고 고정 IP-port에 포함된 임시 IP-port 개수를 확인하는 실험과 추출된 고정 IP-port를 사용하여 분석 하였을 때 얻은 분석률과 정확도를 조사하였다. 실험 시 다른 임계값의 영향을 최소화 하기 위해 임계값($T_{clienthost}$)은 1로 설정하고 임계값($T_{timelimit}$)은 무한대로 설정하였다.

(그림 6)은 임계값($T_{duration}$)에 따른 상태 전이 다이어그램의IP-port의 총 개수(Total IP-port)와 알고리즘에 의해 추출된 고정 IP-port의 개수(Fixed IP-port), 그리고 고정 IP-port에 속해있는 임시 IP-port의 개수(temporal)를 보여준다. 앞서 설명했듯이 고정 IP-port 알고리즘의 목표는 임시 IP-port를 제외한 고정 IP-port를 추출 하는 것이다. 임계값($T_{duration}$)을 9시간 이상으로 실험하였을 때, 임시 IP-port가 추출되지 않은 것을 확인 할 수 있었다. 이보다 낮은 임계값($T_{duration}$)은 임시 IP-port의 추출을 야기시켜 고정 IP-port 추출 시스템의 오버헤드를 증가 시킨다.

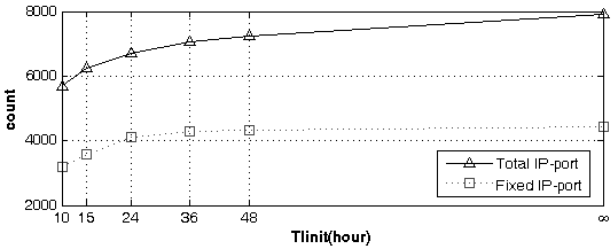
(그림 7)은 (그림 6)에서 추출한 고정 IP-port를 사용하여 Day 5트래픽을 분석한 결과의 분석률(Completeness)와 정확도(Accuracy)를 나타낸다. 그림에서도 확인 할 수 있듯이, 임계값($T_{duration}$)을 증가 시켜 임시 IP-port의 고정 IP-port에서의 비율을 낮추더라도 분석률에는 크게 영향을 미치지 않는 것을 확인할 수 있다. 또한 임계값($T_{duration}$)이 커질수록



(그림 6) 임계값($T_{duration}$)에 따른 IP-port 개수의 변화
($T_{clienthost}=1$, $T_{timelimit}=\infty$, Day1-5)



(그림 7) 임계값($T_{duration}$)에 따른 분석률 및 정확도 변화
($T_{clienthost}=1$, $T_{timelimit}=\infty$, Day1-5)



(그림 8) 임계값($T_{clienthost}$)에 따른 분석률및 정확도 변화
($T_{duration}=0$, $T_{timelimit}=\infty$, Day1-5)

좀 더 정확한 고정 IP-port를 추출 할 수 있기 때문에 정확도가 상승하는 것을 확인할 수 있다

임계값($T_{clienthost}$)은 고정 IP-port 추출에 있어 정확도에 영향을 준다. 임계값($T_{clienthost}$)이 높을 수록, 즉 여러 호스트와 통신 하는 고정 IP-port 일수록 더 높은 정확도를 보인다. 하지만 임계값이 높아질 수록 추출되는 고정 IP-port의 개수가 줄어들어 분석률은 떨어진다. 따라서 실험을 통해 적절한 임계값을 설정해야 한다. 적절한 임계값($T_{clienthost}$) 설정을 위해 앞선 실험과 비슷한 환경에서 임계값($T_{clienthost}$)에 따른 정확도와 분석률을 조사하였다.

(그림 8)에서 확인 할 수 있듯이 임계값($T_{clienthost}$)이 증가하면 분석률(Completeness)은 감소하지만, 정확도(Accuracy)는 증가한다. 트래픽을 발생하는 종단 호스트에서 트래픽을 수집하여 고정 IP-port를 추출하기 때문에 임계값($T_{clienthost}$)이 낮더라도 높은 정확도(Accuray)를 보인다. 따라서 본 네트워크에 있어 적절한 임계값($T_{clienthost}$)은 2이다.

임계값($T_{timelimit}$)은 추출 알고리즘이 실행 될 때 앞서 설명한 임계값($T_{timelimit}$)을 만족하지 못하여 한 상태에서 영원히 잔류하게 하는 것을 방지한다. 이러한 데이터들은 추출 알고리즘의 오버헤드를 증가 시켜 실시간 분석을 어렵게 한다.

(그림 8)은 고정 IP-port 추출 알고리즘에서 임계값($T_{timelimit}$)을 변경하였을 때 변동하는 모든 IP-port(Total IP-port)와 추출된 고정 IP-port(Fixed IP-port)의 개수의 변화를 보여준다. 임계값($T_{timelimit}$)을 48시간으로 정하였을 때 고정 IP-port의 손실 없이 전체 IP-port의 유지 량을 줄일 수 있었다.

본 장에서 설명한 세 가지 임계값들은 <표 1>에 제시된 데이터를 대상으로 실험한 값이다. 이러한 임계값들은 분석 대상 네트워크의 환경에 맞게 재조정되어 사용될 수 있다.

6.2 트래픽 분석

본 논문에서 제안한 고정 IP-port의 성능을 평가하기 위해 <표 1>에 기술한 데이터를 이용하여 실험을 하였다. 실험은 Day 1-4 동안 실시간으로 고정 IP-port를 추출하고 Day5 에 대해 실시간으로 추출과 분석을 동시에 수행 하였다. 실험에서 사용한 임계값은 6.1장에서 찾은 값으로 설정하였다.

<표 2>는 고정 IP-port 기반 트래픽 분석의 결과와 본 연구진이 개발한 LCS(Longest Common String) 기반의 Payload 시그니처 분석기[12]의 분석 결과를 보여준다. Payload

〈표 2〉 고정 IP-port 기반 트래픽 분석 결과

		Day 5	
		Fixed IP-port	Payload
Completeness	Flow(K)	24.83%	86.06%
	Packet(M)	5.15%	76.75%
	Byte(G)	3.99%	75.72%
Accuracy	Flow	99.91%	97.49%
	Packet	99.92%	94.18%
	Byte	99.91%	95.4%

시그니쳐 분석기와 비교해 보았을 때, 본 방법론은 높은 정확도를 가지지만 분석률은 상당히 낮았다. 분석률이 낮은 원인은 매우 낮은 비율의 정답지(Ground truth)데이터를 사용하였고 고정 IP-port로 분석 할 수 없는 동적, 임시 IP-port의 비율이 많았기 때문이다. 하지만, 약 7%의 정답지 데이터에서 추출 된 고정 IP-port를 사용하여 <표 2>와 같은 결과를 얻은 것은 고정 IP-port 기반 트래픽 분석이 의미가 있는 것을 알 수 있다.

FP로 분석되는 경우는 대표적으로 두 가지로 나눌 수 있다. 첫 번째는 특정 서버를 해당 서버의 목적과 다른 목적으로 통신하는 경우이다. Widnows 운영체제에서 DNS 서버의 접근은 svchost.exe가 담당한다. 하지만, abdobe사에서 개발한 특정 프로세스(mdnsresponder.exe)는 특수한 목적을 가지고 주기적으로 DNS 서버에 접근한다. 이러한 경우, DNS서버(svchost.exe)로 등록된 고정 IP-port는 FP로 나타난다. 두 번째 예는 FP의 가장 많은 비율을 차지하는 웹 기반 응용 프로그램이다. 현재 많은 사람들이 사용하는 응용 프로그램들은 웹 기반이고 이러한 응용은 자체 웹 브라우저를 사용한다. 즉, 응용 프로그램이 직접 웹 서버와 통신을 한다. 이러한 경우, 웹 서버(iexplore.exe)로 등록된 고정 IP-port가 FP로 나타난다. 이와 같은 예는 nateon, afreeca palyer 등과 같은 웹 기반 응용 프로그램에서 나타났다.

고정 IP-port기반 분석 방법은 매우 정확하지만, 기존의 다른 방법보다 낮은 분석률(Completeness)를 가진다. 따라서, 현재의 단순한 추출 알고리즘이 아닌, 좀 더 섬세한 추출 알고리즘이 필요하다. 또한, 매우 정확한 특징과 트래픽의 상관 관계를 이용한다면 좀 더 추가적인 분석이 가능하다.

<표 3>은 추출 한 고정 IP-port와 응용의 개수를 나타낸다. 분석 기간 동안 총 82개의 응용에 대해서 조사하였다. 그 중 72(약 88%)의 응용에서 3360개의 고정 IP-port를 추출할 수 있었다.

〈표 3〉 고정 IP-port 추출 알고리즘 성능

	Day 1-5
IP-port	3360 / 7187
Application	72 / 82

(Fixed State/ Total State)

6.3 오버헤드

고정 IP-port 기반 분석 시스템의 낮은 오버헤드를 보이기 위하여 페이로드 시그니처 기반 분석 시스템과 실행시간을 비교 하였다. <표 4>는 실험 결과를 보여준다.

<표 4>에서 확인 할 수 있듯이, 트래픽 페이로드를 일일이 확인해야 하는 페이로드 기반 트래픽 분석 방법보다 월등히 좋은 성능을 가진다. 빠른 분석 시간은 실시간 분석을 가능하게 한다.

〈표 4〉 트래픽 분석 시간 비교

	Min.	Max.	Avg.
Fixed IP port	24	210	81
Payload	1,252	45,765	17,235

(단위: msec)

7. 결론 및 향후 연구

인터넷 트래픽 분석은 네트워크 관리 측면에 있어서 그 중요성이 강조되고 있다. 많은 연구에서 제안된 트래픽 분석 방법들은 큰 오버헤드로 인하여 실제 네트워크 트래픽에 적용하기 어렵다. 따라서 본 논문에서는 매우 정확하고 가벼운 고정 IP-port 기반 분석 방법을 제안하였다. 고정 IP-port는 오직 한 응용에서 고정적으로 사용하는 {IP address, port number, transport protocol}의 쌍으로써 각각의 응용을 분석해서 자동적으로 수집할 수 있다. 이러한 고정 IP-port를 효과적으로 추출하기 위한 알고리즘을 제안하였다. 이 알고리즘은 다양한 환경에 적용할 수 있도록 유연한 임계값을 가지는 것이 특징이다. 따라서 본 논문에서는 가볍고 빠르며 정확한 실시간 트래픽 분석 시스템을 개발 하였다. 분석시스템은 추출, 분석, 검증 시스템으로 구성되어 있으면 모든 시스템은 실시간 및 Off-line 동작이 가능하다. 실제 네트워크 트래픽에 적용하였을 때 24.83%의 분석률과 99.91%의 정확도를 가졌다.

본 논문에서는 단순한 분석 알고리즘을 제안하는 것뿐만 아니라 다양한 알고리즘을 적용하고 검증 할 수 있는 분석, 검증 시스템을 제안하였다. 앞으로 이러한 시스템을 이용하여 고정 IP-port 기반 분석 방법을 기존에 연구된 방법들과 비교 분석하여 좀더 견고한 새로운 알고리즘 개발에 연구할 계획이다. 또한 본 연구에서는 고정 IP-port를 추출하기 위한 임계값을 실험을 통하여 결정하였지만, 다양한 네트워크에 적용되는 유효 임계값 범위를 정하는 연구가 필요하다.

참 고 문 헌

[1] Myung-Sup Kim, Young J.Won, James Won-Ki Hong, “Application-Level Traffic Monitoring and an Analysis on IP Networks,” ETRI Journal Vol.27, No.1, February, 2005.

[2] S. Sen, J. Wang, “Analyzing peer-to-peer traffic across large networks,” Internet Measurement Conference (IMC), Proc. Of the 2nd ACM SIGCOMM Workshop on Internet measurement, pp.137-150, 2002.

[3] W. Li et al. Efficient application identification and the temporal and spatial stability of classification schema. Computer Networks, 2009.doi:10.1016/j.comnet.2008.11.016.

[4] S. Sen, O. Spatscheck, D. Wang, “Accurate, Scalable In-Network Identification of P2P Traffic Using Application Signatures,” WWW 2004, New York, USA, May, 2004.

[5] S. Zander, T. Nguyen, and G. Armitage. Automated Traffic Classification and Application Identification using Machine Learning, In LCN’05, Sydney, Australia, Nov., 15-17, 2005.

[6] Jeffrey Erman, Martin Arlitt, Anirban Mahanti, "Traffic Classification Using Clustering Algorithms," Proc. of SIGCOMM Workshop on Mining network data, Pisa, Italy, Sep., 2006, pp.281-286.

[7] Andrew W. Moore and Denis Zuev, "Internet Traffic Classification Using Bayesian Analysis Techniques," Proc. of the ACM SIGMETRICS, Banff, Canada, Jun., 2005.

[8] T. Karagiannis, K.P apagiannaki and M.F aloutsos, “BLINC: Multilevel Traffic Classification in the Dark,” in Proc. of ACM SIGCOMM, August, 2005.

[9] Szabó, G., Orincsay, D., Malomsoky, S., Szabó, I.: On the validation of traffic classification algorithms. In: PAM. (2008) 72-81.

[10] Risso, F. Baldi, M. Morandi, O. Baldini, A. Monclus, P. Lightweight, Payload-Based Traffic Classification: An Experimental Evaluation. In proceeding of Communications, 2008. ICC '08. IEEE International Conference, 2008.

[11] Byung-Chul Park, Young J. Won, Myung-Sup Kim, James W. Hong, "Towards Automated Application Signature Extraction for Traffic Identification," Proc. of the IEEE/IFIP Network Operations and Management Symposium (NOMS) 2008, Salvador, Bahia, Brazil, Apri., 7-11, 2008, 160-167.

[12] 박준상, 박진완, 윤성호, 오영석, 김명섭, "응용 레벨 트래픽 분류를 위한 시그니처 생성 시스템 및 검증 네트워크의 개발", 2009년 제31회 한국정보처리학회 춘계학술발표대회(KIPS), 부산, 한화리조트, Apr., 23-24, 2009, 제16권 제1호, pp.1288-1291.

윤 성 호

e-mail : sungho_yoon@korea.ac.kr

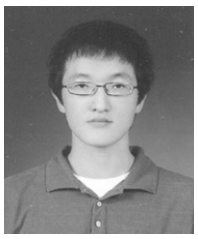
2009년 고려대학교 컴퓨터정보학과(학사)

2009년～현 재 고려대학교 컴퓨터정보학과 석사과정

관심분야 : 네트워크 관리 및 보안, 트래픽 모니터링 및 분석

박 준 상

e-mail : junsang_park@korea.ac.kr
2008년 고려대학교 컴퓨터정보학과(학사)
2008년～현 재 고려대학교 컴퓨터정보학과 석사과정
관심분야 : 네트워크 관리 및 보안, 트래픽 모니텅링 및 분석



박 진 완

e-mail : jinwan_park@korea.ac.kr
2009년 고려대학교 컴퓨터정보학과(학사)
2009년～현 재 고려대학교 컴퓨터정보학과 석사과정
관심분야 : 네트워크 관리 및 보안, 트래픽 모니텅링 및 분석

이 상 우

e-mail : sangwoo_lee@korea.ac.kr
2003년～현 재 고려대학교 컴퓨터정보학과 학사과정
관심분야 : 네트워크 관리 및 보안, 트래픽 모니텅링 및 분석

김 명 섭

e-mail : tmskim@korea.ac.kr
1998년 포항공과대학교 전자계산학과(학사)
1998년～2000년 포항공과대학교 컴퓨터공학과(석사)
2000년～2004년 포항공과대학교 컴퓨터공학과(박사)
2004년～2006년 Post-Doc., Dept. of ECE, Univ. of Toronto, Canada.
2006년～현 재 고려대학교 컴퓨터정보학과 조교수
관심분야 : 네트워크 관리 및 보안, 트래픽 모니터링 및 분석, 멀티미디어 네트워크