https://doi.org/10.22670/knom.2022.25.2.51

네트워크 트래픽 분석을 통한 규칙 기반 사용자 행위 탐지 시스템 설계

박 지 태◆, 백 의 준*, 신 창 의**, 최 정 우*, 김 명 섭°

A Design of Rule-based User Action Detection System for Network Traffic Analysis

Jee-Tae Park◆, Ui-Jun Baek*, Chang-Yui Shin**, Jeong-Woo Choi*, Myung-Sup Kim°

요 약

최근 네트워크 기술과 환경의 성장에 따라 다양한 어플리케이션이 발생하고 있으며, 네트워크 트래픽 양도 급격하게 증가하고 있다. 이러한 추세에 따라 네트워크 내 원활한 서비스 제공과 안전한 네트워크 보안을 위해서는 효율적인 네트워크 관리 방법이 필요하며, 이를 위해 오래전부터 다양한 연구가 수행되어 왔다. 여러 연구들 중 사용자 행위 탐지 연구는 네트워크 내 사용자들의 어플리케이션 사용 행위에 대한 탐지 및 모니터링을 수행하며, 네트워크 관리, 보안 및 지출 관리 등의 여러 분야에 도움을 준다. 본 논문에서는 정확한 사용자 행위 탐지를 목표로 규칙 기반의 사용자 행위 탐지 시스템을 제안한다. 제안하는 방법의 타당성을 검증하기 위해 실제 어플리케이션 트래픽을 수집하고, 본 연구의 선행 연구와 성능 비교 실험을 수행한다.

Key Words : Network Traffic Classification, User Action Detection, SaaS Application

ABSTRACT

Recently, various applications are occurring according to the growth of network technology and environment, and the amount of network traffic is rapidly increasing. In accordance with this trend, an efficient network management is required for smooth service provision and safe network security in the network, and various researches have been conducted for a long time for this purpose. Among several researches, user action detection research detects and monitors application usage behavior of users within the network, and helps in various fields such as network management, security, and expenditure management. In this paper, we propose a rule-based user action detection system with the goal of accurate user action detection. In order to verify the validity of the proposed method, we collect real application traffic and conducted performance comparison experiments with our previous research.

※ 이 논문은 2020년도 산업통상자원부 및 한국산업기술평가관리원(KEIT) 연구비 지원에 의한 연구이며 (No. 20008902, IT비용 최소화를 위한 5채널 탐지기술 기반 SaaS SW Management Platform(SMP) 개발), 2021년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과로 수행되었음 (2021RIS-004)
◆ First Author : Department of Computer and Information Science, Korea University, pjj5846@korea.ac.kr
° Corresponding Author : Department of Computer and Information Science, Korea University, tmskim@korea.ac.kr
* Department of Computer and Information Science, Korea University, {pb1069, choigoya97}@korea.ac.kr
** Defense Agency for Technology and Quality, superego99@dtaq.re.kr
논문번호 : KNOM2022-02-10, Received December 2, 2022; Revised December 10, 2022; Accepted December 16, 2022

Ⅰ. 서 론

오늘날에는 네트워크 환경의 증대와 기술의 발전으로 다양한 어플리케이션이 등장하고 있으며, 모바일, 클라우드 기반 등의 새로운 형태의 어플리케이션도 나타나고 있다[1-3]. 네트워크 서비스를 원활하게 제공하기 위해서는 효율적인 네트워크 관리 방안이 필요하며, 오래전부터 네트워크 관리를 위한 다양한 연구가 수행되어 왔다. 그 중 사용자 행위 탐지 연구는 네트워크 내 보안 및 관리 측면에서 중요한 역할을 한다. 관리자는 네트워크 트래픽을 입력으로 대상 어플리케이션에 대한 실제 사용 행위를 모니터링하며, 이를 통해 네트워크 관리에 유용한 여러 가지 정보를 얻을 수 있다. 또한, 해커들은 경제적, 정치적 등의 다양한 목적으로 공격 대상에게 악성 행위를 수행하며, 이러한 악성 행위는 점차적으로 고도화 되고 있다. 특히 IDS와 같이 네트워크 보안 관제 시스템에 탐지되지 않기 위해 사전에 공격 대상의 취약점을 분석하며, 이를 위해 수행되는 악성 행위가 정상 행위인 것처럼 위장한다[2,3]. 네트워크 보안 관제 시스템은 네트워크 내 사용자 행위에 대한 지속적인 모니터링을 통해 정상 행위로 위장된 악성 행위를 신속하고 정확하게 탐지해야 한다.

사용자 행위 탐지에 대한 연구는 각 연구 별로 목적에 따라 다르게 수행되어 왔으며, 대상 어플리케이션의 특성에 따라 사용자 행위 정의도 다르게 나타난다. 사용자 행위 정의는 행위 탐지 이전에 어떤 행위를 탐지 할 것인지 정의하는 것으로 다양하게 정의될 수 있다. 예를 들어 WeChat과 같은 메신저 형태 어플리케이션의 경우, 채팅방에 들어가기, 채팅 시작하기, 파일 전송 등과 같은 행위가 정의될 수 있으며, Twitter와 같은 SNS 형태 어플리케이션의 경우 게시물 올리기, 게시물 공유 등과 같은 행위가 정의 될 수 있다.

본 논문에서는 클라우드 서비스를 기반으로 하는 SaaS(Software as a Service) 어플리케이션을 대상으로 사용자 행위 탐지 연구를 수행한다. 클라우드 서비스는 네트워크가 연결되어 있는 환경에서 가상화된 컴퓨터 리소스를 여러 가지 서비스 형태로 제공하며, 그 중 SaaS는 리소스를 소프트웨어 형태로 제공하는 클라우드 서비스이다. 과거에는 사용자가 대상 어플리케이션을 1회성 구매 후 1회 설치를 통해 사용하였지만, SaaS의 경우 네트워크가 연결되어 있는 환경에서 라이선스 구독 후, 구독한 기간에 따라 사용 할 수 있다. 사용자 입장에서 번거로운 설치 과정 없이 바로 사용 가능하기 때문에 편리하며, 적절한 라이선스, 기간, 인원을 선택 할 경우에 기존의 1회성 구매에 비해 비용을 절감 시킬 수 있다는 장점이 있다. 특히 비용 절감은 회사, 학교와 같은 대규모의 네트워크 환경에서 더욱 크게 적용하기 때문에, 최근에는 많은 단체에서 SaaS를 사용하는 추세이다. 하지만 실제 사용하는 인원에 비해 더 많은 허용 인원을 가진 라이선스 혹은 필요하지 않은 기능이 포함된 라이선스를 구독할 경우 불필요한 지출이 발생할 수 있다. 따라서 SaaS 어플리케이션을 사용 할 때, 불필요한 지출을 줄이고 적절한 비용 관리를 위해 SaaS 어플리케이션에 대한 사용자 행위 탐지를 기반으로 하는 사용자 행위 모니터링이 필요하다.

본 논문에서는 클라우드 서비스를 기반으로 하는 SaaS 어플리케이션을 대상으로 효율적인 네트워크 관리와 불필요한 지출을 줄이는 것을 목표로 사용자 행위 탐지 시스템을 설계 한다. 먼저 대상 어플리케이션에 대한 4 가지의 사용자 행위를 사전에 정의하고, 대상 어플리케이션에 대한 트래픽 분석을 수행한다. 분석 결과를 바탕으로 사전에 정의한 사용자 행위를 탐지하기 위해 규칙 기반 행위 탐지 시스템을 제안한다.

본 논문은 1장 서론에 이어 2장에서는 관련 연구를 설명하고, 3장에서 SaaS 어플리케이션에 대한 행위 장의 및 제안하는 시스템의 전반적인 구조에 대해 설명한다. 4장에서는 제안하는 시스템을 검증하기 위해 수행된 실험에 대해 기술하며, 5장에서는 결론 및 향후 연구에 대해 설명하고 마친다.

Ⅱ. 관련 연구

2.1. 네트워크 어플리케이션 트래픽 분류

네트워크 어플리케이션 트래픽 분류 연구는 오래전부터 연구가 진행되어 왔으며, 대표적으로 포트 기반, 페이로드 기반 분석 방법이 있다[1-3]. 포트 기반 분석 방법은 고정된 포트 정보를 활용하여 어플리케이션을 분류하는 방법으로, 동적 포트 사용으로 현재는 잘 사용하지 않는 방법이다. 페이로드 기반 분석 방법은 암호화 되어있지 않은 페이로드 정보를 활용하여 고정된 페이로드 스트링 값을 활용하여 어플리

케이션을 분류하는 방법이다. 하지만 암호화 트래픽 사용의 증가로 페이로드 값이 모두 암호화되어 나타나기 때문에 현재는 잘 사용하지 않는다[2-5].

이러한 문제점을 해결하기 위해 학습 기반 분석 방법이 가장 활발하게 연구되고 있다[4, 5]. 이 방법은 어플리케이션 트래픽의 여러 가지 정보를 활용하여 특징들을 추출하고, 추출된 특징을 머신러닝 및 딥러닝 알고리즘을 활용하여 학습한다. 이 때, 페이로드 정보는 암호화되어있다고 가정하기 때문에 주로 플로우의 통계 정보나 헤더 정보 등을 활용하여 특징을 추출한다. 이 방법은 트래픽의 여러 정보를 고려하여 특징을 추출할 수 있으며, 분류 정확도가 다른 방법론에 비해 높다는 장점이 있다[4, 5]. 하지만 학습 특징에 대한 높은 의존도와 어플리케이션 트래픽 분류 분야에서 라벨링된 학습 데이터를 쉽게 구하기 어려우며, 많은 시간과 비용이 든다는 문제점이 있다[5].

2.2. 사용자 행위 탐지 연구

사용자 행위 탐지 연구는 오래전부터 네트워크 보안, 관리를 목적으로 수행되어 왔다[6-10]. 사용자 행위를 탐지하기 위해서는 먼저 대상으로 하는 어플리케이션의 사용자 행위를 정의해야 한다. 사용자 행위 정의는 어플리케이션의 형태(Ex. 메신저, 눈, 파일 작업 등)에 따라서 다양하게 정의 될 수 있다 또한 같은 어플리케이션이라고 하더라도 수행한 연구 목적에 따라 다르게 정의된다[6, 7].

논문 [8]에서는 메신저 형태의 카카오톡(KakaoTalk)을 대상 어플리케이션으로 선정하였다. 카카오톡은 메신저 기반의 신속성과 경량화를 반영하여 TCP/IP 기반의 독자적인 프로토콜 LOCO를 사용한다. 논문 [8]에서는 카카오톡을 대상으로 11 가지의 사용자 행위를 정의하고, 정의된 행위를 Random Forest 알고리즘을 활용하여 99.7%의 정확도로 분류한다.

논문 [9]에서는 메신저 형태의 WeChat 을 대상 어플리케이션으로 선정하였다. WeChat은 카카오톡과 유사하게 메신저 기반의 신속성과 경량화를 반영하여 HTTP, TCP 기반의 독자적인 프로토콜 MMTLS를 사용한다. 9 가지의 사용자 행위를 정의하고, 정의된 행위를 6 가지의 학습 기반의 알고리즘을 활용하여 비교 실험을 수행한다. 여러 가지 알고리즘 중 Random Forest에서 96%의 F1-measure로 가장 좋은 결과를 도출한다.

논문 [10]에서는 SNS 형태의 어플리케이션으로 Instagram을 대상으로 9 가지의 사용자 행위를 정의하고, 정의된 행위를 학습 기반의 SVM 알고리즘을 활용하여 행위를 탐지한다.

본 연구의 선행 연구인 논문 [11]에서 Adobe Creative Cloud를 활용하여 페이로드 시그니처를 추출하고 탐지 실험을 수행한다. 하지만 논문에서는 단순하게 Adobe Creative Cloud에 대한 추출된 페이로드 시그니처를 제시하고, 실제로 수행한 실험에 대해서는 객관적인 자료를 제시하고 있지 않다. 또한 SNI 정보만을 가지고 행위를 탐지 할 경우, 특정 행위가 고정되지 않은 SNI 정보를 사용하는 경우에 탐지를 못하거나 잘못 탐지 할 수 있다는 문제점이 있다.

따라서 본 논문에서는 SaaS 어플리케이션에 대한 사용자 행위 탐지를 목표로, 가장 널리 사용되는 SaaS 어플리케이션 중 하나인 Adobe Creative Cloud를 대상 어플리케이션으로 선정하였다. 또한 선행 연구인 [11]의 시그니처 기반의 행위 탐지 방법과 성능 비교 실험을 수행하여 본 논문의 타당성을 검증한다.

Ⅲ. 본 론

3.1. 사용자 행위 정의

본 장에서는 대상 어플리케이션에 대한 사용자 행위를 정의한다. 이전에 언급한대로 사용자 행위는 연구 목적에 따라 여러 가지로 정의 될 수 있다 본 논문에서는 SaaS 어플리케이션의 불필요한 지출 줄이는 것을 목표로 연구를 수행하기 때문에 사용자의 실사용 기록과 관련된 4 가지 행위(Ex. 어플리케이션 시작, 로그인, 로그아웃 어플리케이션 종료)를 탐지 할 사용자 행위로 정의한다.

3.2. 규칙 기반 행위 탐지 시스템 설계

제안하는 규칙 기반 행위 탐지 시스템의 전체 구조는 크게 두 가지 모듈 (i.e. 규칙 생성 및 행위 탐지 모듈)로 구성되어 있으며, 그림 1에 나타나있다.

규칙 생성 모듈은 대상 어플리케이션에 대한

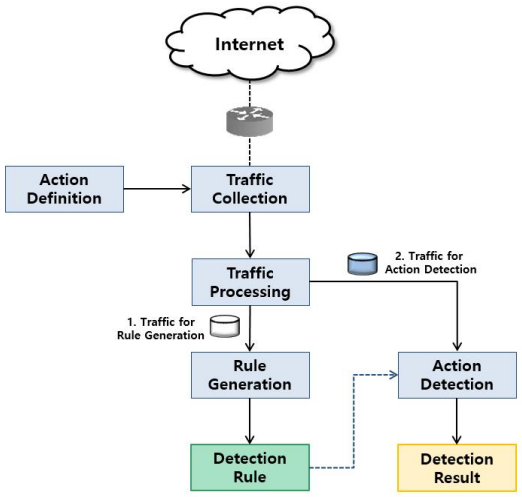

그림 1. 규칙 기반 행위 탐지 시스템 구조
Fig.1. A Structure of Rule based Action Detection System

분석 후 각 행위 별로 공통으로 발생하는 여러 가지 트래픽 정보를 활용하여 탐지 규칙을 생성하는 과정이며, 행위 탐지 모듈은 수집 된 트래픽과 생성된 탐지 규칙을 입력으로 실제 사용자 행위에 대한 탐지를 수행하는 과정이다.

먼저 Action Definition은 이전의 3.1장에서 설명한대로 대상 어플리케이션의 탐지 행위를 정의한다. 다음으로 Traffic Collection에서 대상 어플리케이션에 대한 트래픽을 수집하며, Wireshark를 사용하여 pcap 파일을 수집한다. 이후에 진행되는 과정에 따라 두 가지(i.e. 규칙 생성, 행위 탐지)로 나누어 수집된다. 행위 탐지에는 단순히 사용자 행위를 탐지하는 것이 아니라 어느 시점에 어느 호스트에서 어떤 정보를 통해 어떤 행동이 탐지되는지를 고려해야하기 때문에 트래픽 수집 단계에서 각 행위가 수행된 시간과 행위를 수행한 호스트 IP가 함께 기록되며, 로그 파일 형태로 저장한다. 즉, 규칙 생성 단계에서는 기록된 정보와 수집된 규칙 생성용 트래픽을 활용하여 보다 정확한 규칙을 생성하고, 행위 탐지 단계에서는 기록된 정보와 수집된 행위 탐지용 트래픽을 활용하여 생성된 규칙을 검증한다.

Traffic Processing에서 패킷 형태의 트래픽 파일을 플로우 형태의 파일로 변환한다. 플로우는 수집 된 패킷 중에서 5-tuples(i.e. Source IP, Source Port, Protocol, Destination IP, Destination Port) 정보가 같은 패킷들의 집합으로 정의한다. 수집 된 트래픽 셋은 이후에 수행되는 과정에 상관없이 동일한 전처리 과정을 거친다.

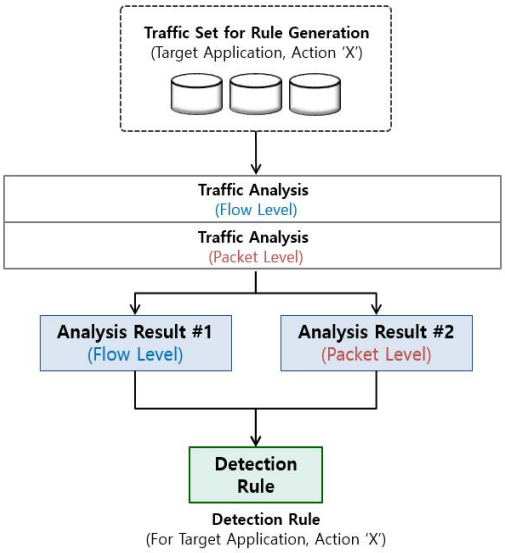

그림 2. 탐지 규칙 생성 과정
Fig.2. A Process of Detection Rule Generation

Rule Generation은 전처리 과정을 거친 트래픽에 대한 분석을 수행하고, 이를 기반으로 규칙을 생성하는 과정이며, 그림 2에 나타나있다. 탐지 규칙은 대상 어플리케이션의 행위들을 탐지하는 조건으로 각 행위 별로 공통적으로 발생하는 여러 가지 정보(Ex. 헤더, SNI 정보 등)로 구성되어 있으며, 행위 탐지 규칙에 대한 예시는 표 1에 나타나있다. 규칙 생성 과정은 대상 어플리케이션의 여러 가지 트래픽 데이터 셋을 활용하여 플로우, 패킷 레벨의 분석을 수행한다. 수행되는 트래픽 분석은 대상 어플리케이션의 개별 행위에 대해 여러 가지 트래픽 셋을 수집하고, 수집된 트래픽 셋을 확인하여 공통으로 발생하는 플로우 혹은 패킷을 찾는 것을 목표로 한다. 하지만 수행되는 분석은 수작업으로 공통 플로우, 패킷을 찾기 때문에 여러 가지 문제점을 발생시킨다. 따라서 추후에 기존 분석 방법의 문제점을 해결하기 위한 방법을 연구 할 예정이다. 수행 후 통해 도출되는 각각의 트래픽 셋 별 분석 결과를 취합하고, 공통적으로 발생하는 정보를 추출하여 탐지 규칙으로 정의한다.

생성된 행위 탐지 규칙은 크게 두 가지로 나

누어져있으며, 탐지 할 대상 어플리케이션과 행위로 이루어진 탐지 대상과 행위에서 공통으로 도출되는 트래픽 정보로 이루어진 탐지 규칙으로 구성되어 있다. 또한 탐지 규칙은 트래픽 형태에 따라 플로우, 패킷 기반 정보와 현재 탐지된 상태에 따라 선행 탐지 되어야 하는 행위를 나타내는 상태 정보로 구성된다. 어플리케이션 정보는 탐지 할 어플리케이션 정보를 뜻하며, 탐지 행위는 탐지 할 행위 정보를 나타낸다. 예를 들어 Microsoft Office 365의 어플리케이션 시작에 대한 규칙일 경우, 어플리케이션 정보는 Microsoft Office 365, 탐지 행위는 어플리케이션 시작으로 정의한다.

표 1. 행위 탐지 규칙 예시
Table 1. An Example of Action Detection Rule

규칙		정보 및 설명	예시
탐지 대상	어플리케이션 규칙	어플리케이션 정보	Adobe Creative Cloud
		탐지 행위	Login
탐지 규칙	플로우 기반 정보	클라이언트 IP	Any
		클라이언트 포트	Any
		프로토콜	TCP
		서버 IP	Any
		서버 포트	443
	패킷 기반 정보	확인 할 패킷 번호 (CPN)	4
		패킷 방향	CS (Client to Server)
		SNI	www.adobe.com
	상태 정보	선행되어야 할 행위 정보	Application Start

규칙 내 플로우 기반 정보는 각 행위 별로 공통적으로 도출된 플로우 정보이며, 클라이언트는 행위를 수행한 호스트, 서버는 호스트와 통신 한 목적지를 나타낸다. 특정 어플리케이션에서 행위 수행 할 때 고정된 플로우를 사용하여 통신 할 경우에 해당 플로우의 헤더 정보를 규칙으로 정의한다.

패킷 기반 정보는 각 행위 별로 공통적으로 도출되는 패킷 정보이며, 대부분의 어플리케이션에서 사용하는 TCP 기반의 TLS/SSL 패킷 페이로드는 암호화가 되어있기 때문에 암호화가 되어있지 않은 Handshake 패킷의 정보를 확인한다. 규칙 내 패킷 기반 정보 중 확인 할 패킷 번호는(Check Packet Number, CPN) 몇 번째의 패킷을 확인 하는지에 대한 정보이며, 패킷 방향은 확인 할 패킷의 통신 방향을 나타내며, 통신 방향은 CS(Client to Server)과 SC(Server to Client)으로 구분된다. SNI(Server Name Indication)는 TLS의 Handshake 과정 초기에 클라이언트가 어느 호스트명에 접속하려는지 서버에 알리는 역할을 하며, 문자열 값으로 구성되어 있다.

상태 정보는 탐지 할 행위가 수행되기 전에 어떠한 다른 행위가 선행되어야 하는지를 나타낸다. SaaS 어플리케이션에 대한 사용자 행위는 개별적으로 실행되지 않고 행위에 따라 순차적으로 실행되며, 다른 행위와 연관성을 가지기 때문에 현재 상태에 따라 탐지되는 정보가 달라 질 수 있다. 예를 들어, 4 가지의 행위 별로 어플리케이션 시작은 “None”, 로그인은 “어플리케이션 시작”, 로그아웃은 “로그인”, 어플리케이션 종료는 “어플리케이션 시작”이 선행되어야 하는 행위이다. 상태 정보는 오탐지를 줄이기 위해 정의하는 정보로, 선행된 행위를 확인하여 불필요한 오탐지 및 중복 탐지를 줄이는 역할을 한다.

생성된 규칙은 대상 어플리케이션의 하나의 행위에 대한 탐지 규칙으로, 최종적으로 하나의 어플리케이션에서 4 가지의 행위를 대상으로 4 가지의 개별 행위 탐지 규칙이 도출된다. 이후, 생성된 4 가지의 개별 행위 탐지 규칙은 대상 어플리케이션에 대한 하나의 탐지 규칙으로 취합된다.

Action Detection은 생성된 하나의 어플리케이션 행위 탐지 규칙과 사전에 수집된 행위 탐지용 트래픽 셋을 입력으로 사용자의 행위를 탐지하는 과정이다. 탐지 결과는 정탐지, 오탐지, 미탐지로 분류되며, 호스트 IP, 어플리케이션 정보, 행위 정보, 행위 수행 시점의 정보가 모두 일치 할 경우 정탐지로 판단하며, 하나의 정보라도 틀릴 경우 오탐지, 탐지하지 못할 경우 미탐지로 판단한다.

탐지 결과를 판단하는 기준은 사전에 트래픽 수집 단계에서 저장된 로그 파일을 활용한다. 세 가지의 탐지 결과를 활용하여 Recall, Precision, F1-measure를 계산하며, 수식은 아래의 수식 (1), (2), (3)에 나타나있다.

$$Recall = \frac{TP}{TP+FP} \quad (1)$$

$$\mathrm{Precision} = \frac{TP}{TP+FN} \quad (2)$$

$$F1-measure = \frac{2 \times \mathrm{Precision} \times Recall}{\mathrm{Precisio} + Recall} \quad (3)$$

도출된 탐지 결과에서 F1-measure 값이 특정 값 α보다 낮을 경우 잘못 생성된 규칙으로 판단하여 폐기하고 다른 정보를 통해 규칙 생성 과정을 반복한다. α 값은 임의로 설정하는 값으로, 본 연구에서 수행된 실험에서는 α 값을 0.95로 설정하였다.

Ⅳ. 실험 및 평가

제안하는 방법을 검증하기 위해 생성된 규칙을 활용한 행위 탐지 실험을 수행한다. 실험에는 대상 어플리케이션은 Adobe Creative Cloud를 사용하며, 사전에 언급한대로 어플리케이션 시작, 로그인, 로그아웃, 어플리케이션 종료의 4 가지 행위를 탐지 행위로 정의한다.

표 2. 트래픽 데이터 셋 정보
Table 2. An Information of Traffic Data Set

Purpose	App.	Trace	Traffic Information	
			Flow	Packet
Rule Generation	Adobe Creative Cloud	#1	985	8,022
		#2	1,021	8,258
		#3	1,125	10,552
		#4	845	7,125
		#5	625	5,255
		#6	1,225	13,758
		#7	328	6,255
		#8	512	5,220
		#9	524	8,802
		#10	425	7,775
Action Detection		#1	3,028	50,952
		#2	2,425	92,022
		#3	2,012	48,125
		#4	1,995	80,122
		#5	2,452	100,255

제안하는 방법의 타당성을 검증하기 위해 본 논문의 선행 연구인 시그니처 기반의 행위 탐지 방법[11]과 비교 실험을 수행한다. 시그니처를 활용한 행위 탐지 방법은 대상 어플리케이션에 대한 패킷 기반의 분석을 수행하고, 도출된 SNI 정보를 활용하여 행위를 탐지하는 방법이다.

실험에 사용한 트래픽 셋에 대한 정보는 표 2에 나타나있으며, Wireshark를 활용하여 트래픽 셋을 수집하였다. 총 15가지 Trace의 트래픽 셋을 수집하였으며, 규칙 생성을 위한 10 Trace와 행위 탐지를 위한 5 Trace로 구분된다. 규칙 생성 트래픽 셋에서는 Trace 별로 4 가지의 개별 행위를 1회 수행하였으며, 행위 탐지 트래픽 셋에서는 Trace 별로 4 가지의 개별 행위를 5회 수행하였다. 실험 결과는 행위 탐지 트래픽 셋을 대상으로 각 Trace 별로 전체 20회(4 가지 행위, 5회) 수행된 행위를 대상으로 정탐지(TP), 오탐지(FP), 미탐지(FN)을 확인한다.

실험 결과는 표 3에 나타나있으며, 각 Trace 별로 전체 20회의 행위 중 정탐지, 오탐지, 미탐지에 따라 계산한 Recall, Precision, F1-measure 값과 실험에 사용한 5 가지 Trace 결과 값의 평균을 기술하였다.

실험 결과로는 선행 연구에서 약 70~100%의 Recall과 70~89%의 Precision이 도출되었으며, 평균으로 약 78%의 Recall, Precision, F1-measure 값이 나타난다. 제안하는 방법은 약 88~100%의 Recall과 80~100%의 Precision이 도출되었으며, 평균 92%의 Recall, Precision, F1-measure 값이 나타난다.

표 3. 행위 탐지 결과
Table 3. Result of Action Detection

Method	Trace	Detection Result		
		Recall (%)	Precision (%)	F1-measure
Previous Method [11]	#1	87.50	77.77	82.35
	#2	78.95	71.42	74.99
	#3	78.75	89.69	83.86
	#4	100	50	66.67
	#5	82.50	71.42	76.56
	Avg.	85.54	72.06	76.89
Proposed Method	#1	100	80	88.89
	#2	94.44	89.47	91.89
	#3	95	100	97.43
	#4	88.89	94.44	91.58
	#5	94.44	89.47	91.89
	Avg.	94.55	90.68	92.34

제안하는 규칙 기반의 행위 탐지 방법은 시그니처 기반의 행위 탐지 방법에 비해 성능이 향상되었음을 알 수 있으며, 특히 Precision에서 상대적으로 차이가 많이 난다. 이는 시그니처 기반의 행위 탐지 방법은 SNI 정보만을 사용하기 때문에, 고정적인 SNI 정보를 사용하지 않거나 잘못된 SNI 정보를

시그니처로 정의 할 경우, 오탐지가 높게 발생 할 수 있다.

실제로 탐지 결과를 살펴보면 시그니처 기반의 분석 방법에서 SNI 정보를 시그니처로 잘못 정의되거나 실제로 특정 행위에서 고유하게 발생하는 정보가 아니라 다른 행위에서도 빈번하게 발생하는 SNI 정보일 경우가 다수 발생하여 오탐지 비율이 높게 나타난다.

Ⅴ. 결 론

본 논문에서는 사용자 행위 탐지 연구에 대해 소개하고, 행위 탐지 연구의 필요성에 대해 설명한다. 또한, 관련 연구에 행위 탐지 연구에 대한 기존의 여러 연구와, 본 연구의 선행 연구에 대해 소개하고, 선행 연구에 대한 문제점에 대해 언급한다.

본 논문에서는 관련 연구에서 언급한 선행 연구의 문제점을 해결하기 위해 규칙 기반의 행위 탐지 시스템을 제안한다. 제안하는 시스템은 규칙 생성과 행위 탐지의 두 가지 모듈로 구성되어있으며, 본론에서 각 모듈의 세부 구조에 대해 설명하였다.

제안하는 방법의 타당성을 검증하기 위해 선행 연구와 동일한 트래픽 셋을 활용하여 비교 실험을 수행 하였다. Adobe Creative Cloud를 대상으로 10 가지의 규칙 생성 트래픽 셋과 5 가지의 행위 탐지 트래픽 셋을 수집하였으며, 5 가지의 행위 탐지 트래픽 셋을 대상으로 성능 비교 실험을 수행하였다. 실험 결과로 제안하는 규칙 기반의 행위 탐지 방법이 전반적으로 더 높은 성능을 보이며, 특히 선행 연구와 제안하는 방법은 Precision에서 상대적으로 크게 차이가 났다. 이는 앞서 언급한 선행 연구의 문제점으로 오탐지 비율이 높게 나타나는 것으로 판단된다.

하지만 제안하는 방법에서 규칙 생성 할 때, 수행되는 트래픽 분석, 공통 특징 추출, 규칙 생성 등의 과정이 수작업으로 수행되기 때문에 많은 시간과 노력이 소모된다. 또한, 수작업으로 트래픽 분석 및 공통 특징 추출을 하기 때문에 선행 연구의 문제점에서 제시한 부정확한 정보를 규칙으로 잘못 정의하는 경우도 발생하였다. 또한, SNI 정보에 대한 의존도가 너무 높기 때문에, SNI 암호화와 같은 상황이 특정 어플리케이션에서 발생 할 경우에 탐지 정확도가 크게 낮아질 것으로 예상된다.

따라서 향후 연구로는 수동적 규칙 생성 방법을 개선하여 규칙 생성 시 수행되는 각각의 과정들을 일련의 과정으로 정리하고, 수작업으로 수행되는 분석 및 추출 과정을 자동으로 수행 될 수 있는 자동 규칙 생성 방법을 연구 할 예정이다. 또한 SNI 정보에 대한 의존도를 낮추기 위해 트래픽의 통계적 정보와 같은 다른 트래픽 특성을 함께 고려하여 규칙을 생성하는 방안을 연구 할 예정이다. 그리고 보다 다양한 어플리케이션을 대상으로 다른 방법론과 비교 실험을 통해 본 연구의 타당성을 추가로 검증할 예정이다.

References

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

[2] A. Dainotti, A. Pescape and K. Claffy, “Issues and Future Directions in Traffic Classification,” Network IEEE, Vol. 26, no. 1, 2012, pp. 35-40.

[3] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, “A Survey of Intrusion Detection Techniques in Cloud,” J. Network Computer Applications, vol. 36, no. 1, 2013, pp. 42 – 57.

[4] T. T. T. Nguyen and G. Armitage, “A Survey of Techniques for Internet Traffic Classification using Machine Learning,” IEEE Communications Surveys and Tutorials, Vol. 10, 2008, pp. 56-76.

[5] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, “Malware Traffic Classification using Convolutional Neural Network for Representation Learning,” in Proc. of 2017 International Conference on Information Networking (ICOIN), IEEE, Jan, 2017, pp. 712 – 717.

[6] M. Conti, L. V. Mancini, R. Spolaor and N. V. Verde, ″Analyzing Android Encrypted Network Traffic to Identify User Actions,″ in IEEE Transactions on Information Forensics and Security, vol. 11, no. 1, Jan. 2016, pp. 114-125.

[7] A. Shahraki, M. Abbasi, A. Taherkordi and A. D. Jurcut, ″Active Learning for Network Traffic Classification: A Technical Study,″ in

IEEE Transactions on Cognitive Communications and Networking, vol. 8, no. 1, March 2022, pp. 422-439.

[8] K. Park and H. Kim. "Encryption Is Not Enough: Inferring user activities on Kakaotalk with traffic analysis" International Workshop on Information Security Applications (WISA), 2015, pp. 254-265, Springer, Cham

[9] C. Hou, J. Shi, C. Kang, Z. Cao and X. Gang, "Classifying User Activities in the Encrypted WeChat Traffic," 2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC), 2018, pp. 1-8.

[10] H. Wu, Q. Wu, G. Cheng and S. Guo, "Instagram User Behavior Identification Based on Multidimensional Features," IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2020, pp. 1111-1116,

[11] 이민성, 최정우, 권윤주, 박지태 "페이로드 시그니처를 이용한 Adobe 소프트웨어 사용자 행위 탐지". 2021년도 통신망운용관리 학술대회 (KNOM 2021), 2021, pp. 24-25.

박 지 태 (Jee-Tae Park)

2017년 : 고려대학교 컴퓨터정보학과 학사

2017년 - 현재 고려대학교 컴퓨터정보학과 석박사통합과정

<관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석,

백 의 준 (Ui-Jun Baek)

2018년 : 고려대학교 컴퓨터정보학과 학사

2018년 - 현재 고려대학교 컴퓨터정보학과 석박사통합과정

<관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석

신 창 의 (Chang-Yui Shin)

2003년 : 육군사관학교 운영분석학과 학사

2007년 : 고려대학교 전자컴퓨터공학과 석사

2022년~현재 : 고려대학교 컴퓨터정보학과 박사과정

<관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 분석

최 정 우 (Jung-Woo Choi)

2022년 : 고려대학교 컴퓨터정보학과 학사

2022년~현재 : 고려대학교 컴퓨터정보학과 석사과정

<관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석

김 명 섭 (Myung-Sup Kim)

1998년 : 포항공과대학교 전자계산학과 학사

2000년 : 포항공과대학교 전자계산학과 석사

2004년 : 포항공과대학교 전자계산학과 박사

2006년 : Dept. of ECS, Univ of Toronto Canada

2006년~현재 : 고려대학교 컴퓨터정보학과 교수

<관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석, 멀티미디어 네트워크