페이로드 시그니처를 이용한 마이크로소프트 Office 365 서비스 탐지

이민성, 박지태, 최정우, 김명섭
고려대학교
{min0764, pjj5846, choigoya97, tmskim}@korea.ac.kr

Microsoft Office 365 Service Detection based on payload signature

Lee Min-Seong, Park Jee-tae, Choi jeong-woo, Kim Myung-Sup
Korea Univ.

요 약

SaaS는 클라우드 기반 소프트웨어 제공 모델로서 최근 많은 기업들이 사용을 하고 있다. SaaS 사용이 활성화 되면서 많은 기업에서 다루고 있다. 기업은 관리 부서를 통해 소프트웨어를 구매 및 관리해야 한다. 하지만 승인되지 않은 클라우드 소프트웨어를 사용하게 되면서 기업에서도 관리하지 못하는 Shadow IT 문제가 발생하고 있다. 이러한 리스크를 관리하면서 소프트웨어의 구매 및 비용 절감의 효율적인 자산 운용을 위하여 SaaS 사용 모니터링이 필요하다. 따라서 본 논문에서는 SaaS 소프트웨어 중에서도 마이크로소프트에서 제공하는 Office 365의 트래픽을 분석하여 어떤 서비스를 사용하고 있는지 탐지하는 방법론에 대해 제안한다.

Ⅰ. 서 론

SaaS는 클라우드 기반 소프트웨어 제공 모델로서, 마이크로소프트의 Office 365, 구글의 G Suite, Adobe의 Creative Cloud 등이 있다. SaaS 어플리케이션은 초기 도입 비용이 상대적으로 적고 필요에 따라 사용을 조절할 수 있기 때문에 기업에서 지속적으로 사용을 하고 있다. 기존 소프트웨어를 SaaS 어플리케이션 형태로 제공하는 소프트웨어 회사들도 증가하고 있다. 이에 따라 소프트웨어의 관리의 새로운 방식이 필요하게 되고, SaaS 어플리케이션을 모니터링 하여 관리를 할 수 있는 방법의 연구가 진행되고 있다.

모니터링을 통한 관리가 필요한 가장 중요한 이유는 Shadow IT 문제이다. Shadow IT란 승인하지 않은 클라우드 소프트웨어를 구입하고, 이를 IT 관리부서나 책임자가 파악하지 못하는 현상을 뜻한다. 기업에서 SaaS 어플리케이션을 구매하고 이용할 때, 클라우드 소프트웨어의 관리를 하지 못하는 부분들이 발생하게 된다. 이러한 문제를 해결하기 위하여 기업에서는 증가하고 있는 기업의 리스크를 관리 할 수 있고 소프트웨어의 구매 및 비용 절감 등의 효율적인 자산 운용을 위한 Shadow IT를 포함 할 수 있는 SaaS 사용 모니터링이 필요하다.

트래픽 분석방법론에는 IP / Port 기반 분석 방법, 페이로드 시그니처 기반 분석 방법, 통계 정보를 이용한 분석 방법, 머신 러닝을 이용한 분석 방법 등이 있다. 각각의 방법론들은 장단점이 존재하지만 본 논문에서는 IP / Port 기반, 페이로드 시그니처 기반 분석 방법을 통하여 사용된 마이크로소프트 서비스를 탐지한다.

우리가 자주 사용하고 쉽게 접할 수 있는 SaaS 소프트웨어는 마이크로소프트에서 제공하는 서비스들이다. 파워포인트, 엑셀, 워드 등 대부분의 기업에서도 문서 작성을 위해, 발표를 위해 제일 많이 사용되고 있는 서비스들이다. 본 논문에서는 SaaS 사용 모니터링을 위하여 SaaS 소프트웨어 중에서도 마이크로소프트에서 제공하는 Office 365의 서비스에 대한 트래픽 분석을 진행한다. 트래픽 분석을 통해 사용자가 Office 365의 어떤 서비스(파워포인트, 엑셀, 워드)를 이용하고 있는지 알 수 있는 분석 방법을 제시한다.

본 논문의 구성은 서론에 이어, 2장에서 마이크로소프트 서비스 탐지를 위한 트래픽 분석 방법론에 대해 기술하고, 마지막으로 결론 및 향후 연구에 대해 기술한 후 마친다.

Ⅱ. 본 론

본 장에서는 마이크로소프트 Office 365 트래픽을 분석하여 사용자가 어떤 서비스를 사용하고 있는지 탐지하는 방법론에 대하여 기술한다. Office 365의 트래픽을 분석하기 위해서, 단순히 마이크로소프트와 통신하는 트래픽만을 분석해야 한다는 생각을 할 수 있다. 하지만 아카마이 서버와 통신하는 트래픽도 분석을 진행해야 한다. 아카마이는 콘텐츠 전송 네트워크로 대부분의 컨텐츠를 사용자에게 제공해 주어야 하는 기업들이 사용하고 있다. 사용성이 높고 효율이 좋은 컨텐츠를 사용자에게 전달하기 위해 사용되고 있다. 특히 SaaS의 경우, 사용자가 많이 사용하게 되면 트래픽이 많이 발생하게 된다. 하지만 컨텐츠 전송 네트워크를 사용하여 트래픽의 부담을 덜어주고, 사용자에게는 좋은 컨텐츠를 전송해 줄 수 있다. 트래픽 분석을 진행하면 마이크로소프트 트래픽과 함께 아카마이와 통신하는 트래픽도 다양하게 발생을 하며, 서비스를 탐지하는 데에 있어 중요한 역할을 한다. 실제 분석 결과, 아카마이 서버에서 발생하는 트래픽을 분석하였을 때, 여러 가지 마이크로소프트 관련 정보를 유추할 수 있는 페이로드가 발생하고 있다.

트래픽 분석에 있어 IP / Port 기반 분석 방법과 페이로드 시그니처 기반 분석 방법을 이용한다. SaaS 트래픽의 경우, SSL/TLS 기반으로 암호화된 패킷을 사용하고 있다.[1][2] IP / Port 기반 분석과 페이로드 시그니처 분석이 어려울 수 있으나, SaaS트래픽에 대한 연구가 많지 않고, 향후 연

※ 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구(No.2018-0-00539-001,블록체인의 트랜잭션 모니터링 및 분석 기술개발)와 2020년도 산업통상자원부 및 한국산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임 (No. 20008902, IT비용 최소화를 위한 5채널 탐지기술 기반 SaaS SW Management Platform(SMP) 개발)

구를 위하여 가장 접근하기 좋은 방법을 택하였다. 본 논문에서 제시하는 방법론의 경우 마이크로소프트 서비스를 사용하는데 웹을 사용한다는 것과 로그인을 통해 서비스를 사용하고 있다는 전제하에 진행되었다. 트래픽 수집은 웹을 통해 접속하여 마이크로소프트 Office 365 서비스 중 파워포인트, 워드, 엑셀, 3가지의 서비스를 이용하면서 수집하였다.

마이크로소프트 Office 365 서비스 중에서도 일상생활 및 기업에서 가장 많이 사용되고 있는 파워포인트, 엑셀, 워드에 대한 탐지를 중점적으로 분석한다. 사용자가 3가지의 서비스를 이용하게 되었을 때 나타나는 페이로드 분석 특징을 마이크로소프트, 아마카이 서버의 두 가지로 나누어 볼 수 있다. 서비스를 사용할 때, 마이크로소프트와 통신하는 서버에서 발생하는 고정적인 IP / Port를 도출 할 수 있었다. 해당 IP에서 발생하는 페이로드 분석 결과, 페이로드에서도 3가지 서비스에 대한 페이로드 시그니처를 정의할 수 있었다. 하지만, 여러 가지 상황에서 트래픽을 수집해 보았을 때, 마이크로소프트 서버에서 발생하는 트래픽만 가지고는 서비스를 완벽하게 탐지할 수 없었다. 따라서 Office 365의 클라우드 서비스의 컨텐츠 전송을 도와주는 아카마이 서버에서 발생하는 트래픽에 대한 분석을 진행하였다. 아카마이 서버에서는 IP대역이 항상 다양하여 고정된 IP를 찾을 수 없었다. 표1은 마이크로소프트 서비스를 사용할 때 발생하는 아카마이 서버의 IP 대역이다. 트래픽 분석 결과, 아카마이 서버의 대역은 다르지만, 마이크로소프트 서비스의 페이로드 정보를 가지고 있는 것을 확인하였다.

표 1. 아카마이 서버 IP 대역

Akamai 대역	23.72.0.0 – 23.79.255.255
	104.64.0.0 – 104.127.255.255

아카마이 대역에서 발생하는 트래픽의 페이로드를 분석한 결과, 우리가 탐지해야하는 3가지 서비스에 대한 페이로드 시그니처를 찾을 수 있었다. 마이크로소프트 서버 IP와 아카마이 서버에서 발생하는 페이로드 정보를 통해 사용자가 Office 365를 웹으로 이용할 때, 파워포인트, 엑셀, 워드에 해당하는 3가지 서비스에 대해 탐지가 가능하였다. 표2는 3가지 서비스에 대한 페이로드 시그니처 정보이다.

표 2. 페이로드 시그니처

Type	Destination IP	Payload
Akamai	다양한 IP 대역	c1-powerpoint-15.cdn.office.net
		c1-word-edit-15.cdn.office.net
		c1-excel-15.cdn.office.net
Microsoft	13.107.6.171	powerpoint.officeapps.live.com
		word-edit.officeapps.live.com
		excel.officeapps.live.com

그림 1은 정의된 IP / Port와 페이로드 시그니처 정보를 바탕으로 IP / Port 매칭과 페이로드 매칭을 통해 사용자가 이용한 서비스의 이름을 도출하는 알고리즘이다.

입력으로 전체 플로우가 들어오며, 플로우에서 정의된 마이크로소프트 IP / Port 정보를 매칭한다. 매칭 후 해당 플로우의 페이로드 정보를 확인하여 서비스 이름을 도출한다. 같은 방법으로 아카마이 서버의 대역을 확인하고 페이로드 정보를 통해 서비스 이름을 도출한다.

알고리즘 구현 후 검증을 하기 위해 여러 가지 실험을 진행하였다. 개인

Algorithm : Service Information Analysis
Input : Entire Flow
Output :Service Name
[Notation] F : Target Flow / SRC_F Source IP of the Flow / DST_F Destination IP of the Flow / PT_F: Port Number of the Flow / PL_P : Payload of the Packet / MS_SIG_F : Predefined Payload Signature (MS) / AK_SIG_F : Predefined Payload Signature (Akamai)

```
// Service Information Analysis
        for i=1 to Numbers of Flow in Entire Flows
            if DST_F == "13.107.6.171" && PT_F == 443
                for j=1 to 9 of Packet in Flows
                    if PL_P == MS_SIG_F
                        Set the Service Name as PL_P (User IP : SRC_F)
                        return Service Name
                    else
                        continue
            if the target flow belongs to Akamai
                for j=1 to 9 of Packet in Flows
                    if PL_P == AK_SIG_F
                        Set the Service Name as PL_P (User IP : SRC_F)
                        return Service Name
                    else
                        continue
```

PC, 학내망, 노트북 등, 여러 가지 환경에서도 사용자가 사용한 서비스 이름을 도출하는지 확인하였다. 개인 PC나 노트북의 경우 직접 웹 서비스를 사용하고 트래픽을 수집한 후 검증하기 때문에 도출된 결과가 정확하다는 것을 확인 할 수 있었다. 학내망의 경우, 파워포인트나 워드, 엑셀 작업들을 많이 하게 되는 환경이 갖춰져 있다. 하지만 다양한 호스트 IP에서 발생하는 여러 트래픽들이 수집되기 때문에 각각의 사용자들이 서비스를 사용했는지 확실하게 확인 할 방법은 없다. 하지만, 학교 내 다수의 PC를 사용하여 Office 365 서비스를 사용한 후 호스트 IP와 서비스 이름을 확인하여 정확하게 서비스 이름을 도출하는 것을 확인하였다.

Ⅲ. 결론 및 향후 연구

클라우드 서비스가 사용이 되면서, 대부분의 기업들이 적은 비용을 통해 SaaS 어플리케이션을 사용하게 되었다. 하지만 기업에서 관리하지 못하는 승인되지 않은 클라우드 서비스가 발생하게 되고, 이는 곳 기업의 효율적인 자산 운용에 영향을 미치게 된다. 이러한 Shadow IT 문제를 해결하기 위해 SaaS 모니터링 기술이 필요하다.

본 논문은 SaaS 서비스 중에서도 마이크로소프트 Office 365 서비스를 대상으로 트래픽 분석을 진행하였다. 사용자가 Office 365 서비스를 사용하였을 때 어떤 서비스를 사용하였는지 탐지할 수 있는 방법론에 대하여 제시하였다. 특히, 우리 주변에서 가장 많이 사용되고 있는 파워포인트, 엑셀, 워드 서비스에 대한 분석을 중점적으로 두었다. 제시한 방법론을 바탕으로 다양한 환경에서 실험 및 검증을 진행하였으며, Office 365 사용 서비스에 대한 이름을 도출 할 수 있었다.

향후 연구로는 사용자가 SaaS 서비스를 이용할 때 발생하는 다양한 행동들을 분석할 예정이다. 로그인이나 로그아웃 과정 등이 해당 될 수 있다. 또한, 마이크로소프트 뿐만 아니라 구글의 G-suite, Adobe 등 다양한 소프트웨어들에 대한 분석을 진행 할 예정이다.

참 고 문 헌

[1] 김성민, 박준상, 윤성호, 김종현, 최선호, 김명섭, "SSL/TLS 기반 암호화 트래픽의 서비스 식별 방법", 통신학회 논문지 Vol 40 No.11, Nov.2015, pp.2160-2168.

[2] 김성민, 구영훈, 김명섭, "Session ID - Server IP 캐싱 기반의 SSL/TLS 암호화 트래픽의 서비스 식별 방법", 2015년도 한국통신학회 하계종합학술발표회, 라마다호텔, 제주도, Jun. 23-25, 2015.