HTTP 트래픽의 User-Agent를 활용한 모바일 트래픽 분류

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

Classifying mobile traffic using User-Agent for HTTP traffic

Choi Jeong Woo, Park Jee-tae, Lee Min-Seong, Kim Myung-Sup
Korea Univ.

요 약

최근 대부분의 사람들이 스마트폰을 사용함에 따라 어플을 만들거나 웹 페이지를 만들 때 모바일 환경을 고려해서 개발이 되어야 할 필요성이 증가했다. 특히 하이브리드 및 웹 앱을 만들다보면 Android 와 iOS를 구분해야하는 경우도 생긴다. 그리고 공유기와 같은 NAT 를 사용한 트래픽의 경우에는 같은 IP 라고 하더라도 사용한 시각에 따라서 여러 가지 OS가 도출되는 상황이 발생한다. 따라서 본 논문에서는 스마트폰에서 발생한 트래픽을 기존 PC와 노트북에서 발생한 트래픽과 분류한다. 그 다음 OS 정보들을 바탕으로 해서 NAT 인지 아닌지를 구분한다. 이러한 내용들을 HTTP 트래픽의 User-Agent항목을 활용해서 진행하는 방법을 제안한다.

Ⅰ. 서 론

최근 스마트 폰의 사용이 당연시 되면서 어플을 만들거나 웹 페이지를 만들 때 모바일 환경을 고려해서 개발이 되어야 할 필요성이 증가하고 있다. 특히 하이브리드 및 웹 앱을 만들다보면 Android 와 iOS 를 구분해야 하는 경우가 발생한다. 그리고 공유기와 같은 NAT 를 사용한 트래픽의 경우에는 같은 IP 라고 하더라도 사용한 시각에 따라서 여러 가지 OS가 도출되는 상황이 발생한다. 따라서 본 논문에서는 이러한 경우를 고려해서 NAT 인지 아닌지를 구분하는 것 또한 진행한다. 스마트폰에서 발생한 트래픽과 PC 및 노트북에서 발생한 트래픽을 분류하고 스마트폰에서 발생한 트래픽의 경우 OS 분류한다. 그 다음 분류된 OS 정보들을 바탕으로 해서 NAT인지 아닌지 판별하는 내용까지 진행하는 방법을 제안하고자 한다. 스마트폰 트래픽이 가지는 특성에 관한 연구 또는 스마트폰 트래픽을 기존의 트래픽들과 비교, 분석하는 연구들이 간간히 이루어지고 있다 [1],[2]. TCP SYN 패킷의 헤더 정보들 중에 Window Size필드, Option필드 등을 바탕으로 해서 OS 시그니처를 생성하여 OS 를 구분하는 방법도 있고 [2], HTTP 트래픽의 User-Agent 항목을 활용하여 자동으로 OS 를 구분하는 방법도 있다[1]. 하지만 TCP SYN 패킷의 헤더 정보들 중 Window Size 필드, Option 필드 등을 활용하는 방법의 경우 시간이 지남에 따라 iOS, Android, Windows 모두 '65535'라는 동일한 크기의 Window Size 를 갖는 상황으로 변했다. 따라서 본 논문은 HTTP 트래픽의 User-Agent 항목을 사용하여 스마트폰에서 발생한 트래픽과 기존의 PC 및 노트북에서 발생한 트래픽을 구분하고 분류된 OS 정보들을 바탕으로 해서 NAT 인지 아닌지를 구분하는 방법을 제안한다. 본 논문의 구성은 본론에서는 연구에 사용된 학내망 트래픽에 대한 언급을 하고 먼저 스마트폰에서 발생한 트래픽을 기존의 PC 그리고 노트북에서 발생한 트래픽과 구분하고 모바일 기기의 경우 각각의 OS를 구분한다. 그 다음 분류된 OS 정보들을 바탕으로 해서 NAT 인지 아닌지를 구분하는 방법에 대한 언급을 한다. 그리고 결론에서는 해당 연구를 정리하고 향후 연구에 진행할 내용에 대한 언급을 하는 순서로 진행된다.

Ⅱ. 관련 연구

[1]은 해당 단말의 OS 정보를 판별하기 위한 방법으로 수동형 방법과 능동형 방법이 있고, 수동형 방법이 탐지속도가 빨라 실시간 네트워크 트래픽 분석에 적합하지만 시그니처 기반으로 동작하기 때문에 능동형 방법에 비해 분석률과 정확성이 낮은 문제를 가지고 있다. 이러한 문제점을 해결하기 위해 해당 논문에서는 HTTP 트래픽의 User-Agent 항목에 표기되어 있는 운영체제 정보를 기반으로 운영체제 판별을 위한 헤더 시그니처를 자동으로 추출하는 방법을 제안한다. 정확한 시그니처를 생성하기 위해 패킷을 수집 후 3-way handshake 를 가지는 플로우를 대상으로 시그니처를 추출했다. [2]는 OS 시그니처 추출 방법을 제안한다. TCP SYN 패킷의 헤더 정보를 활용한다. TCP SYN 패킷은 처음 TCP 통신을 시작할 때 전달되는데 헤더의 몇몇 필드들은 OS에서 정의하는 임의의 초기 값들로 채워지게 된다. 주목할 점은 서로 다른 OS 들이 자체의 기준에 따라 이러한 TCP SYN 패킷을 생성할 때 서로 다른 초기 값들을 가지고 있다는 것이다. 따라서 이러한 필드 값들을 활용하여 서로 다른 OS 들을 분류하기 위한 OS 시그니처를 추출하는 것도 가능해진다.

Ⅲ. 본론

본 장에서는 학내망 트래픽 수집과 스마트폰에서 발생한 트래픽 구분 방법 그리고 NAT 장비 판별 방법에 대해 언급한다.

※ 이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2018R1D1A1B07045742)과 2020년도 산업통상자원부 및 한국산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임 (No. 20008902, IT비용 최소화를 위한 5채널 탐지기술 기반 SaaS SW Management Platform(SMP) 개발)

A. 학내망 트래픽 수집

본 절에서는 실험에 사용된 학내망 트래픽에 대해 언급한다. 데이터는 실시간으로 Android 를 사용하는 스마트폰과 iOS 를 사용하는 스마트폰 그리고 iPad 를 사용한 각각의 트래픽을 수집하여 사용하였다.

B. 모바일 트래픽 분류 방법

본 절에서는 모바일 트래픽 분류 방법에 대해 언급한다. 스마트폰에서 발생한 트래픽과 PC 나 노트북에서 발생한 트래픽을 구분하기 위해서 HTTP 트래픽의 User-Agent 항목을 활용해서 OS 정보를 도출하여 모바일 트래픽을 OS 별로 분류한다. HTTP 헤더는 클라이언트와 서버가 요청 또는 응답으로 부가적인 정보를 전송할 수 있도록 해준다. HTTP 헤더는 대소문자를 구분하지 않는 이름과 콜론 ':' 다음에 오는 값 (줄 바꿈 없이)으로 이루어져 있다. 값 앞에 붙은 빈 문자열은 무시된다. [1]여기서 User-Agent 는 HTTP, SIP 에서 서버/클라이언트 간의 통신에서 클라이언트의 소프트웨어를 식별하기 위해 이용된다. 이것을 통해서 서버는 클라이언트에게 적합한 서비스를 제공할 수 있다. 렌더링 엔진, 브라우저가 실행중인 운영체제, 브라우저의 플랫폼 등이 User-Agent 에 나타난다. 예시로 사용한 User-Agent 정보는 다음과 같다. 기종은 아이폰으로 진행했으며 브라우저는 크롬을 이용한 내용이다.

내용

- Mozilla/5.0 (iPhone; CPU iPhone OS 13_3 like Mac OS X) AppleWebKit/605.1.15(KHTML, like Gecko) CriOS/80.0.3987.95 Mobile/15E148 Safari/604.1

(그림 1) 아이폰으로 사용한 User-Agent 정보

이러한 내용에서 OS 별로 특징적인 문자열을 특정해서 그것을 시그니처로 사용하여 OS 를 구분한다. MAC OS 의 경우에는 Mac OS X/Macintosh 가 특정되고, iPad OS 의 경우에는 Mac OS X/Macintosh/IPad, iOS 의 경우에는 Mac OS X/iPhone 그리고 Android 의 경우에는 Linux/Android 가 특정된다. Mac OS, iPad OS 그리고 iOS 의 경우에 Mac OS X 가 공통으로 나와서 Apple 제품으로 이용했다는 것을 알 수는 있지만 그것이 iPhone 인지 iPad 혹은 Mac Book 인지는 구분할 수는 없다. Apple 제품들을 구분하려면 또 다른 시그니처인 Macintosh/IPhone 을 이용해서 구분한다. 하지만 여기서 MAC OS 와 ipad OS 는 Macintosh 라는 시그니처가 중복돼서 나온다. 이러한 경우에서는 iPad OS 에서만 식별되는 시그니처인 Ipad 라는 시그니처를 이용해서 구분한다. 다음 표 2 는 각각의 OS별 식별되는 시그니처들을 정리해놓은 것이다.

OS	Signature1	Signature2	Signature3
MAC OS	Mac OS X	Macintosh	
iPad OS	Mac OS X	Macintosh	Ipad
IOS	Mac OS X	IPhone	
Android	Linux	Android	

표 1 User-Agent 의 OS Signature

C. NAT 장비 판별 방법

본 절에서는 NAT 를 사용한 트래픽을 판별하는 방법에 대해 언급한다. 우선 NAT 는 부족해지는 IPv4 (Internet Protocol version 4) 주소부족 문제를 해결하기 위해 단기적인 해결방안으로 제안된 방식으로 RFC 1918 의 사설주소 영역을 이용하여 공인 IP 주소로 접근할 수 있도록 해주는 기술이다. 일반적으로 NAT는 홈 네트워크 또는 사용자 네트워크에 설치되어 사설 주소들을 공인 IP 주소의 송신자 또는 수신자로 변환하여 준다 [4]. NAT 장비 판별 방법은 HTTP 트래픽의 User-Agent 항목을 활용하여 NAT 장비를 판별하는 것이다. 위의 절에서 구분된 OS 정보들을 바탕으로 누적된 데이터에서 하나의 IP에서 여러 OS 정보가 탐지된다면 확인된 OS 의 수만큼의 호스트들이 NAT 내부에 연결이 되어있다는 것을 의미한다. 즉 한 개의 IP에서 발생되지만 HTTP 트래픽의 User-Agent 정보에서 서로 다른 OS 의 정보가 탐지된다면 그 탐지된 OS 의 수만큼의 호스트들이 NAT 내부에 연결되어있다는 것으로 판단한다.

Ⅳ. 결론

본 논문에서는 HTTP 트래픽의 User-Agent 항목을 활용해서 스마트폰의 OS 별 시그니처를 활용해서 스마트폰에서 발생한 트래픽을 구분하고 분류된 OS 정보를 활용해서 NAT 인지 아닌지를 구분하는 방법에 대해서 제시하였다. 각기 다른 OS 를 사용하는 모바일 기기에서는 HTTP 트래픽 의 User-Agent 항목의 내용 중에서 일부 문자열이 다르게 식별된다. 그러한 문자열들을 특정해서 시그니처로 설정하여 서로 다른 OS 를 사용하는 모바일 기기들을 구분하였다. 그리고 누적된 데이터에서 한 개의 IP에서 발생되지만 여러 OS의 정보가 탐지된다면 이러한 경우 NAT 를 사용했고 탐지된 OS 정보의 수만큼의 호스트가 NAT 내부에 연결되어있다는 것으로 판단하였다. 이러한 방법을 사용한다면 어플을 개발하거나 웹 페이지를 만들 때 특히 하이브리드 및 웹 앱을 만들 때처럼 모바일과 기존의 PC 및 노트북을 구분해야하거나 Android와 iOS를 구분해야하는 상황에서 유용하게 사용될 것으로 기대된다. 그리고 NAT 에 대한 구분까지 진행됨으로 인해서 한 가지 IP에서 여러 가지 OS 가 탐지되는 상황에서의 혼란도 방지 할 수 있을 것으로 기대된다.

향후 연구로는 최근 많이 사용되고 있는 Microsoft Office 365 와 같은 SaaS (Service as a Service) 서비스의 관리에 대한 내용을 진행해보려 한다. 우선 SaaS 서비스는 공급자나 서비스 제공자가 서버상에 애플리케이션을 호스팅하고, 고객은 웹 브로우져 등 온라인을 통해 사용한만큼 비용을 지불하고, 소프트웨어를 서비스로 이용할 수 있도록 하는 소프트웨어 배포 모델을 가리킨다[3]. 특히 기업에서 SaaS 서비스를 사용하여 기업의 초기 비용 절감과 필요에 따른 유연한 사용으로 지속적으로 많이 사용되고 있다. 하지만 기업에서는 서비스를 가져와서 사용하기에 관리가 어렵다는 단점이 있다. 따라서 기업에서도 사용한 서비스를 관리할 수 있는 새로운 방식이 필요하다. 본 논문에서 제시한 모바일 기기와 기존의 PC, 노트북을 구분하는 방법을 활용해서 모바일로 사용한 Microsoft Office 365의 트래픽을 관리하는 것에 대한 연구를 진행해볼 예정이다. 예를 들면 PoewrPoint를 사용했는지 Word를 사용했는지와 같은 Feature를 구분하는 내용이나 Login/Logout과 같은 내용을 통해 사용자가 자주 사용하지 않는 Feature를 확인하여 지속적으로 Microsoft Office 365 의 사용을 관리할 수 있도록 도움을 줄 수 있는 연구를 진행해 볼 예정이다.

참 고 문 헌

[1] 허민, 이현신, 김명섭, "HTTP 트래픽을 이용한 운영체제 시그니처 자동 생성에 관한 연구", KNOM Conference 2011, Pohang, korea, April 21-22, 2011

[2] 정태열, Le Quoc Do, 홍원기, "OS 시그니처 기반의 스마트폰 트래픽 분류 방법에 관한 연구", 한국통신학회 동계종합학술발표회, 2012

[3] 김형환 외 12명, "SaaS 기술 개발 동향", 전자통신 동향분석, 제24권, 제4호, 2008.

[4] 선종현, "NAT 환경에서의 IP ID를 이용한 DDoS 공격 호스트 탐지 알고리즘에 관한 연구", 학위 논문, 2010