

응용 레벨 모바일 트래픽 분류를 위한 모바일 트래픽 수집 구조 연구

최영락^{0,1}, 리건², Mahmoud Mahdi³, 박병철², 홍원기^{1,2}

포항공과대학교¹ 정보전자융합공학부,² 컴퓨터공학과

³ Computer Science and Networks, Higher School of Communications of Tunis SUP'COM

^{1,2} {dkby, gunine, fates, jwkhong}@postech.ac.kr, ³ mahmoud.mahdi@supcom.rnu.tn

요 약

스마트 폰, 스마트 태블릿과 같은 모바일 장치가 대중화됨에 따라, 다양한 모바일 앱으로부터 WiFi/3G/4G 등의 여러 네트워크를 통해 생성되는 모바일 트래픽 양이 급증하는 추세에 있다. 응용 레벨 모바일 트래픽 분류 작업은 분류 결과를 통해 서비스 제공자 입장에서 급증하는 모바일 트래픽을 대비하기 위한 전략 수립의 근거가 될 뿐만 아니라, 어플리케이션 개발자 및 사용자 입장에서 모바일 트래픽 특성에 알맞게 효율적으로 모바일 장치를 활용할 수 있다는 점에서 매우 중요하다. 본 연구에서는 가상 사설망 구성 및 모바일 TMA (Traffic Measurement Agents)를 활용하여 다양한 네트워크 및 모바일 플랫폼을 지원 가능한 모바일 트래픽 수집 구조를 제안한다. 제안하는 구조를 통해 선행 연구에서 모바일 단말의 제약된 성능 및 NAT (Network Address Translator) 환경에서도 손실없는 트래픽 수집 및 모바일 단말 별 트래픽 수집이 가능해진다. 제안한 수집 구조가 실제 적용 가능한지 확인하기 위해 VPN 서버를 구성하고 Android 및 iOS 장비에 대해 WiFi/3G 트래픽이 올바르게 수집되는지 확인하였으며, 제안하는 모바일 TMA 에서 사용할 기능이 구현 가능한지 확인하였다.

1. 서론

최근 스마트 혁명이라 일컫는 스마트 폰, 스마트 태블릿과 같은 모바일 장치의 급속한 확산을 통해 모바일 트래픽 양이 급증하였으며, 향후 모바일 트래픽은 더욱 증가량 폭이 늘어날 것으로 예상된다. 서비스 제공자 입장에서 모바일 트래픽 분석은 모바일 네트워크의 상태를 정확하게 파악하여 고품질의 네트워크 서비스 제공, 과도한 네트워크 자원을 점유하는 모바일 트래픽 관리 및 모바일 네트워크를 통한 악의적인 공격을 방어하는 데 있어 매우 중요하다. 또한, 분석된 모바일 트래픽 패턴 정보를 기반으로 모바일 앱 개발자들은 모바일 네트워크를 보다 효율적으로 활용하는 앱 개발이 가능해지고, 모바일 단말 사용자들에게 여러 모바일 네트워크를 트래픽 패턴에 따라 선택 가능하도록 하는 지표를 제공하는 것이 가능해진다.

모바일 트래픽을 응용 레벨에서 분석하기 위해서는 분석하고자 하는 모바일 어플리케이션을 분류할 수 있는 정보를 생성하는 과정이 필요하다. 선행 연구에서는 [1] 모바일 TMA (Traffic Measurement Agents)를 활용하여 각 모바일 어플리케이션 별로 해당 어플리케이션에 대한 트래픽을 모아 놓은 Ground-Truth 데이터를 수집하고자 하였다. 해당 연구에서는 모바일 장치 성능의 한계로 인하여 모바

일 장치에서 100% 수집하지 못하던 트래픽을 인터넷 Junction 에서 수집한 트래픽과 비교하는 작업을 통해 해당 모바일 어플리케이션에 대한 최대한 많은 트래픽을 수집하여 분석하였다. 그러나 인터넷 Junction 에서 수집된 트래픽은 다른 장치들에서 발생된 트래픽과 섞여있어 해당 모바일 장치의 100% 트래픽을 수집하였다고 보장하기 어렵다는 단점이 있다. 또한, 선행 연구에서 수행하였던 방법은 WiFi에만 적용 가능하고 3G/4G 등 타 네트워크에는 적용이 불가능하다는 단점이 있었다.

본 연구에서는 선행 연구에서의 단점을 극복하기 위해 가상 사설망 구성 및 모바일 TMA 를 활용한 모바일 트래픽 수집 구조를 제안하고자 한다. 제안하는 수집 구조는 가상 사설망을 활용하여 모바일 단말에서 생성된 모든 트래픽을 가상 사설망 서버로 우회 (redirect)시켜 해당 트래픽을 수집하는 방법을 활용하여 손실없이 트래픽 수집이 가능하다. 또한, 모바일 TMA 를 활용하여 모바일 장치에서 실행되는 모바일 앱 정보를 모니터링하여 해당 모바일 장치에서 실행 중인 모바일 앱들이 생성한 트래픽을 추출하고자 하였다. 제안하는 시스템 구조를 통해 모바일 앱이 여러 네트워크를 통해 발생시킨 모든 트래픽들을 모바일 단말 별로 수집이 가능하여 모바일 트래픽 분류 정보를 생성할 때, 그리고 모바일 트래픽 분류 정확도를 검증하기 위한

Ground Truth 데이터 활용에 도움을 줄 것으로 기대된다.

본 논문의 구성은 다음과 같다. 2 장에서는 모바일 트래픽 분류 및 분석에 대한 관련 연구에 대해 기술한다. 3 장에서는 제안하는 모바일 트래픽 수집 구조에 대해 설명한다. 4 장에서는 제안한 수집 구조에 대한 환경 구성 및 트래픽 수집, 기능 확인을 통해 실제 적용 가능한지 살펴본 결과를 설명한다. 마지막으로 5 장에서는 요약 및 향후 연구에 대해 기술한다.

2. 관련 연구

모바일 트래픽 양이 증가 추세에 있으며, 앞으로도 계속하여 증가할 것이라고 예측되고 있다 [2]. 많은 기존 연구들에서는 해당 모바일 트래픽에 대해 측정하고 분석하는 과정을 통하여 모바일 트래픽이 어떤 특성을 보이는지를 제시하고 있다 [3, 4, 5]. 해당 연구 결과들을 살펴보면 HTTP 트래픽이 모바일 트래픽의 대부분에 해당하였으며, 웹 브라우징 및 비디오 재생과 관련된 트래픽이 상위 비율에 해당되었다. 그러나 해당 연구들에서 사용한 Port 매칭 및 HTTP User-agent 문자열 매칭 방식은 기존 트래픽 분류 방법과 비교해 볼 때, 정확한 트래픽 분류 결과를 보장하지 못한다고 할 수 있다. [6]에서는 수집된 트래픽을 기반으로 스마트폰 사용 패턴을 측정하는 데 목표를 두어 다양한 모바일 플랫폼, 어플리케이션에 따른 특성, 그리고 지리적인 위치에 따른 스마트폰 어플리케이션 사용 패턴 측정 결과를 설명하였다.

선행 연구 [1]에서 제안하였던 시스템 구조는 모바일 트래픽 분류를 위한 정보를 생성하기 위한 구조에 초점을 두고 있다. 해당 연구에서는 모바일 단말에서 수집한 패킷에 대해 어떤 모바일 어플리케이션으로부터 발생되었는지에 대한 정보를 수집하고, 인터넷 Junction 에서 수집된 트래픽과 비교하는 작업을 통해 모바일 단말의 성능 제약으로 인한 패킷 손실을 보완하여 모바일 트래픽을 모바일 어플리케이션 별로 수집하였다. 그러나 실험에 사용된 모바일 TMA 프로그램은 모바일 단말에서 제약된 성능에 비해 패킷 수집 및 어플리케이션 비교와 같은 다소 무거운 작업을 실행하여 패킷 손실뿐 아니라 모바일 단말 사용에 지장을 줄 수 있다. 또한 제안하였던 시스템 구조는 WiFi 에만 적용 가능하여 3G/4G 등 타 모바일 네트워크에서 발생한 트래픽 정보 수집은 불가능하였다.

따라서 본 연구에서는 다양한 플랫폼으로 구성된 모바일 단말들이 여러 모바일 네트워크를 사용하여 발생하는 모든 응용 레벨 트래픽을 수집하여 모바일 트래픽 분류에 사용 가능한 Ground Truth 데이터를 생성하는 데 초점을 두어 모바일 트래픽 수집 시스템 구조를 제안한다. 제안하는 시스템 구조를 통해 단말별 각 모바일 앱에 대한 모바일 트래

픽을 수집하여, 이를 기반으로 모바일 트래픽 분류가 가능한 분류자를 생성한다면 [7], 보다 다양하고 자세한 모바일 트래픽 사용 패턴 분석이 가능할 것이다.

3. 제안 트래픽 수집 구조

3 장에서는 제안하는 트래픽 수집 구조를 설명한다. 3.1 절에서는 해당 수집 구조를 적용할 교내 네트워크 환경을 소개하고, 3.2 절에서는 가상 사설망에 따른 수집 구조에 대해 설명한다. 3.3 절에서는 수집 구조에서 사용되는 모바일 TMA 의 내부 구조를 설명한다.

3.1. 교내 네트워크 망 구성

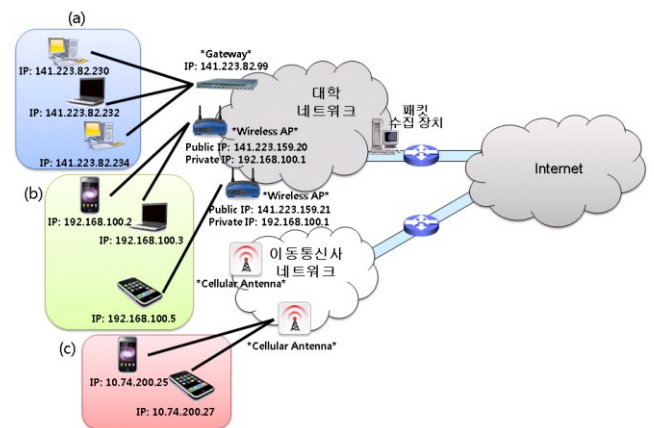


그림 1. 교내 인터넷 연결 구조 (IP 주소는 예시)

트래픽 수집은 사용되는 네트워크 망이 어떻게 구성되어 있는지에 따라 수집 위치 및 방법 등이 달라질 수 있다. 그림 1 은 교내 (POSTECH) 에서 인터넷을 사용하는 장치들이 접근 가능한 네트워크 종류 및 인터넷 연결 구조를 도식화한 것이다. 교내에서 인터넷을 사용하는 장치들은 네트워크 종류에 따라 크게 (a) 유선 네트워크 및 공인 (Public) IP 주소를 사용하는 장치들, (b) 무선 AP 에 접속하여 사설 (Private) IP 주소를 사용하는 장치들, (c) 이동통신사 네트워크를 직접 접근하는 장치들 (주로 모바일 장비)의 3 가지로 구분된다. 현재 교내에서 트래픽 분류 실험을 위해 사용되는 패킷 수집 장치는 대학 네트워크와 인터넷 사이에 해당하는 Junction 에 설치되어 있어 무선 AP 를 통해 발생하는 트래픽은 NAT [8]설정으로 인해 사설 IP 주소 (모바일 단말에서 사용하는 IP 주소)가 아닌 공용 IP 주소 (무선 AP 의 IP 주소)를 기반으로 트래픽 수집만 가능하였다. 따라서 특정 모바일 장치에서 생성된 트래픽을 수집하기 위해서는 해당 트래픽을 찾기 위한 부가적인 방법을 필요로 하였고, 선행 연구에서 제안된 방법으로는 100%의 트래픽을 찾아내지 못하는 어려움이 있었다.

3.2. 가상 사설망을 활용한 단말별 트래픽 수집

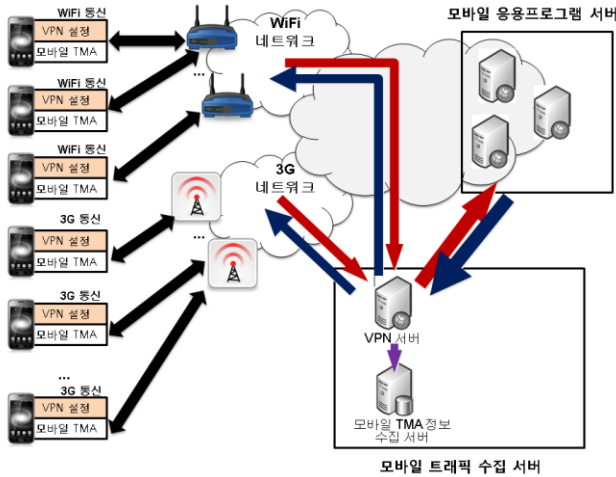


그림 2. 모바일 트래픽 수집을 위한 제안 구조

그림 2 는 제안하는 모바일 트래픽 수집을 위한 구조를 나타낸다. 제안하는 구조에서는 모바일 단말에서 실행되는 응용프로그램 및 소켓 정보만을 수집하는 모바일 TMA 를 설치하고, 가상 사설망에 해당하는 VPN (Virtual Private Network) 을 구성하고, VPN 서버에서 모바일 트래픽을 단말별로 수집한다. 모바일 단말에 VPN 을 설정하였을 경우 모바일 단말이 네트워크를 사용하는 방식은 모바일 단말에서 해당 네트워크를 직접적으로 접근하는 것이 아니라, 모바일 단말의 모든 트래픽은 VPN 서버를 통해 우회하여 VPN 서버에서 해당 네트워크에 접근하게 된다. 이 때, 모바일 단말 사용자들이 어떤 모바일 네트워크 (예: WiFi, 3G, 4G) 를 사용하더라도 송수신되는 모바일 트래픽이 VPN 서버를 거치므로, VPN 서버에서 모바일 트래픽을 수집하여 모바일 트래픽을 손실없이 수집이 가능해진다. 또한, VPN 을 통해 각 단말당 가상 채널 구축이 가능해 지므로, VPN 서버에서는 각 가상 채널에 대해 트래픽 수집을 수행함으로써 모바일 단말별 트래픽 수집이 가능해진다.

선행 연구에서는 인터넷 Junction 상에서 수집된 모든 트래픽 가운데 특정 모바일 단말로부터 발생한 트래픽을 찾고자 하였기에, 해당 모바일 트래픽을 찾는 데 많은 시간이 걸리고, 100%의 트래픽 추출 정확도를 보증하지 못하였다. 하지만, 제안하는 수집 구조는 모바일 단말별로 가상 채널을 할당하므로 단말별로 모바일 트래픽 수집이 가능해지기에 단말 별로 모바일 트래픽을 추출하는 과정을 수행하지 않아도 된다. 이는 NAT 로 구성된 네트워크 환경에서도 모바일 단말별 트래픽 수집이 가능해진다는 이점을 제공한다.

또한, 선행 연구에서 트래픽을 수집한 위치에 해당하는 인터넷 Junction 은 WiFi 네트워크와 모바일 응용프로그램 서버 사이에만 위치하고 있어, 3G

및 4G 와 같은 타 모바일 네트워크를 통해 발생하는 트래픽 수집은 불가능하였다. 3G 및 4G 네트워크의 경우 패킷 송수신 프로토콜이 WiFi 등과 같은 IP 네트워크와 차이가 있어 이에 대한 직접적인 패킷 수집은 불가능하다. 그러나 제안하는 수집 구조는 가상 사설망을 통해 다른 모바일 네트워크를 사용하는 모바일 트래픽도 VPN 서버를 거치도록 설정하였기 때문에 해당 트래픽에 대한 수집이 가능해진다.

3.3. 모바일 TMA 를 활용한 앱별 트래픽 수집

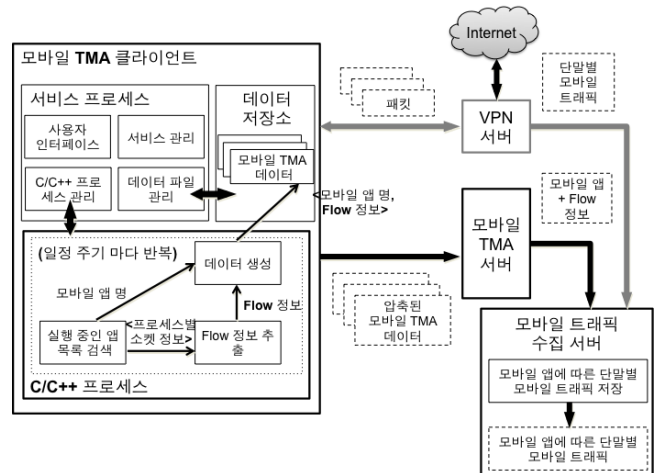


그림 3. 모바일 TMA 기능 및 앱 별 모바일 트래픽 수집 과정

그림 3 은 제안하는 수집 구조에서 사용되는 모바일 TMA 구조 및 기능과 함께 모바일 TMA 를 통해 전달된 데이터 및 VPN 서버에서 수집한 단말별 트래픽을 기반으로 모바일 앱에 따른 단말별 모바일 트래픽을 수집하는 과정을 도식화한 것이다. 제안하는 구조에서 사용되는 모바일 TMA 는 모바일 단말에서 트래픽을 수집하지 않고, 모바일 단말에서 실행되는 모바일 앱을 기준으로 생성되는 Flow 정보를 찾아 해당 정보를 모바일 TMA 서버에 주기적으로 전송한다. Flow 정보는 5-tuple 로 구성된 <source IP, source Port, destination IP, destination Port, Protocol type> 데이터를 담고 있으며, 모바일 앱 이름은 모바일 운영체제에서의 실행되는 모바일 앱에 대한 프로세스 명에 해당한다. 해당 과정은 상대적으로 짧은 주기마다 실행되어 실행되는 모든 모바일 앱에 대한 Flow 정보를 보관된 해당 정보는 모바일 TMA 서버에 주기적으로 전송되어 모바일 트래픽 응용 레벨 분류를 위한 데이터 활용을 가능하게 한다.

선행 연구에서는 인터넷 Junction 에서 모바일 트래픽을 수집하더라도 NAT 환경으로 인하여 수집된 트래픽이 어느 단말에서 발생하였는지를 알지 못하였다. 이를 해결하기 위해 모바일 단말에서 패킷 수집을 수행하였고, 패킷 수집 및 수집된 패킷에

대해 모바일 앱을 찾는 과정을 수행하여 모든 패킷을 수집하지 못하는 단점이 있었다. 그러나 제안하는 구조에서는 VPN 구성을 통해 단말별 모바일 트래픽 수집이 가능해져, 모바일 TMA 에서 패킷을 수집하지 않고, 실행 중인 모바일 앱 및 해당 모바일 앱의 Flow 정보만 추출하여 선행 연구에서보다 효과적으로 모바일 트래픽 수집이 가능해진다.

4. 트래픽 수집 환경 구성

4 장에서는 3 장에서 제안한 수집 구조를 실제 구성 가능한지 알아보기 위해 VPN 서버를 구성하여 WiFi 및 3G 트래픽이 올바르게 수집되는지를 확인한 결과 및 Android 및 iOS 장비에서 모바일 TMA 개발에 필요한 기능이 구현 가능한지를 살펴본 결과를 설명한다.

4.1. VPN 구성

VPN 을 구성하여 제안한 구조와 같이 모바일 트래픽 수집이 가능한지를 살펴보기 위해 교내 공용 IP 가 할당된 리눅스 서버에 VPN 서버를 구성하고, Android 및 iOS 운영체제가 동작 중인 모바일 장치에 VPN 을 설정한 후, WiFi 및 3G 트래픽을 발생시켜 VPN 서버에서 수집되는지 확인해 보았다. VPN 서버 구성을 위해 사용되는 대표적인 프로토콜로는 PPTP, L2TP, IPSec 등이 있다. 본 논문에서는 PPTP 프로토콜을 지원하는 poptop [9]를 활용하여 리눅스 서버에 VPN 서버를 구성하였다. VPN 구성 및 모바일 트래픽 수집에 사용된 환경은 다음과 같다.

- 리눅스 서버 OS: CentOS 6.2
- VPN 구성 프로그램: pptpd v1.3.4
- Android 모바일 장치: Galaxy S (Gingerbread)
- iOS 모바일 장치: iPhone 4 (iOS Ver. 5.0.3)
- 패킷 수집 프로그램: tcpdump v4.1

그림 4는 iPad 및 Galaxy S에서 VPN 설정을 구성하는 화면을 나타낸다. iOS 및 Android 모두 VPN 설정을 지원하여, 제안하는 구조와 같이 VPN 을 구성하고, VPN 서버를 통한 네트워크 통신이 가능하였다. 그림 5는 iPhone 및 Galaxy S 모두 VPN 설정을 구성하였을 때, VPN 서버가 설치된 리눅스에서 구성되는 가상 인터페이스를 조회한 결과에 해당한다. ppp0, ppp1 과 같이 별도의 VPN 채널이 구성되어 tcpdump [10]를 활용해 패킷 수집이 가능하였다. 환경 구성 후 WiFi 및 3G 트래픽을 생성하여 수집한 결과, 모두 VPN 서버에서 문제없이 수집됨을 확인하였다.

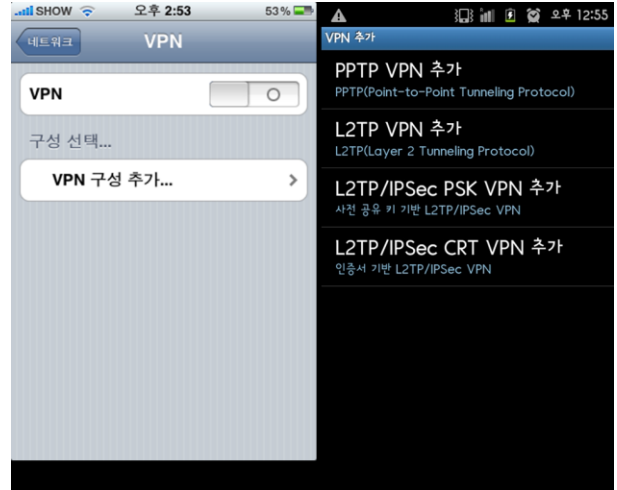


그림 4. 모바일 단말에서의 VPN 구성 화면

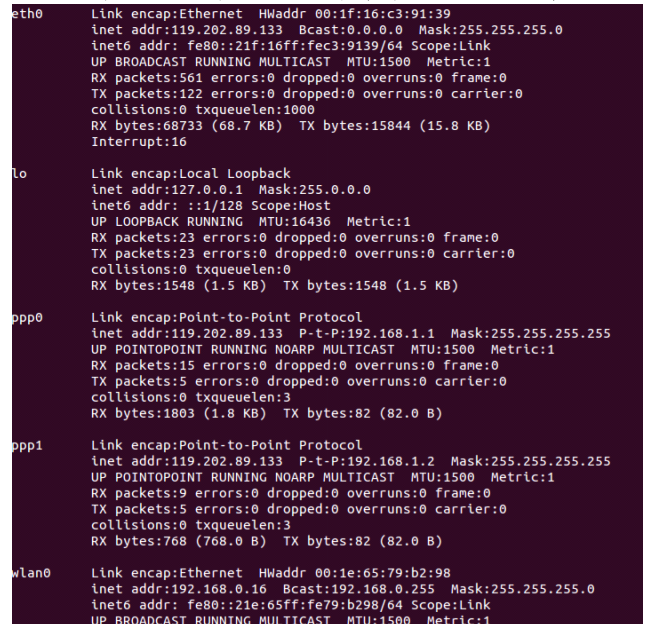


그림 5. 단말별 VPN 채널에 따른 가상 인터페이스 조회 결과

4.2. 모바일 TMA 기능 지원 확인

모바일 TMA 기능을 모바일 단말에서 실행하기 위해서는 각 모바일 운영체제에서 해당 기능들을 지원하는지 확인하는 과정이 필요하다. 제안하는 구조에서 모바일 TMA 에서 수행하는 핵심 기능은 크게 (1) 전체 모바일 앱 목록 검색 및 (2) 각 모바일 앱에 대한 Flow 정보 조회에 해당한다. Android 및 iOS 운영체제에서 위 2 가지 기능을 모두 지원하는지 확인해 보았다. 확인 결과, 두 운영체제 모두 해당 기능을 위해서는 최상위 (root) 권한을 필요로 하였으며, 따라서 Android 의 경우 루팅을, iOS 의 경우 탈옥 (Jailbreaking)이라는 과정을 필요로 하였다.

Android 운영체제는 Linux 커널을 기반으로 구현된 운영체제로, 해당 기능들은 Linux 에서 실행 중인 프로세스 목록 검색 및 각 실행 중인 프로세스에 대한 Flow 정보를 조회하는 기능과 거의 동일

하게 구현되어 있었다. Linux 와 마찬가지로 /proc 디렉토리에 실행 중인 모바일 프로세스에 대한 모든 id 목록을 볼 수 있었으며, 각 id 목록은 디렉토리로 이루어져 있었다. 각 id 에 대한 디렉토리 안에는 cmdline 파일을 조회하여 모바일 프로세스 명을 알 수 있었고, net/tcp 및 net/udp 파일을 조회하여 Flow 소켓 정보 조회가 가능하였다. Android NDK 를 활용하여 native C programming 을 하는 것으로 해당 기능 구현이 가능함을 조사 결과 알 수 있었다.

iOS 운영체제는 BSD 계열 커널 소스를 기반으로 구현된 운영체제로, Linux 및 Android 와 달리 /proc 디렉토리가 존재하지 않아 위에서 사용하였던 방법으로는 해당 기능 구현이 불가능하였다. 그러나 iOS 에서 컴파일 가능한 gcc 를 설치하여 C programming 을 할 경우, BSD 에서 지원하는 시스템 프로그래밍이 가능한 것으로 조사되었으며, 해당 방식을 통해 생성된 lsof 바이너리를 다운로드하여 실행하여 iOS 에서 실행 중인 모바일 앱 목록과 함께 해당 앱이 사용 중인 Flow 정보를 확인할 수 있었다. 따라서 모바일 TMA 에서 사용하고자 하는 두 기능 역시 iOS 에서 구현이 가능함을 확인하였다.

5. 결론

본 연구에서는 선행 연구에서 모바일 트래픽을 수집할 때의 어려움을 극복하기 위해 가상 사설망 구성 및 모바일 TMA 를 활용한 모바일 트래픽 수집 구조를 제안하고, 제안한 구조가 구현 가능한지를 확인하였다. 제안하는 구조를 통해 모바일 앱에 대한 단말별 모바일 트래픽 수집이 가능해졌을 뿐만 아니라, WiFi 및 3G/4G 네트워크를 사용할 때의 모바일 트래픽 수집 역시 가능해졌다. 또한, 단말별 트래픽을 추출하는 작업을 필요로 하지 않아 선행 연구에서 사용하였던 모바일 트래픽 수집 구조보다 더 빠르게 모바일 트래픽 분석에 사용될 데이터 생성이 가능하다는 장점을 얻을 수 있었다.

향후 연구로는 제안한 구조를 통해 WiFi 및 3G/4G 네트워크에서 발생된 모바일 트래픽을 수집 가능하도록 Android 및 iOS 모바일 TMA 를 개발하고, VPN 구성을 통해 모바일 트래픽에 대한 Ground Truth 데이터를 얻고자 한다. 해당 모바일 앱 및 네트워크 별 모바일 트래픽을 기반으로 모바일 네트워크 및 플랫폼에 따른 트래픽 차이 분석하고, 선행 연구인 자동화된 분류자를 생성하는 시스템과 연동하여 모바일 트래픽을 정확하게 분류 가능한 기술을 확보하고자 한다.

6. 참고 문헌

- [1] 최영락, 정재윤, 박병철, 홍원기, “응용 레벨 모바일 트래픽 모니터링 및 분석을 위한 시스템 연구”, KNOM Review Vol 14, No 2, Dec. 2011, pp. 10-21.
- [2] Cisco, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011-2016”, Feb. 14, 2012.
- [3] G. Maier, F. Schneider, and A. Feldmann, “A First Look at Mobile Hand-held Device Traffic”, Passive and Active Measurement, Zurich, Switzerland, Apr. 7-9, 2010, pp.161-170.
- [4] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, "A First Look at Traffic on Smartphones", Internet Measurement Conference (IMC), 2010, pp 281-287.
- [5] A. Gember, A. Anand, and A. Akella, “A Comparative Study of Handheld and Non-Handheld Traffic in Campus Wi-Fi Networks”, Passive and Active Measurement, Atlanta, USA, Mar. 20-22, 2011, pp.173-183.
- [6] Q. Xu, J. Erman, A. Gerber, Z. M. Mao, J. Pang, S. Venkataraman, "Identifying Diverse Usage Behaviors of Smartphone Apps", Internet Measurement Conference (IMC) 2011, Germany, Nov 2-4 2011, pp. 329-344.
- [7] Y. Choi, J. Y. Chung, B. Park, and J. W. Hong, "Automated Classifier Generation for Application-Level Mobile Traffic Identification," 13th IEEE/IFIP Network Operations and Management Symposium (NOMS 2012), Mini-conference, Maui, Hawaii, USA, April 16-20, 2012, pp. 1075-1081.
- [8] P. Srisuresh and M. Holdrege, “IP Network Address Translator (NAT) Terminology and Considerations”, RFC 2663, August 1999.
- [9] Poptop – The PPTP Server for Linux, available on <http://poptop.sourceforge.net/>.
- [10] Tcpdump & libpcap, <http://www.tcpdump.org/>.