KNOM Review '16-02 Vol.19 No.02
https://doi.org/10.22670/knom.2016.19.2.26

Sky-Scope : 스카이프 응용 트래픽 탐지 시스템

이 성 호•, 구 영 훈*, 이 민 회**, 박 지 태***, 김 명 섭#

Sky-Scope : Skype Application Traffic Detection System

Sung-Ho Lee•, Young-Hoon Goo*, Min-Hoe Lee**, Ji-Tae Park***, Min-Myung-Sup Kim#

요 약

오늘날 네트워크 환경이 증가하며 응용 및 서비스 별로 발생시키는 트래픽 패턴에 다양한 종류가 생성됨에 따라 트래픽의 응용 및 서비스 별로 분류할 수 있는 트래픽 분석 방법이 연구되고있다. 특히 Skype는 Microsoft사에서 서비스하는 P2P기반의 VoIP서비스로서 현재 국제적으로 가장 광범위하게 사용되고있다. 이러한 이유로 Skype 트래픽 탐지는 네트워크 망 관리 측면에 있어 그 중요성이 대두되고있다. 기존의 시그니쳐, 기계 학습 기반의 탐지 방법들이 갖고있었던 한계점 극복과 현재의 Skype트래픽 패턴을 보다 정확하게 분석하고 탐지하기 위해 본 논문은 패턴, 리스트, 시그니쳐 기반의 응용 탐지 방법을 조합한 종합적인 Skype트래픽 탐지 시스템을 제안한다. 제안된 시스템은 학내 네트워크망을 통해 수집한 다양한 Skype 트래픽에 적용하여 정확성과 탐지율을 검증하였다.

Key Words : Skype, P2P, Traffic Classification, Network Management, Analysis

ABSTRACT

Today, as the network environment changes, various types of traffic patterns for applications and services are generated. Several methods for applications' and services' traffic classification and analysis are being studied. However, there is yet no solution to the problem. In particular, Skype is among P2P-based VoIP services that is serviced by Microsoft and is currently of the most widely used application, globally. For this reason, the importance of Skype traffic detection is growing for effective network management. This paper presents a comprehensive Skype traffic detection system that combines patterns, list and signatures based applications detection methods to overcome the limitations of existing signatures and machine learning based detection methods for accurate detection and analysis of current Skype traffic patterns. The proposed system was deployed to the campus network and collected various Skype traffic in different environment to verify the its detection rate and accuracy.

Ⅰ. 서 론

최근 인터넷 사용자의 증가와 고속 네트워크의 보급으로 네트워크 트래픽이 급증하였다. 이것은 단순히 WWW, FTP, SMTP, DNS와 같은 전통적인 인터 넷 서비스뿐만 아니라 멀티미디어, P2P (peer-to-peer), P2P기반으로 게임, 스마트 폰 응용 등의 다양한 서비스의 증가로 인한 부분도 큰 몫을

※이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구(No.2015R1D1A3A01018057) 및 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 과학기술인문융합연구사업임(No.NRF-2016M3C1B6929228).
• First Author : Korea University Department of Computer and Information Science, gaek5@korea.ac.kr
Corresponding Author : Korea University Department of Computer and Information Science, tmskim@korea.ac.kr
* Korea University Department of Computer and Information Science, gyh0808@gmail.com
** Korea University Department of Computer and Information Science, minhoe1122@korea.ac.kr
*** Korea University Department of Computer and Information Science, pjj5846@korea.ac.kr
논문번호 : KNOM2016-02-04, Received September 23, 2016; Revised October 10, 2016; Accepted November 13, 2016

차지하기 때문이다. 이에 따른 트래픽이 급증함에 따라 효과적인 네트워크 관리를 위해 트래픽 모니터링 및 분석의 중요성이 커지고 있다.

효과적인 네트워크 관리를 위해 선행되어야 할 것은 해당 트래픽이 어떤 응용 또는 어떤 서비스에서 발생 되었는가를 판별 하는 것이다. 이미 많은 기존 논문에서 응용 레벨 트래픽을 분류에 대한 다양한 알고리즘을 제시하였지만 P2P기반의 Skype 응용에 대한 분석율은 다른 응용에 비해 상대적으로 낮았다. 또한 지속적인Skype 응용의 버전 향상과 함께 트래픽 패턴 역시 계속 변화하고 있기 때문에 현재의 트래픽에 적용하기에는 한계가 있었다. 따라서 이러한 한계점을 극복하기 위한 연구가 더 필요한 실정이다.

본 논문은 payload 시그니쳐 기반 분석, machine learning 기반 분석 등 기존의 방법론에 의존하지 않고 Skype응용의 트래픽 특성을 분석해 해당 응용 트래픽을 정확하게 탐지할 수 있는 시스템을 제안하는 것을 목적으로 한다. Skype 응용은 P2P 방식의 메신저로써 사용자간 채팅, 음성통화, 화상통화, 전화 교환망을 통한 일반 전화, 파일전송 등의 기능을 제공한다. 안정적인 통 화품질 서비스 제공과 일반 전화망에 비교해 저렴 한 가격은 오늘날 전 세계적으로 가장 많이 사용하 는 메신저로 만들었다. 하지만 엔터프라이즈 네트워 크 관리자 입장에서 보면 Skype의 트래픽들은 기본 적으로 암호화가 되어있고, 해당 응용 설치 시 동적 포트 번호가 할당되고, 일반적인 프로토콜을 사용하지 않아 Skype의 트래픽을 정확하게 탐지하는 것은 일반적인 분석 방법론 으로는 불가능하다. 하지만 Skype 응용 설치 때마다 달라지는 클라이언트 포트를 알아낼 수 있다면 대부분의 트래픽 을 쉽게 분류 할 수 있으며 각 호스트의 응용 사용 시간 측정 및 제어 등 다양한 방면에서 활용이 가능하다.

본 논문에서는 Skype 응용 설치과정에서 동적으로 할당되는 클라이언트의 Skype 포트번호를 알아내기 위해 Skype 응용의 로그인 과정을 분석하였다. 로그인 과정을 각 단계별로 구분하여 분석하고 각 단계에서 발생하는 패킷의 내용을 조사함으로써 엔터프라이즈 네트워크 내의 Skype 사용자(호스트) 각각의 동적 포트번호{IP, port}를 추출 할 수 있었다. 또한 추출된 사용자들의 리스트를 내부 Skype 사용자와 연관되어 트래픽을 발생시키는 외부 호스트의 {IP, port}를 추출 하였다. 이렇게 추출된 내부 및 외부 사용자의 {IP, port} 리스트를 기반으로 Skype 트래픽을 탐지하는 탐지 모듈을 개발하여 적용한 결과 신뢰성 있는 결과를 얻어 낼 수 있었다.[1] 또한 제안하는 시스템에서는 리스트 기반의 분석 방법과 함께 Skype 응용 트래픽에서 고정적으로 발생하는 패턴들을 정의하고 이러한 패턴들을 트래픽 탐지에 적용하는 패턴 기반의 탐지 방법과 시그니쳐 기반의 탐지 방법을 적용할 수 있는 모듈을 개발해 함께 적용했다. 이러한 리스트, 패턴, 시그니쳐 3가지 응용 탐지 방법을 종합해 적용해 본 결과 기존의 리스트 기반의 방법으로 탐지할 수 없었던 Skype 응용 트래픽을 정확하게 탐지할 수 있었다.

본 논문에서 정의하는 Skype 응용 트래픽은 해당 응용 (Skype.exe프로세스)에서 발생하는 모든 트래픽으로 정의한다. Skype 트래픽의 탐지는 대상 네트워크 링크로부터 수집되는 모든 트래픽을 입력으로 하여 Skype응용 트래픽만 탐지하여 분류해 내는 것을 목표로 한다. 본 논문에서는 학내 네트워크의 인터넷 링크를 트래픽 수집 지점으로 설정하였다.

본 논문은 본 장 서론에 이어, 2장에서 관련 연구에 대해 언급하고, 3장에서 본 논문에서 제안하는 Skype 응용 트래픽 탐지를 위한 시스템의 구조와 탐지 알고리즘에 관해 설명한다. 이어 4장에서는 Skype 응용 트래픽 탐지 시스템에 학내 망에서 수집한 Skype 트래픽을 적용한 실험 결과를 서술한다. 5장에서 결론 및 향후 연구로써 본 논문을 마친다.

Ⅱ. 관련 연구

이미 기존의 많은 연구들에서 응용 트래픽 분류를 위한 DPI 기반, 기계학습 기반, 패킷 사이즈와 포트 기반 방식들이 제안되고 있지만 Skype응용에 대해서는 분류의 정확성이 높지 않거나 향후 연구로 남겨두고 있다[2,3,4,5].

[2],[3]에서는 트래픽 분석을 위해 payload 기반 분류 방법론을 제안했으나 Skype응용에 대해서는 그 데이터가 암호화 되어있어 분류가 어렵다고 기술하고 있다. Payload기반 분류 방법은 데이터 부분이 암호화 되어있으면 분류가 어렵겠지만, 시그니쳐 자동 생성 시스템을 활용할 경우 암호화된 데이터에서도 고품질의 응용 시그니쳐를 생성할 수 있고, 이러한 시그니쳐들은 트래픽 분류 시 중요한 단서가 될 수 있다.

[4],[5]에서는 기계학습 기법을 이용해 P2P응용 프

로그램을 분류하는 방법을 제안하고 있다. 하지만 제안하는 방법론은 별도의 학습을 위한 데이터 셋이 필요하며 Skype응용에 한해서는 분류 정확성이 높지 않은 결과를 보여주고 있다. 본 논문에서 제안하는 시스템은 별도의 학습을 위한 데이터 셋이 필요하지 않으며 동적 포트번호를 추출과 동시에 정확하게 분류 할 수 있는 장점을 지니고 있다.

[6]에서는 패킷 사이즈 분포를 이용한 분류 방법론을 제안하고 있다. 하지만 Skype 구 버전인 3.0에 대해 적용하였고, 동적인 Skype 호스트 각각의 포트를 추출하기 어렵다는 단점을 가지고 있다. 본 논문에서도 일부 패킷 사이즈를 이용한 분류 방법을 제안하고 있지만 Skype 최신 버전에 대해 적용이 가능하며 Skype 호스트 각각의 포트를 추출함에 따라 엔터프라이즈 네트워크 내에 발생하는 Skype응용의 트래픽에 대해 다양한 측정이 가능하다. 제안하는 시스템에서는 Skype응용의 트래픽 분류시스템을 위해 Skype사용자 각각의 {IP, port}를 추출함과 동시에 분류 작업을 수행하며 분류의 정확성과 속도를 목표로 하였다.

[7]에서는 Skype 응용의 로그인 세션에서 탐지 가능한 UDP 시퀀스 시그니처와 IP 대역을 활용한 탐지 방법론을 제안하고 있다. 하지만 UDP 시퀀스 시그니처는 현재의 최신 버전의 Skype 응용에 대해 제한적으로만 적용가능하고 IP 대역 기반의 탐지 방법 역시 정확도가 낮은 문제점이 있다. 본 논문에서 제안하는 시스템 역시 IP 대역 기반의 탐지 방법을 사용하고 있지만 다른 3가지의 탐지 방법과 함께 적용하기 때문에 기존 방법에 비해 오탐률을 낮출 수 있었다.

Ⅲ. Skype 응용 트래픽 탐지 방법

본 장에서는 제안하는 Skype 응용 트래픽 탐지 방법에 대해 설명한다. Skype 트래픽 탐지 방법은 2 단계로 구성되어 있다.

첫 단계에서는 Skype 응용의 로그인 세션 분석을 통해 Skype를 사용하는 호스트의 IP(SC-IP)와 Skype 클라이언트에서 호스트에 할당하는 동적 포트(SC-Port)를 탐지하고 {SC-IP, SC-Port} 정보를 리스트로 구축한다. 이 과정에서는 패턴 기반 탐지 방법이 적용된다. 다음 단계에서는 리스트를 통해 탐지된 {SC-IP,SC-Port} 정보를 바탕으로 해당 호스트에서 사용하는 Skype 응용의 서비스 기능들을 분류하고 탐지한다. 해당 과정에서는 리스트, 시그니처 기반의 탐지 방법이 적용되고 추가적으로 IP간의 상관 관계를 이용한 correlation 방법과 IP 대역을 기반으로 한 IP range 방법을 적용 한다.

본 논문에서는 정확한 Skype 응용 트래픽 탐지를 위해 Skype 트래픽을 그림 1과 같이 사용되는 포트와 프로토콜에 따라 크게 4가지 종류로 분류한다.

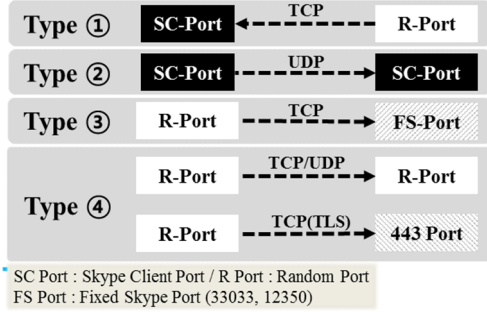

그림 1. Skype 응용 트래픽의 4가지 종류
Fig. 1. 4-Types of Skype Application Traffic

Type 1, 2, 3의 트래픽 flow는 Skype 응용 클라이언트의 동적 포트인 SC-Port와 Skype 고정 포트를 통해 통신하기 때문에 반드시 탐지해야하는 종류의 flow이다. Type 4의 flow들은 주로 Skype 서비스 사용 과

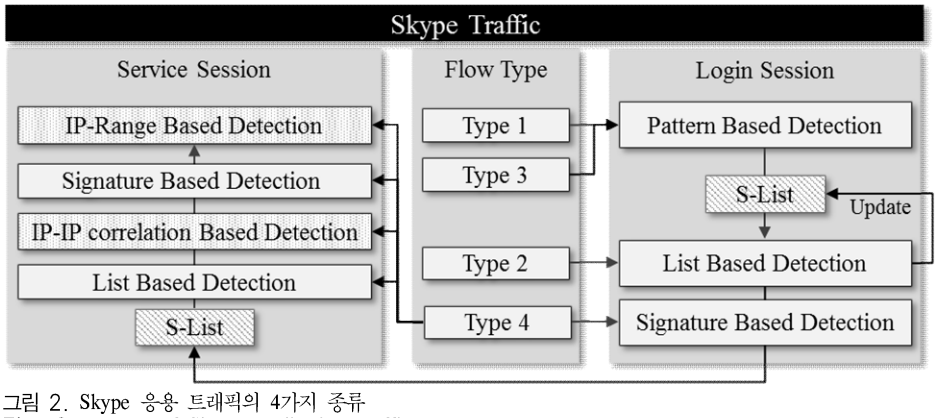

그림 2. Skype 응용 트래픽의 4가지 종류
Fig. 2. 4-Types of Skype Application Traffic

정에서 많이 발생하는 flow 이다. 특히 임의의 포트 간(R-Port/R-Port) UDP 연결을 통해서 실질적인 대부분의 데이터가 발생하기 때문에 서비스 세션 탐지를 위해서는 Type 4의 flow 또한 탐지할 수 있어야 한다. 반면 Type 4의 TLS flow는 Skype 응용의 로그인, 서비스 세션 모두에서 공통적으로 발생하는 flow로서 Skype 응용 뿐만 아니라 타 응용에서도 많이 사용되는 프로토콜이기 때문에 정확한 Skype 응용 트래픽 탐지를 위해서는 Skype 응용에서 발생하는 TLS flow만을 정확하게 탐지할 수 있어야 한다.

전체적인 Skype 응용 트래픽 탐지 방법은 그림 2와 같이 우선 로그인 세션 탐지 과정을 수행하고 이후 서비스 세션 탐지 과정을 수행하는 과정을 따른다.

로그인 세션 트래픽 분석의 시작은 패턴 기반의 탐지 방법이다. Type 1과 Type 3의 flow들을 패턴 기반의방법을 통해 탐지하고 결과를 바탕으로 초기 S-List를 구축한다. S-List는 Skype 사용 호스트의 IP와 SC-Port 정보가 나열된 리스트로서 리스트 기반 탐지 방법에 사용된다. Type 2의 flow들은 리스트 기반의 방법으로 탐지된다. 리스트 기반 탐지 과정에서 탐지되는 모든 Type 2의 UDP flow들은 발신 및 수신 쪽의 포트가 모두 SC-Port이기 때문에 이를 바탕으로 S-List에 매칭된 flow의 정보를 추가해 리스트를 최신화 한다. 로그인 세션 트래픽 중 Type 4의 flow들은 시그니쳐 기반의 방법을 통해 탐지된다. 시그니쳐 기반의 탐지 방법에서 사용되는 시그니쳐는 시그니쳐 자동 생성 시스템을 통해 생성된다.

서비스 세션 트래픽에서 발생하는 Type 4의 flow에 대해서는 리스트와 시그니쳐 기반의 탐지 방법을 적용한다. 로그인 세션 탐지 과정에서 작성된 S-List를 적용해 리스트 기반 탐지 방법을 수행한다. 추가적으로 IP 상관관계, 시그니쳐, IP 대역 기반 탐지 방법을 순차적으로 수행한다.

3-1. Skype 로그인 세션 탐지

그림 2의 내용과 같이 Skype 응용 트래픽 탐지의 첫 단계는 Skype 로그인 세션 탐지이다. 로그인 세션이 중요한 이유는 로그인 세션 분석을 통해 Skype 사용 호스트의 SC-Port를 찾을 수 있기 때문이다. SC-Port를 통해 Skype 응용에서 발생하는 로그인 트래픽의 TCP flow(Type 1)와 모든 UDP flow(Type 2)를 탐지할 수 있고 Skype 사용 호스트를 판별해 리스트(S-List)를 구축할 수 있다. 이러한 SC-Port를 탐지하기 위해 패턴 기반의 탐지 방법을 적용한다.

Skype 응용 로그인 트래픽에서 발생하는 패턴을 정의하기 위해 약 20세트 이상의 로그인 트래픽을 수집하고 해당 세트에서 항상 발생하는 패킷 사이즈 분포와 5-Tuple 정보를 바탕으로 패턴을 정의했다.

표 1. Skype 응용 로그인 트래픽 패턴
Table 1. Skype Application Login Traffic Pattern

Pattern #	Src. IP	Src. Port	Prot.	Dst. IP	Dst. Port	Pkt Size Distribution [Pkt # : (+/-)Byte]
1	SC-IP	R-Port	TCP	ANY	33033	3:+66 / 4:-64/ 6:-416
2	ANY	ANY	TCP	SC-IP	SC-Port	4:64
3	SC-IP	R-Port	TCP	ANY	12350	3:+66/ 5:-361/ 6:+462/ 7:-68/ 9:-78
4	SC-IP	R-Port	TCP	ANY	12350	3:+66/ 5:-81/ 6:-76/ 7:-74

패턴은 표 1과 같이 정의된다. 1, 3, 4번의 패턴은 Type 3의 flow들을 탐지하기 위해 구성된다. 33033과 12350 포트는 Skype 응용에서 고정적으로 사용되는 포트(Fixed Skype Port)이기 때문에 패턴을 통해 정확하게 탐지할 수 있다. 2번 패턴은 Type 1의 flow를 탐지하기 위한 패턴으로서 로그인 세션 분석에서 가장 중요한 Skype 사용 호스트의 SC-Port를 탐지할 수 있다.

그림 3는 Skype 응용의 로그인 과정에서 발생하는 Skype Client와 Super Node간 주요 연결들을 나타낸다. 이 중 그림3 오른쪽 하단의 'SC-Port Detection'에 해당되는 연결을 통해 Skype 사용 호스트의 SC-Port를 탐지할 수 있다.

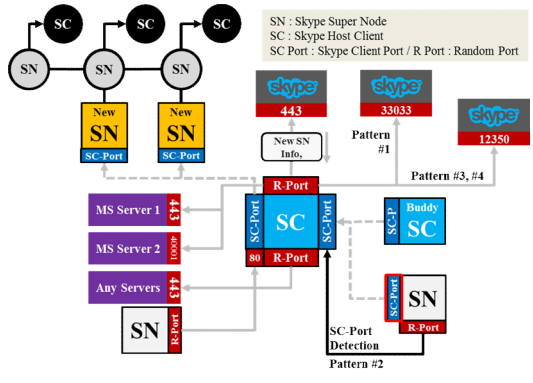

그림 3. Skype Client Port(SC-Port) 탐지 방법
Fig. 3. Skype Client Port(SC-Port) Detection Methods

탐지된 Skype 사용 호스트의 {SC-IP,SC-Port} 정보는 Skype 리스트에(S-List)에 등록된다. S-List는

Skype 사용 호스트들의 IP와 Port 정보를 나열한 리스트로서 이후 리스트 기반 탐지 방법에 사용된다. 패턴 기반의 탐지 방법을 통해 Skype 사용 호스트의 정보가 초기 S-List에 등록되면 리스트와 시그니쳐 기반 탐지 과정을 순차적으로 적용한다.

3-2. Skype 서비스 세션 탐지

Skype 응용 트래픽에 대한 로그인 세션 탐지 과정 후 서비스 세션 탐지를 수행한다. Skype 응용의 서비스 세션에서는 Type 4의 flow들이 발생한다. 표 2는 Skype 응용에서 사용할 수 있는 주요 통화 기능에서 발생하는 Type 4 트래픽들의 TCP/UDP 패킷 비율을 나타낸다.

표 2. Skype 서비스 기능별 TCP/UDP 패킷 비율
Table 2. Rate of TCP/UDP packets by Skype service function

Protocol	Voice Call Pkts	Video Call Pkts
TCP	15~19%	7~10%
UDP	81~85%	90~93%

표 2와 같이 음성/영상 통화에서는 UDP 패킷의 비율이 높은 것을 확인할 수 있고, 실제 Skype 응용의 서비스 트래픽 분석 결과 수신자와 발신자간 실제 통화 데이터의 80%가 UDP 세션을 통해 전송되는 것을 확인했다. 따라서 Skype 응용의 서비스 세션을 탐지하기 위해서는 실제 데이터를 전송하는 UDP flow를 탐지하는 것이 가장 중요하다.

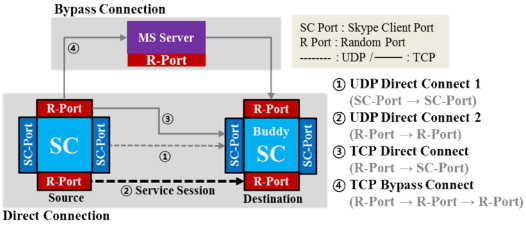

그림 4. Skype 응용 서비스 트래픽 종류
Fig. 4. Skype Application Service Traffic Type

Skype 응용 서비스 트래픽 중 음성/전화 통화 과정에서 발생하는 주요 flow들은 그림 4와 같다. 발생하는 flow 들은 크게 2가지 수신자와 발신자간 직접적으로 연결되는(Direct Connection) flow와 Skype 서버나 Super Node를 우회하는(Bypass Connection) flow로 분류된다.

직접 연결 flow는 3가지 종류로 나누어진다. 수진자와 발신자간 SC-Port를 사용하는 연결(1)과 R-Port(Random Port)를 사용하는 UDP 연결(2)이 있다. R-Port와 SC-Port간 TCP 연결(3)은 불규칙적으로 나타나기 때문에 탐지에서 제외한다. 우회 연결 flow는 R-Port와 R-Port를 이용한 TCP 연결이 존재한다. 4가지 종류의 서비스 flow 중에서 서비스 세션으로 판단할 수 있는 flow는 (2)번 flow 이다. Skype 응용의 서비스 트래픽을 분석했을 때 실제 데이터를 전송하는 UDP flow들 중에서 대부분의 데이터를 전송하는 flow는 직접 연결 flow 중 (2)번 flow이다. 따라서 (2)번에 해당되는 R-Port / R-Port 기반의 UDP flow 탐지를 할 수 있다면 Skype 응용의 서비스 세션 탐지가 가능하다. 하지만 문제점은 해당 UDP flow가 R-Port 기반의 통신을 한다는 점이다. SC-Port를 사용한다면 리스트 기반의 탐지 방법을 통해 탐지할 수 있지만 임의의 포트를 사용하기 때문에 리스트 기반의 탐지 방법으로는 서비스 세션을 탐지할 수 없다.

서비스 세션을 탐지하기 위해서는 IP간 상관 관계를 이용한 IP-IP correlation 방법과 시그니쳐 기반의 탐지 방법이 같이 적용되어야 한다.

IP-IP correlation 방법은 발신/수신 IP를 통한 탐지 방법으로 이전에 탐지된 flow들의 2-Tuple(Source IP, Destination IP) 정보를 바탕으로 flow들을 탐지하는 분석 방법이다. 따라서 이전에 리스트 기반의 탐지 방법을 통해 탐지된 그림 4의 (1)번 flow의 2-Tuple 정보를 바탕으로 (2)번 flow를 탐지할 수 있다. (1)번 flow는 SC-Port를 사용한 연결이기 때문에 리스트 기반 탐지 방법에 의해 탐지할 수 있다. 하지만 상관 관계 기반의 탐지 방법은 2-Tuple 정보만을 사용하기 때문에 정확성이 낮은 위험이 있다. 따라서 보다 정확한 탐지를 위해 시그니쳐 탐지 방법을 같이 적용한다.

표 3. Skype 서비스 기능별 시그니쳐 기반 탐지율
Table 3. Signature-based Detection Rate by Skype Service Function

Skype Service	# of Signatures	Flow Completeness	Packet Completeness	Byte Completeness
Voice Call	14	69%	99%	99%
Video Call	20	66%	98%	99%

분석 결과 시그니쳐 기반의 방법 만을 적용했을 때

flow 기준에서의 탐지율은 상대적으로 낮지만 Packet과 Byte 단위에서는 거의 모든 서비스 트래픽을 탐지할 수 있다. 따라서 시그니쳐 기반의 방법을 같이 사용한다면 IP-IP correlation 탐지 방법만 적용했을 때 나타날 수 있는 Type 2 error flow들을 제외하고 Skype 서비스 세션 flow를 정확하게 탐지할 수 있다. Type 4의 Skype 응용 서비스 트래픽은 UDP flow 외에도 다양한 TCP flow들로 구성되어 있다. 이 중 TLS flow들은 서비스 세션과 로그인 세션에서 또한 많이 발생하는 flow이다. 따라서 Skype 응용의 TLS flow를 정확하게 탐지하는 방법이 중요하다. TLS 세션은 Skype 외에 다양한 응용에서 활용되는 프로토콜이기 때문에 정확하게 Skype의 TLS flow를 탐지하는 방법이 필요하다. 실험 결과 IP 대역을 기반으로 한 탐지 방법과 시그니쳐 기반의 탐지 방법을 같이 적용했을 때 보다 정확한 Skype 응용의 TLS flow를 탐지할 수 있었다.

IP 대역 기반의 탐지 방법은 Skype 응용 flow가 발생하는 IP 대역의 범위를 설정하고 범위 내에서 발생하는 443 Port를 사용하는 flow를 탐지한다. IP 대역만을 탐지 조건으로 활용하기 때문에 IPIP correlation 분석 방법과 마찬가지로 정확성이 낮은 단점이 있다. 따라서 이러한 단점을 보완하기 위해 시그니쳐 기반의 탐지 방법을 함께 적용한다. 생성된 시그니쳐 중 TLS를 탐지할 수 있는 시그니쳐는 표 4와 같다.

표 4. Skype TLS flow 탐지 시그니쳐 예
Table 4. Skype TLS flow Detection Signature Example

Sig. #	1	2	3
Protocol	TCP	TCP	TCP
Src.IP	Any	65.55.252.167	Any
Src.Port	443	443	Any
Dst.IP	Any	Any	111.221.123.231
Dst.Port	Any	Any	Any
Content	“.skype”,“|02||03||01||00||01||a3|”	“msg.skype.com”,“msg.skype.live.com”	“registrar.skype.com”,“ui.skype.com”

본 논문에서 제안하는 Skype 응용 탐지 시스템은 패턴, 리스트, 시그니쳐 기반의 탐지 방법뿐 만 아니라 정확도 향상을 위해 flow에 따라 추가적으로 상관관계, IP 대역 기반의 탐지 방법을 함께 적용한다. 다양한 탐지 방법을 적용하는 궁극적인 이유는 Skype 응용에서 발생하는 다양한 flow 패턴을 보다 정확하고 유연하게 탐지하기 위해서이다.

Ⅳ.Sky-Scope 시스템의 구조와 분석 실험

본 논문에서 제안하는 Sky-Scope 시스템은 Skype 응용 트래픽을 정확하게 탐지하기 위한 오프라인 시

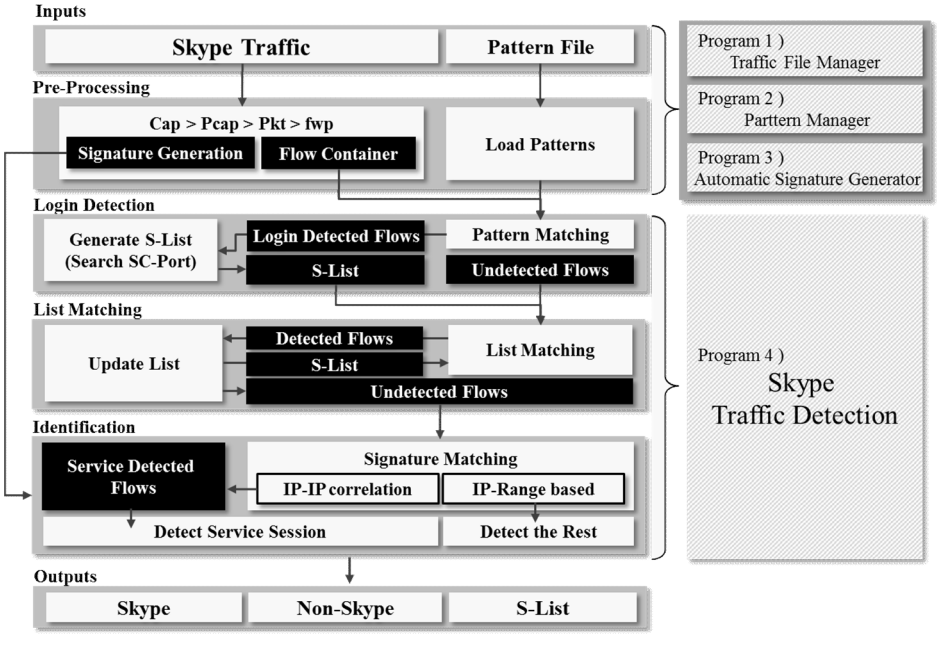

그림 5. Sky-Scope 시스템 구조
Fig. 5. Sky-Scope System Structure

스템이다. 시스템은 그림 5와 같이 크게 4가지 프로그램으로 구성되고 이 중 트래픽 탐지 프로그램은 다시 3가지 모듈로 구성된다.

Skype 응용 트래픽 캡쳐 파일과 3장에서 정의한 패턴 파일을 입력으로 받아 전처리 과정을 수행한다. 전처리 과정에서는 트래픽 캡쳐 파일을 flow 형식으로 바꿔주는 작업과 시스템 내로 패턴 파일을 로드해 파싱하고 시그니쳐를 생성하는 작업을 수행한다. 전처리 과정은 프로그램 1~3번에 해당한다. 전처리 과정이 완료되면 Skype 트래픽 탐지 프로그램에서 Skype 응용 트래픽 탐지 작업을 수행한다. 탐지 프로그램은 3가지 모듈로 구성된다.

Login Detection 모듈에서는 입력으로 받은 Skype 응용 트래픽에 패턴 기반 탐지 방법을 적용해 Skype 응용에 로그인한 호스트를 판단한다. Login Detection 과정에서 가장 중요한 점은 그림 2의 SC-Port Detection flow를 탐지하는 것이다. 표 1의 2번 패턴 매칭을 통해 SC-Port를 탐색할 수 있다. 따라서 Login Detection 과정을 통해 수집한 트래픽 내에 존재하는 모든 Skype 응용 로그인 호스트를 찾아낼 수 있고 탐지된 로그인 호스트의 IP(SC-IP)와 포트(SC-Port) 정보를 바탕으로 초기 리스트(S-List)를 작성한다.

List Matching 모듈에서는 Login Detection 모듈에서 생성된 초기 S-List를 입력으로 받아 Skype 응용 트래픽을 탐지한다. S-List에 등록된 Skype 호스트와 통신을 하는 모든 UDP flow(Type 2)에 대해서 Skype 응용으로 판단하고 등록되지 않은 SC-IP, SC-Port 정보를 탐지해 S-List에 업데이트한다. 업데이트 된 S-List를 적용해 리스트 기반 탐지를 반복해 최종 S-List를 확정한다. S-List를 통해 탐지되는 UDP flow를 모두 Skype 응용 트래픽으로 판단하는 이유는 S-List에는 Skype 사용 호스트의 SC-Port 정보가 등록되어있기 때문에 SC-Port를 통해 통신하는 모든 flow(Type 2) 또한 Skype 응용의 flow로 판단할 수 있다.

Identification 모듈에서는 시그니쳐 기반의 탐지를 바탕으로 IP 상관 관계 분석과 IP 대역 기반의 탐지 방법을 같이 수행한다. IP-IP correlation 탐지 수행을 위해 List Matching 모듈에서 확정된 S-List에 등록된 호스트 들을 바탕으로 2-Tuple 정보를 생성한다. 생선된 2-Tuple 정보와 전처리 단계에서 생성된 시그니쳐를 함께 적용해 그림 4의 (2)번에 해당하는 Skype 응용의 서비스 세션 flow를 탐지한다. 탐지된 서비스 세션 flow를 bps(Bit per Second) 기준으로 분류해 사용한 Skype 서비스 기능을 탐지한다. Bps를 통한 서비스 기능 분류는 음성 통화는 8bps 이상, 영상 통화는 128bps를 기준으로 한다.

서비스 세션 flow를 제외한 나머지 flow들(Type 4)에 대해서는 IP 대역과 시그니쳐 기반의 탐지 방법을 함께 적용해 탐지한다. 사용한 IP 대역은 표 5와 같다. IP 대역에는 MS/Skype 외에도 Google, AKAMAI, AppNexus와 같은 CDN 또한 포함된다. 그러나 시그니쳐 기반의 방법과 함께 적용하기 때문에 Type 2 error(False Postive) flow를 필터링하고 오탐을 방지할 수 있다.

표 5. Skype 응용 트래픽의 사용 IP 대역
Table 5. Skype Application Traffic IP Range

Service Name	IP Range
Skype	91.190.219.0 ~ 91.190.219.255
	91.190.219.0 ~ 91.190.219.255
Akamai	23.0.0.0 ~ 23.15.255.255
	23.32.0.0 ~ 23.67.255.255
	23.192.0.0 ~ 23.223.255.255
	104.64.0.0 ~ 104.127.255.255
	118.214.0.0 ~ 118.215.191.255
	125.56.128.0 ~ 125.56.255.255
	184.24.0.0 ~ 184.31.255.255
	184.84.0.0 ~ 184.87.255.255
AppNexus	103.243.22.0 ~ 103.243.221.255
Microsoft	13.64.0.0 ~ 13.107.255.255
	23.96.0.0 ~ 23.103.255.255
	40.64.0.0 ~ 40.71.255.255
	40.74.0.0 ~ 40.125.127.255
	52.145.0.0 ~ 52.191.255.255
	65.52.0.0 ~ 65.55.255.255
	104.40.0.0 ~ 104.47.255.255
	104.208.0.0 ~ 104.215.255.255
	111.221.29.0 ~ 111.221.29.255
	111.221.64.0 ~ 111.221.127.255
	134.170.0.0 ~ 134.170.255.255
	137.116.0.0 ~ 137.116.255.255
	157.54.0.0 ~ 157.60.255.255
	204.79.195.0 ~ 201.79.197.255

Skype 응용 트래픽 탐지가 끝나면 최종 결과물로 Skype로 탐지된 트래픽과 S-List가 출력된다. Skype가 아닌 Noise 트래픽을 입력 트래픽에 섞어 탐지를 수행할 경우 Non-Skype 트래픽이 추가로 출력된다. 본 논문에서 제안하는 Skype 응용 트래픽 탐지 시스템의 타당성을 검증하기 위해 Skype 응용에 대한 검증 실험을 진행했다. 검증 실험은 학내망 내의 두 호스트에서 Skype 응용의 로그인과 서비스 기능 별로 총 19개의 Skype 트래픽 세트를 수집해 Sky-Scope

시스템에 적용해 탐지율(Recall)을 측정하는 방법으로 수행되었다. 19개의 트래픽 세트는 TMA/S(Traffic Measure Agent/Server) 통해 검증된 Skype 응용 정답지(Ground Truth) 트래픽이다. 실험 트래픽 정보는 표 6과 같다.

실험 트래픽은 Skype 응용의 로그인 트래픽 11개 서비스 트래픽 8개로 구성된다. 제안하는 Sky-Scope 시스템에 실험 트래픽을 적용했을 때 호스트의 SC-IP와 SC-Port를 정확히 탐지하는지 검증하고 로그인 세션에 대한 Duration, Flow, Packet, Byte 사이즈와 해당 호스트의 Buddy List 정보를 확인한다. 또한 서비스 세션에 대해 데이터 flow 탐지 및 서비스 기능을 분류하는지 확인하고 통신을 한 상대방 호스트의 SC-IP, SC-Port 정보를 확인한다.

실험 결과는 표 7과 같다. Sky-Scope 시스템에 실험 트래픽을 적용했을 때 패턴 기반 탐지 결과 두 로그인 호스트를 정확하게 탐지했고 리스트 기반 탐지를 수행해 총 211 개의 {SC-IP, SC-Port} 정보가 S-List에 등록되었다. 시그니쳐 자동 생성 시스템에 실험 트래픽 적용 결과 총 75개의 시그니쳐가 생성되었다. 생성된 시그니쳐를 이용한 시그니쳐 기반의 탐지와 IP-IP correlation 방법을 적용해 서비스 세션 탐지 결과 총 8개의 서비스가 정확하게 탐지되었다. Type 4의 TLS 트래픽에 대해서는 IP 대역 기반의 시그니쳐 매칭 탐지를 수행한 결과 최종적으로 flow 기준 98.6%, 패킷 기준 99.7%, 바이트 기준 99.8%로 거의 모든 Skype 응용 트래픽을 탐지하는 것을 검증하였다.

표 7. Skype 응용 트래픽 적용 실험 결과
Table 7. Skype Application Traffic Experiment Result

Detection Method	Flow Completeness	Pkt Completeness	Byte Completeness
1. List	39.6% (578/1,461)	26.7% (2,501/9,380)	6.6%
2. Signature & IP-IP Correlation	56.3% (824/1,461)	69.4% (6,514/9,380)	91.5%
3. Signature &IP-Range	2.6% (39/1,461)	3.6% (338/9,380)	1.7%
Total Completeness	98.6%	99.7%	99.8%

분석 결과 시그니쳐와 IP 상관 관계 분석을 통해 가장 많은 Skype 응용 트래픽을 탐지할 수 있었다. 제안하는 Sky-Scope에서 탐지하지 못한 20개의 flow는 모두 Type 4에 해당하는 TLS flow였다. 검증 결과 탐지되지 않은 TLS flow는 시그니쳐 매칭 및 IP대역 모두 일치하지 않았다. 이러한 미탐지 flow(FN)에 대해서는 IP대역을 추가하거나 시그니쳐를 업데이트 할 경우 탐지 가능하다. 또는 추가적인 지속시간 기반의 탐지 방법 또한 현재 적용할 계획이기 때문에 추후 Sky-Scope 시스템상에서 탐지 가능할 것이라 판단된다.

표 6. Skype 응용 실험 트래픽
Table 6. Skype Application Experimental Traffic

Trace Name	Function	Host SC-IP	SC-Port	Receive Host IP	Receive Host SC-Port	Size(KB)
Login_h1-1	Login	163.152.219.203	30838			1,414
Login_h1-2	Login	163.152.219.203	30838			2,565
Login_h1-3	Login	163.152.219.203	30838			3,165
Login_h1-4	Login	163.152.219.203	30838			4,334
Login_h1-5	Login	163.152.219.203	30838			3,031
Login_h2-1	Login	163.152.219.191	8944			365
Login_h2-2	Login	163.152.219.191	8944			327
Login_h2-3	Login	163.152.219.191	8944			301
Login_h2-4	Login	163.152.219.191	8944			276
Login_h2-5	Login	163.152.219.191	8944			430
Login_h2-6	Login	163.152.219.191	8944			306
VoiceCall_h1toh2-1	Voice Call	163.152.219.203	30838	163.152.219.191	8944	994
VoiceCall_h1toh2-2	Voice Call	163.152.219.203	30838	163.152.219.191	8944	1,687
VoiceCall_h2toh1-1	Voice Call	163.152.219.191	8944	163.152.219.203	30838	824
VoiceCall_h2toh1-2	Voice Call	163.152.219.191	8944	163.152.219.203	30838	1,650
VideoCall_h1toh2-1	Video Call	163.152.219.203	30838	163.152.219.191	8944	7,012
VideoCall_h1toh2-2	Video Call	163.152.219.203	30838	163.152.219.191	8944	7,154
VideoCall_h2toh1-1	Video Call	163.152.219.191	8944	163.152.219.203	30838	7,209
VideoCall_h2toh1-2	Video Call	163.152.219.191	8944	163.152.219.203	30838	7,213

Ⅵ. 결론 및 향후 과제

본 논문에서는 Skype 응용 트래픽 탐지를 위한 Sky-Scope시스템의 구조와 탐지 알고리즘에 대해 서술하고 간단한 실험을 통해 시스템의 타당성을 검증했다. 패턴, 리스트, 시그니쳐 기반의 탐지 방법을 통해 제안하는 시스템은 이전의 다른 응용 탐지 시스템과 비교해 보다 다양하고 정교하게 Skype 응용 트래픽을 분석할 수 있다. 또한 실제 Skype 응용 트래픽을 적용한 실험 결과 또한 높은 응용 탐지율을 나타내고 있다. 이전의 다른 시스템과 다르게 Skype 응용의 로그인 세션을 통해 호스트를 탐지하고 나아가 서비스 세션 및 종류까지 탐지할 수 있는 시스템을 설계하고 검증했다는 점에 본 연구의 의의가 있다고 판단된다.

하지만 현재 탐지하지 못하는 Skype 응용 flow들도 존재하기 때문에 향후 시스템 개선 및 안정화 작업을 통해서 이러한 문제점을 보완하고 추가적인 탐지 방법들 또한 적용해볼 계획이다. 또한 현재 시스템은 오프라인에서 미리 수집한 트래픽을 기반으로 동작하기 때문에 상대적으로 시스템을 구성하는 프로그램들에 가해지는 부하가 적다. 향후 실시간으로 망 내에서 동작하는 시스템으로 발전시켜 시스템의 성능을 향상시킬 계획이다.

References

[1] S. W. Park, H. S. Lee, M. J. Choi and M. S. Kim,"Real-time Identification of Skype Application Traffic using Behavior Analysis", *KICS*, vol.36, No.2, pp.131-140, Feb. 2011.

[2] Risso, F., Baldi, M., Morandi, O., Baldini, A., Monclus, P.,"Lightweight, payload-based traffic classification: An experimental evaluation", *In Proceedings of IEEE International Conference on Communications ICC*, pp.5869-5875, 2008.

[3] N. Cascarano, L. Ciminiera, F. Risso,"Improving Cost and Accuracy of DPI Traffic Classifiers", *25th ACM Symposium On Applied Computing(SAC 2010)*, pp.643-648, March. 2010.

[4] H.Liu, W.Feng, Y.Huang, X.Li,"A peer-to-peer traffic identification method using machine learning", *in International Conference on Networking, Architecture, and Storage, NAS*, pp.155-160, July. 29-31, 2007.

[5] Zhu Li, Ruixi Yuan, and Xiaohong Guan,"Accurate Classification of the Internet Traffic Based on the SVM Method" *Proc. of the IEEE International Conference on Communications*, pp.1373-1378, Jun. 24-28. 2006.

[7] Ying-Dar Lin, Chun-Nan Lu, Yuan-Cheng Lai, Wei-Hao Peng, Po-Ching Lin," Application classification using packet size distribution and port association", *Journal of Network and Computer Applications*, Vol.32, pp.1023-103, Sep. 2009.

[8] Z. Yuan, C. Du, X. Chen, D. Wang, Y. Xue,"SkyTracer: Towards fine-grained identification for Skype traffic via sequence signatures", *Computing, Networking and Communications (ICNC) 2014 International Conference*, Feb. 3-6. 2014.

이 성 호 (Sung-ho Lee)

2016년~현재 : 고려대학교 컴퓨터 정보학과 석사과정

〈관심분야〉 네트워크 관리 및 보안, 트래픽 모니터링 및 분석, 트래픽 분류

구 영 훈 (Young-Hoon Goo)

2016년~현재 : 고려대학교 컴퓨터 정보학과 석사과정

〈관심분야〉 네트워크 관리 및 보안, 트래픽 모니터링 및 분석, 트래픽 분류

이 민 회 (Min-Hoe Lee)

2012년~현재 : 고려대학교 컴퓨터 정보학과

〈관심분야〉 네트워크 관리 및 보안, 트래픽 모니터링 및 분석, 트래픽 분류

박 지 태 (Ji-Tae Park)

2013년~현재 : 고려대학교 컴퓨터 정보학과

〈관심분야〉 네트워크 관리 및 보안, 트래픽 모니터링 및 분석, 트래픽 분류

김 명 섭 (Myung-Sup Kim)

1998년 포항공과대학교 전자계산학과 졸업

2000년 포항공과대학교 컴퓨터 공학과 석사

2004년 포항공과대학교 컴퓨터 공학과 박사

2006년 Post-Doc. Dept. of ECE, Univ. of Toronto, Canada

2006~2015년 고려대학교 컴퓨터정보학과 부교수

2016년~ 현재 고려대학교 컴퓨터정보학과 교수

〈관심분야〉 네트워크 관리 및 보안, 트래픽 모니터링 및 분석, 멀티미디어 네트워크