P2P 기반의 VoIP 서비스 응용 트래픽 탐지 시스템

이성호, 구영훈, 박지태, 지세현 김명섭
고려대학교

{gaek5, gyh0808, pjj5846, sxzer, tmskim}@korea.ac.kr

P2P based VoIP Service Application Traffic Detection System

Sung-Ho Lee, Young-Hoon Goo, Ji-Tae Park, Se-Hyun Ji and Myung-Sup Kim
Korea Univ.

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 Skype 는 Microsoft 사에서 서비스하는 P2P 기반의 VoIP 서비스로서 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 Skype 트래픽 패턴을 보다 정확하게 분석하고 탐지하기 위해 본 논문은 패턴, 리스트, 시그니쳐 기반의 응용 탐지 방법을 조합한 종합적인 Skype 트래픽 탐지 시스템을 제안한다. 제안된 시스템은 학내 네트워크망을 통해 수집한 다양한 Skype 트래픽에 적용하여 정확성과 탐지율을 검증하였다.

Ⅰ. 서 론

최근 인터넷 사용자의 증가와 고속 네트워크의 보급으로 네트워크 트래픽이 급증하였다. 이것은 단 순히 WWW, FTP, SMTP, DNS 와 같은 전통적인 인터 넷 서비스뿐만 아니라 멀티미디어, P2P (peer-to-peer), P2P 기반으로 게임, 스마트 폰 응용 등의 다양한 서비스의 증가로 인한 부분도 큰 몫을 차지하기 때문이다. 이에 따른 트래픽이 급증함에 따라 효과적인 네트워크 관리를 위해 트래픽 모니터링 및 분석의 중요성이 커지고 있다.

Skype 의 트래픽들은 기본 적으로 암호화가 되어있고, 해당 응용 설치 시 동적 포트 번호가 할당되고, 일반적인 프로토콜을 사용하지 않아 Skype 의 트래픽을 정확하게 탐지하는 것은 일반적인 분석 방법론으로는 불가능하다. 하지만 Skype 응용 설치 때마다 달라지는 클라이언트 포트를 알아낼 수 있다면 대부분의 트래픽 을 쉽게 분류할 수 있으며 각 호스트의 응용 사용 시간 측정 및 제어 등 다양한 방면에서 활용이 가능하다.

Ⅱ. 본론

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 Skype 를 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, 시그니쳐 기반의 탐지 방법이 적용되고 추가적으로 IP 간의 상관 관계를 이용한 correlation 방법과 IP 대역을 기반으로 한 IP range 방법을 적용 한다.

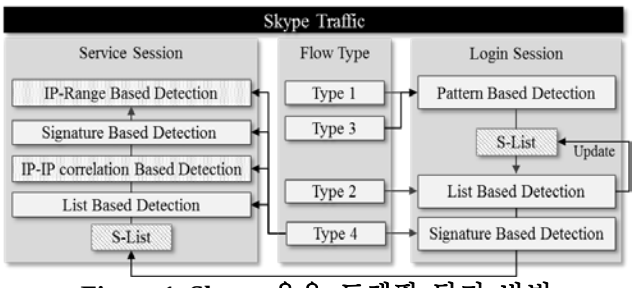

Figure 1. Skype 응용 트래픽 탐지 방법

본 논문에서는 정확한 Skype 응용 트래픽 탐지를 위해 Skype 트래픽을 그림 1 과 같이 사용되는 포트와 프로토콜에 따라 크게 4 가지 종류로 분류한다. 로그인 세션 트래픽 분석의 시작은 패턴 기반의 탐지 방법이다. Type 1 과 Type 3 의 flow 들을 패턴 기반의방법을 통해 탐지하고 결과를 바탕으로 초기 S-List 를 구축한다. S-List 는 Skype 사용 호스트의 IP 와 SC-Port 정보가 나열된 리스트로서 리스트 기반 탐지 방법에 사용된다.

Type 2 의 flow 들은 리스트 기반의 방법으로 탐지된다. 리스트 기반 탐지 과정에서 탐지되는 모든 Type 2 의 UDP flow 들은 발신 및 수신 쪽의 포트가 모두 SC-

이 논문은 2016 년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 과학기술인문융합연구사업 (No.NRF-2016M3C1B6929228)과 2015 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2015R1D1A3A01018057).

Port 이기 때문에 이를 바탕으로 S-List 에 매칭된 flow 의 정보를 추가해 리스트를 최신화 한다. 로그인 세션 트래픽 중 Type 4 의 flow 들은 시그니쳐 기반의 방법을 통해 탐지된다. 시그니쳐 기반의 탐지 방법에서 사용되는 시그니쳐는 시그니쳐 자동 생성 시스템을 통해 생성된다.

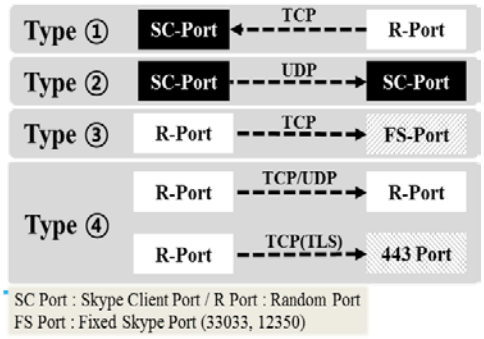

Figure 2. Skype 응용 트래픽의 4 가지 종류

Ⅲ. 실험

본 논문에서 제안하는 Sky-Scope 시스템은 Skype 응용 트래픽을 정확하게 탐지하기 위한 오프라인 시스템이다. 시스템은 그림 3 과 같이 크게 4 가지 프로그램으로 구성되고 이 중 트래픽 탐지 프로그램은 다시 3 가지 모듈로 구성된다.

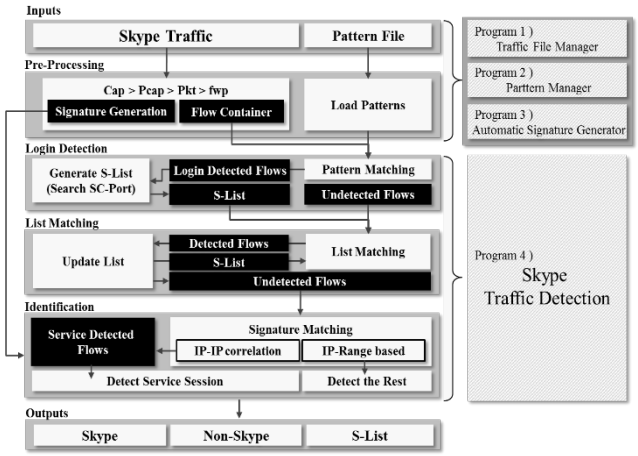

Figure 3. Sky-Scope 시스템 구조

전처리 과정에서는 트래픽 캡쳐 파일을 flow 형식으로 바꿔주는 작업과 시스템 내로 패턴 파일을 불러와 파싱하고 시그니쳐를 생성하는 작업을 수행한다. 전처리 과정은 프로그램 1~3 번에 해당한다. 전처리 과정이 완료되면 Skype 트래픽 탐지 프로그램에서 Skype 응용 트래픽 탐지 작업을 수행한다.

Login Detection 모듈에서는 입력으로 받은 Skype 응용 트래픽에 패턴 기반 탐지 방법을 적용해 Skype 응용에 로그인한 호스트를 판단한다. List Matching 모듈에서는 Login Detection 모듈에서 생성된 초기 S-List 를 입력으로 받아 Skype 응용 트래픽을 탐지한다. Identification 모듈에서는 시그니쳐 기반의 탐지를 바탕으로 IP 상관 관계 분석과 IP 대역 기반의 탐지 방법을 같이 수행한다.

Skype 응용 트래픽 탐지가 끝나면 최종 결과물로 Skype 로 탐지된 트래픽과 S-List 가 출력된다. Skype 가 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 19 개의 Skype 트래픽 세트를 수집해 Sky-Scope 시스템에 적용해 탐지율(Recall)을 측정하는 방법으로 수행되었다.

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%

Table 1. Skype 응용 트래픽 적용 실험 결과

실험 결과는 표 1 과 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-List 에 등록되었다. 시그니쳐 자동 생성 시스템에 실험 트래픽 적용 결과 총 75 개의 시그니쳐가 생성되었다. 생성된 시그니쳐를 이용한 시그니쳐 기반의 탐지와 IP-IP correlation 방법을 적용해 서비스 세션 탐지 결과 총 8 개의 서비스가 정확하게 탐지되었다. Type 4 의 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.

Ⅳ. 결론 및 향후 연구

본 논문에서는 Skype 응용 트래픽 탐지를 위한 Sky-Scope 시스템의 구조와 탐지 알고리즘에 대해 서술하고 간단한 실험을 통해 시스템의 타당성을 검증했다. 하지만 현재 탐지하지 못하는 Skype 응용 flow 들도 존재하기 때문에 향후 시스템 개선 및 안정화 작업을 통해서 이러한 문제점을 보완하고 추가적인 탐지 방법들 또한 적용해볼 계획이다.

참 고 문 헌

[1] Sung-Ho Yoon, Kyu-Seok Shim, Su-Kang Lee, and Myung-Sup Kim , "Framework for Multi-Level Application Traffic Identification," Proc. of the Asia-Pacific Network Operations and Management Symposium (APNOMS) 2015, Busan, Korea, Aug. 19-21, 2015, pp.424-427.

[2] 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.