Session ID - Server IP 캐싱 기반의 SSL/TLS 암호화 트래픽의 서비스 식별 방법

김성민, 구영훈, 김명섭
고려대학교

{gogumiking, gyh0808, tmskim}@korea.ac.kr

A Method for Service Identification of Encrypted Traffic based on SSL/TLS with Session ID - Server IP Caching

Sung-Min Kim, Young-Hoon Goo, Myung-Sup Kim
Korea Univ.

요 약

오늘날 네트워크 트래픽이 복잡, 다양해짐에 따라 발생하는 네트워크 보안문제 해결을 위해 다양한 암호화 프로토콜 중 하나인 SSL/TLS 가 널리 사용되고 있다. 하지만 현재의 트래픽 분석 시스템은 암호화 트래픽을 프로토콜 레벨에 한정적으로 분석하고 있으며, 암호화 이전의 SSL/TLS Handshake 과정을 분석하여 생성한 시그니쳐를 기반으로 SSL/TLS 응용 서비스를 식별할 수 있지만, 세션 재연결을 할 때에는 온전한 Handshake 를 맺지 않기 때문에 페이로드 시그니쳐만으로 모든 SSL/TLS 트래픽을 분석할 수 없다. 따라서 본 논문에서는 SSL/TLS Handshake 과정에서 생성되는 세션 ID 와 서버 IP 집합을 별도의 테이블에 기록하고 이를 이용하여 SSL/TLS 응용 서비스를 식별하는 방법을 제안한다. 제안하는 방법은 최대 99%에 가까운 SSL/TLS 트래픽을 분석하였으며, 페이로드 시그니쳐 기반 분석 방법만을 이용하여 SSL/TLS 트래픽을 분석한 결과보다 최대 27%의 분석률을 증가시켜 그 성능과 가능성을 검증하였다.

Ⅰ. 서 론

오늘날 네트워크 응용 서비스의 다양화와 사용자의 증가로 인해 네트워크 트래픽은 복잡 다양하게 발생하고 있다. 이에 따른 네트워크 과부하를 해결하고, 효과적인 네트워크 자원의 운용과 관리를 위해서는 네트워크 트래픽 분석은 필수적인 요소이다. 또한 개인정보 유출, 사생활 침해, 사용자 계정도용 등 네트워크 보안문제가 증가됨에 따라 다양한 암호화 프로토콜이 개발되어왔다. 다양한 암호화 프로토콜 중 대표적인 SSL/TLS(Secure Sockets Layer/Transport Layer Security)[1]는 TCP/IP 와 마찬가지로 연결지향 프로토콜로 서버와 클라이언트 상호간 인증서와 키 교환을 통해 안전한 통신을 보장한다. 하지만 이때 서버와 클라이언트가 실제 주고받는 데이터는 암호화 되어 그 내용이 숨겨져 있기 때문에 네트워크 관리자 입장에서는 호스트가 어떤 응용 서비스를 이용 하는지 식별하기 쉽지 않다. 하지만 SSL/TLS 의 경우 데이터 암호화 이전의 인증서와 키 교환 단계인 SSL/TLS Handshake 과정은 암호화가 되어 있지 않기 때문에 해당 패킷을 분석하면 트래픽을 특정 응용 서비스로 식별 할 수 있다.

따라서 본 논문에서는 SSL/TLS 트래픽의 Handshake 과정에서 페이로드 시그니쳐를 추출하여 이를 기반으로 SSL/TLS 트래픽을 응용 서비스 단위로 식별하고, 이후 세션을 재연결 하는 경우 세션 ID 와 서버의 IP 의 관계를 이용하여 추가적인 분석을 하는 방법을 제안한다.

본 논문의 구성은 2 장에서는 기존 SSL/TLS 트래픽 분석의 연구에 대하여 소개를 하고, 3 장에서 본 논문에서 제안하는 ID-IP 캐싱기반의 SSL/TLS 응용 서비스 식별 방법에 대하여 기술한다. 4 장에서는 3 장을 바탕으로 구축한 분석시스템과 기존 페이로드 시그니쳐 기반 분석 방법과 비교평가를 하고, 마지막으로 5 장에서는 결론 및 향후 연구에 대하여 기술한다.

Ⅱ. 관련 연구

SSL/TLS 은 데이터를 암호화 하여 전송하는 프로토콜 이지만 TCP/IP 와 같은 연결지향 프로토콜 이기 때문에 서버와 클라이언트가 상호간 인증서와 키를 교환하는 과정 Handshake 이 포함되어 있으며, 이 과정은 암호화 되어 있지 않다. 따라서, SSL/TLS Handshake 의 인증서 교환 레코드를 분석하면 특정 응용 서비스의 시그니쳐를 추출 할 수 있으며, 이를 바탕으로 SSL/TLS 트래픽을 응용 서비스 단위로 식별 할 수 있다. 하지만, 일정시간 클라이언트의 응답이 없으면 SSL/TLS 의 세션은 만료가 되고, 재연결을 할 때 이전 Handshake 에서 서버가 생성한 Session ID 를 이용하여 재연결을 하기 때문에 페이로드 시그니쳐 만으로는 서비스를 식별할 수 없다.

다음으로 패킷 크기, 수집 시간 등 플로우의 통계 정보를 기반으로 응용 서비스를 식별하는 통계정보 기반 분석 방법이 있으며, 패킷을 직접 확인하지 않아도 되기 때문에 암호화 되어있는 트래픽도 식별할 수 있는 장점을 이용 하여 SSL/TLS 트래픽을 분류하는 연구[2]가 진행이 되었다. “ AdaBoost”, “ C4.5”, “ Navie Bayes”등 기계학습 기반 알고리즘으로 SSL/TLS 트래픽을 식별하였지만, 기계학습 알고리즘의 단점인 특정 네트워크에 종속적이라는 것과 프로토콜 단위가 아닌

서비스 별 세부적인 식별을 하지 못한다는 한계점을 벗어나지는 못하였다.

Ⅲ. 본 론

본 장에서는 본 논문에서 제안하는 Session ID-Server IP 캐싱 기반의 SSL/TLS 트래픽의 서비스 식별 방법에 대하여 기술한다.

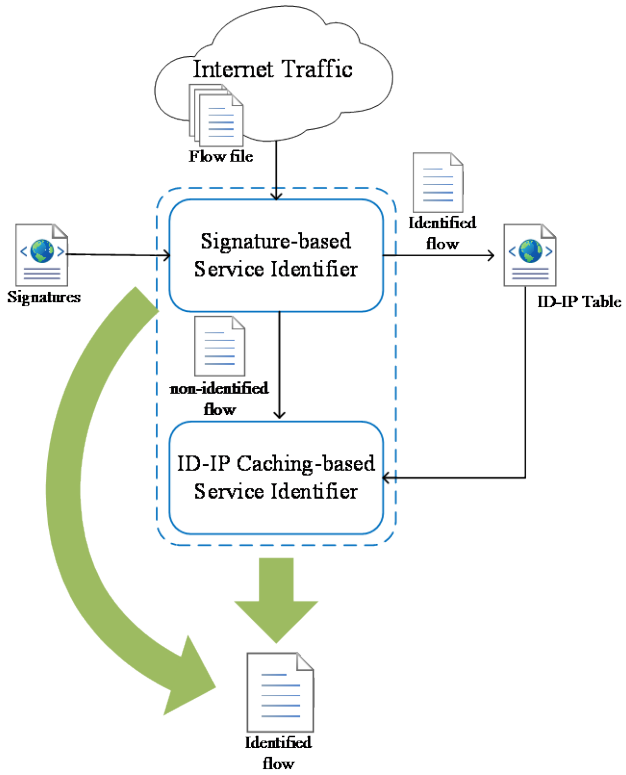

그림 1. SSL/TLS Service Identifier

그림 1 은 제안하는 방법론의 전체 구성도로, 시스템의 입력 데이터는 인터넷상에서 발생한 트래픽을 5-Tuple(Src IP, Dst IP, Src Port, Dst Port, Protocol)을 기준으로 세션별로 모은 파일이다. 시스템은 크게 “ Signature-based Service Identifier”와 “ ID-IP Caching-based Service Identifier” , 두 모듈로 구분될 수 있다. 먼저 Signature-based Service Identifier 는 입력된 SSL/TLS 의 플로우 파일을 페이로드 시그니쳐를 기반으로 분석을 하고, 분석된 플로우의 세션 ID 와 서버 IP 집합을 ID-IP Table 에 기록한다. 이때, 세션 ID 는 서버가 생성하여 클라이언트 측으로 전송하기 때문에 세션 ID 와 서버 IP 는 고유한 집합이라고 할 수 있다. 따라서 ID-IP Caching-based Service Identifier 는 Signature-based Service Identifier 에서 분석하지 못한 SSL/TLS 플로우를 ID-IP Table 에 기록된 집합과 비교하여 특정 응용 서비스를 식별한다.

Ⅳ. 실험 및 평가

본 장에서는 3 장에서 기술한 방법론을 기반으로 구현한 시스템을 이용하여 실제 SSL/TLS 트래픽을 식별하는 실험에 대하여 기술한다. 실험에 사용한 SSL/TLS 응용 서비스는 대표적인 검색 사이트인 Google, 최근 사용량이 늘고 있는 SNS(Social Network Service)인 Facebook 과 Kakaotalk 이며, 그 양은 플로우 단위로 506,412, 패킷 단위로는 59,164,220, 바이트 단위로는 46,224,581,249 이다. 실험 방법은 수집된 트래픽을 페이로드 시그니쳐 기반 분석 방법만을 이용하여 분석한 결과와 이때 생성된 ID-IP Table 을 이용하여 추가적인 분석을 한 결과를 비교, 분석하였으며, 그 결과는 그림 2 와 같다.

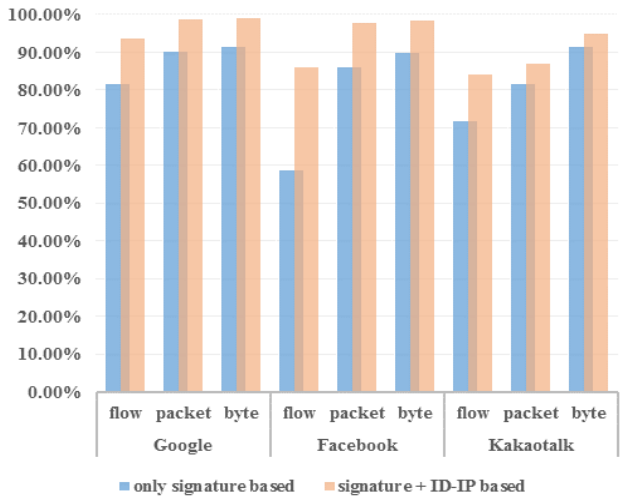

그림 2. SSL/TLS 트래픽 분석 결과

모든 응용, 모든 단위에서 페이로드 시그니쳐 기반 분석 방법만으로 분석한 결과보다 세션 ID 와 서버 IP 캐싱을 이용하여 추가적인 분석을 한 분석 방법이 더 많은 트래픽을 식별하였으며, 플로우 단위로 최대 27%의 분석률이 증가하였다. 이때 분석하지 못한 트래픽은 Handshake 가 제대로 맺어지지 않은 경우, 서버의 IP Table 상에서 클라이언트 정보가 삭제되어 재연결을 거부하는 Alert 메시지 등 네트워크 관리에 있어서 영향이 없는 특수한 경우가 대부분이므로 본 방법론의 타당성을 입증 할 수 있다.

Ⅴ. 결 론

본 논문에서는 SSL/TLS 트래픽을 세션 ID 와 서버 IP 캐싱을 이용하여 서비스 단위로 식별하는 방법을 제안하였으며, 실험을 통하여 기존 페이로드 시그니쳐 기반 분석 방법과 비교해 최대 27%의 SSL/TLS 트래픽을 더 분석함으로 본 방법론의 타당성을 입증하였다.

향후 연구로는 ID-IP Table 의 유효기간 정책을 수립하여 저장공간 관리를 효율화 하여 실시간 분석 시스템에 적용시킬 예정이다.

ACKNOWLEDGMENT

본 논문은 BK21 플러스 사업+ (No. T1300572) 및 2013 년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보. 컴퓨팅기술개발사업(2010-0020728)의 지원을 받아 수행된 연구임

참 고 문 헌

[1] Elgohary, Ashraf, Tarek S. Sobh, and Mohammed Zaki. "Design of an enhancement for SSL/TLS protocols." computers & security 25.4 (2006): 297-306.

[2] McCarthy Curtis, and A. Nur Zincir-Heywood. "An investigation on identifying SSL traffic." Computational Intelligence for Security and Defense Applications (CISDA), 2011 IEEE Symposium on. IEEE, 2011.