DNS 프로토콜을 활용한 사용자 행위 모니터링 시스템 개발

안평주*, 정동현**, 정상훈***, 장재원****, 석지원*****, 이경문******
*한서대학교, **영남이공대학교, ***한양대학교, ****조선대학교,
*****이화여자대학교, ******중부대학교
pung0552@gmail.com, stjhyeon@gmail.com, shone2137@hanyang.ac.kr, qjas03878@gmail.com,
sjw0712@ewhain.net ,gilgil1973@gmail.com

Development of User Behavioral Statistics System Using DNS Protocols

Pyeong-Ju Ahn*, Dong-Hyun Jeong**, Sang-Hoon Jung***, Jae-Won Jang****,
Ji-Won Seok*****, Kyung-Moon Lee******
*Hanseo University, **Yeungnam University College, ***Hanyang University
****Chosun University, *****Ewha Womans University, ******Joongbu University

요 약

본 연구에서는 사용자의 접속 기록을 상세하게 모니터링할 수 있는 DNS 패킷 기반 분석과 사용자별 프로세스 분석 기법을 융합한 네트워크 모니터링 시스템 설계를 제안한다. 네트워크 패킷 수집을 위한 탭 장비와 사용자 디바이스에 플러그인 형태 프로그램 설치를 통하여 어떤 프로세스에서 패킷이 발생하였는지 분석이 가능하다. 이를 통해 네트워크 증설을 위한 데이터 확보, 악성 패킷 분류를 위한 데이터로의 사용 등 다양한 방법으로 활용할 수 있는 확장형 도메인 모니터링 시스템을 제안한다.

1. 서론

최근 몇 년간 인터넷의 발전으로 인해 비약적으로 네트워크 트래픽이 증가함에 따라 안정적인 네트워크 서비스 제공을 위한 네트워크 트래픽 관리의 중요성이 증대되었다. 네트워크 망의 효율적인 증설 및 비용 감축을 위해서는 수많은 네트워크 트래픽의 증가 속에서 주기적인 네트워크 트래픽 변화를 측정할 필요가 있으며 시장에도 이를 측정하기 위한 다양한 제품들이 존재한다 네트워크의 트래픽 관리를 위해서는 사용자의 서비스 이용에 대한 데이터 수집이 필요조건이다. 그리고 이러한 데이터를 수집하고 처리하는 방식에 따라 제품의 기능이 나누어지게 된다. 시장에 존재하는 도구 <표 1>를 조사한 결과, 기존의 네트워크 트래픽 모니터링 제품들은 패킷 흐름을 디바이스에서 모니터링 하는 것에 불과하거나 도메인 별 분류가 불가능하여 사용자가 직관적으로 트래픽의 속성을 파악하는 것이 어렵다는 문제점을 안고 있다.

본 연구는 이러한 문제점을 보완하여 사용자의 행위 및 접속 기록을 상세하게 모니터링 할 수 있는 방법론을 제시한다.

본 논문에서는 각 계층에서 수집한 패킷 해시 값의 동일성을 검증하여 네트워크 계층의 정보와 디바이스 계층의 정보를 연결하고 이를 통해 어떤 프로세스에서 패킷이 발생하였는지에 대해 분석하는 TNS [1] 를 소개한다. 또한 DNS 패킷 분석과 사용자 별 프로세스 분석 기법을 융합해 어플리케이션 별로 세분화된 데이터를 분석하는 새로운 네트워크 모니터링 프로그램 설계를 제안한다.

2. 배경

2.1 관련 연구 사례

DNS 질의 및 응답 패킷은 클라이언트 IP, 서버 IP, 포트 및 해당 서비스의 이름을 비롯한 다양한 정보를 포함한다. DNS 패킷에 대한 성행 연구는 DNS 패킷의 특징을 기반으로 트래픽 분류를 시도한 연구와 DNS 패킷의 정보를 실시간 시각화 처리하는 모델 연구, DNS 패킷을 활용한 악성 행위 탐지에 대한 연구로 분류된다. 트래픽 분류와 관련된 연구로는 DNS 패킷 특징을 이용하여 빠르게 IP 흐름을 분류하는 알고리즘

[1] TNS : Traffic name space

기능 \ 제품		C사	P사	N사	B사	B사	M사	I사	W사	S사	A사	L사	D사	T사	E사	F사	TNS
구동 환경	설치 위치	디바이스 & 네트워크						디바이스			네트워크						디바이스 & 네트워크
	크로스 플랫폼	X	X	O	O	O	X	O	O	X	X	X	O	X	O	X	O
	웹인터페이스 ,GUI	O	O	O	O	O	O	X	O	O	O	O	O	O	O	O	O
기능	PCAP 지원	X	X	O	X	O	X	X	O	X	X	O	X	X	X	X	O
	IP TO IP 트래픽 측정	O	O	O	O	O	X	O	O	O	O	O	O	X	O	O	O
	L7 추적	O	O	O	X	O	X	O	O	O	O	X	X	X	O	X	△
	도메인 별 트래픽 추적	O	X	X	X	X	X	O	X	O	X	X	X	X	O	X	O
	프로세스 별 트래픽 추적	X	O	O	X	X	X	X	X	O	X	X	X	X	O	X	O
	DNS 기반 패킷 분석	O	X	X	X	X	X	X	X	X	X	X	X	X	X	X	○
	통계 및 시각화	O	O	O	O	O	O	X	△	O	O	O	O	O	O	O	O
	사용자별 트래픽 추적	X	X	X	X	X	X	X	X	O	X	X	X	X	X	X	O

<표 1> 벤치마킹 비교표[2]

연구[1]가 있으며 DNS 패킷에서 서비스의 IP 주소와 이름을 통해 헤더 시그니처 생성을 고안한 연구[2]가 있다. 이러한 연구에서는 DNS 패킷을 네트워크 트래픽 분류의 새로운 기준으로 제안한다. 이를 기반으로 DNS 응답 패킷에서 추출한 서비스 IP 와 도메인 이름을 시그니처로 생성하여 트래픽 분류를 실험[3]하였으나 전체 분석률은 29.08%에 불과하였다. 트래픽 분류를 네 가지 방법으로 구체화하여 진행한 연구[4]에서는 DNS 패킷 정보와 호스트 프로파일링을 통해 추출한 정보를 결합하여 트래픽 분류 성능의 개선을 제안하였으나, 가장 성능이 좋았던 모델도 약 10%의 패킷 분류에 실패하였다.

DNS 질의 정보를 웹에서 실시간 시각화 처리하는 모델 구현에 대한 연구[5, 6]에서는 DNS 패킷의 중요성을 강조하여 DNS 질의 정보를 실시간으로 시각화 하는 모델을 제안한다. 그러나 좋은 시각화 모델임에도 불과하고 DNS 질의 정보만을 시각화 하는 것은 DNS 응답이 오지 않은 DNS 질의 정보도 포함하는것과 동시에 통신하는 패킷의 흐름을 파악하기에 어려움이 있었다.

DNS 패킷 기반 악성 행위 탐지에 대한 연구로는 봇넷에서 주로 발생하는 DNS 질의 정보를 분석하는 봇넷 탐지 기술 연구[7], DNS 트래픽을 활용하여 침해사고 대응을 위한 정보 제공에 대한 연구[8], 최상위 도메인의 DNS 질의 로그를 통한 악성 도메인 탐지[9], 패킷 흐름 정보를 바탕으로 비정상적인 DNS 행위의 탐지 알고리즘 제시[10]가 있다. 이러한 종래의 연구를 근거로 DNS 패킷에서 추출한 정보를 네트워크 트래픽 분류의 기준으로 삼을 수 있으며 악성 행위 탐지를 위한 도구로써 활용 가능성을 확인하였다.

본 연구는 선행 연구의 DNS 프로토콜에 기반한 트래픽 분류 방법과 시그니처 생성 방법을 개선하여 DNS 응답 패킷 분석을 통해 네트워크 계층에서 서비스별 트래픽 통계를 제공하고 동시에 사용자 디바이스에 분석을 위한 프로그램을 설치하여 사용자의 네트워크 행위를 상세히 분석함으로써 구체적인 트래픽 분류에 기여하고, 분류 기준을 다양화한다.

2.2 현황 분석

2.2.1. 선별 개요

본 연구에서는 기존의 네트워크 모니터링 제품들을 기능별로 분류하였으며 그 내용은 표<1> 과 같다.

2.2.2. 항목 정의

<표 1> 에서 기존의 제품들을 비교하기 위해 아래와 같은 지표기준을 만들었고 그 설명은 다음과 같다.

1. 설치 위치는 시스템이 어느 위치에 설치되어 있는지를 나타내며, 어느 계층에서 발생한 트래픽을 분류하는지 확인한다.
2. 크로스 플랫폼을 통해 제품의 유동적인 사용이 가능한지 알아본다.
3. 웹 인터페이스 및 GUI 항목을 통해 통계 및 시각화 측면에서 사용자 편의를 고려하였는지 확인한다.
4. pcap 지원은 pcap 파일의 분석 가능 여부를 평가한다. pcap 라이브러리를 통해 라이브로 들어오는 패킷과, 수집하여 저장해둔 파일 형태의 패킷을 분석할 수 있다.
5. IP to IP 트래픽 측정을 통해 IP 간의 트래픽 송수신 정보를 확인하여 패킷 흐름 별 통신 양을 확인한다.
6. L7 추적을 통해 L7 기반의 DNS 모니터링 지원 여부를 확인하며, 페이로드 분석을 통한 트래픽 필터링의 가능 여부를 확인한다.
7. 도메인 별 트래픽 추적을 통해 도메인 별 트래픽 측정 가능 여부를 확인한다.
8. 프로세스별 트래픽 분석을 통해 받아온 패킷이

[2] 벤치 마킹 비교표

어느 프로세스에 해당하는 지의 여부를 알아봄으로써 프로세스 별 트래픽 양을 확인한다.

9. DNS 기반 패킷 분석은 DNS 패킷에 저장된 정보를 기반으로 분석하고 있는지 알아보는 항목이다.
10. 통계 및 시각화항목에서 패킷 분석 결과를 통계 및 시각화하여 제공 여부를 확인한다.
11. 사용자 별 트래픽 추가항목에서 발생한 패킷이 동일한 네트워크에 존재하는 어느 사용자의 트래픽인지 판별한다.

2.3 연구 제안

위에서 소개한 장비 및 연구들은 각각의 장단점이 존재하지만 공통적으로는 사용자별 트래픽 모니터링과 네트워크 계층의 트래픽을 이용한 도메인 분류 기능을 지원하지 않았다. 서비스 관리자 입장에서 사용자 별 트래픽 모니터링과 도메인 분류가 되지 않으면 악성코드를 유발하는 프로세스에 대한 분석이 불가능하며 추가 분석이 요구된다.

본 연구에서는 이러한 문제점을 보완함으로써 DNS 패킷을 기반으로 네트워크 흐름 통계 정보를 시각화하고 표현한 트래픽 모니터링 시스템을 제안한다.

3. 시스템 요구사항

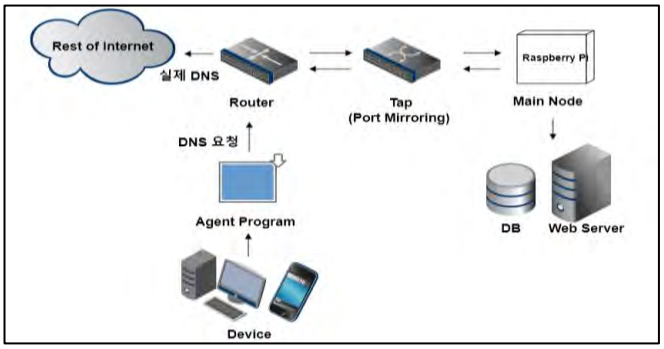
3.1 시스템 개요

본 시스템은 라우터로부터 패킷을 미러링하여 도메인 네임 라벨링을 거쳐 데이터베이스로 전달하고 웹 인터페이스를 통해 시각화하는 시스템이다.

사용자는 디바이스에 패킷 분석 프로그램을 선택적으로 설치하고 웹에서는 디바이스 프로그램에서 전달받은 정보와 네트워크 계층 트래픽 분석 프로그램에서 전달받은 정보를 결합한다. 패킷 미러링을 통해 패킷 데이터를 전달받는 모듈, 전달받은 패킷의 흐름 정보를 추출하는 모듈, DNS 패킷 데이터를 관리하는 모듈로 구성된 네트워크 계층 트래픽 분석 프로그램, 패킷 흐름 추출 모듈, 프로세스 정보 수집 모듈, 프로세스 별 패킷 분류 모듈로 구성된 사용자 계층 트래픽 분석 프로그램, 데이터 베이스에 저장된 데이터를 기반으로 통계 정보를 생성 및 시각화하는 웹 모듈로 구성된 웹 인터페이스 프로그램으로 이루어져 있다. 다음 <그림 1>은 본 연구에서 제안하는 시스템의 구성도를 나타낸다.

3.1.1.네트워크 계층 트래픽 분석 프로그램

패킷 미러링 및 입력 모듈에서는 라우터를 통하는 모든 패킷을 복제하여 패킷 흐름 추출 모듈로 전달한다. 패킷 흐름 추출 모듈에서 전달받은 패킷의



<그림 1> 시스템 구성도

흐름 정보를 확인하여 패킷을 구분하고, DNS 관리 모듈로 전달한다. DNS 관리 모듈에서는 DNS 응답 패킷이 확인될 경우, 시그니처 생성에 필요한 서비스 네임, 서비스 IP 와 같은 데이터를 추출하고 저장한다. 전달받은 흐름 정보와 도메인 네임 정보를 연결한다.

3.1.2. 사용자 계층 트래픽 분석 프로그램

패킷 흐름 추출 모듈에서는 전달받은 패킷에서 흐름 정보를 추출하고 프로세스 별 패킷 분류 모듈로 전달한다. 프로세스 정보 수집 모듈에서는 디바이스에서 실행중인 네트워크 관련 프로세스의 정보를 수집한다. 프로세스 별 패킷 분류 모듈에서는 프로세스 정보 수집 모듈에서 수집한 정보를 바탕으로 패킷을 분류한다. 이후 사용자가 PC, 모바일 환경에서 웹 대시보드에 접근하여 트래픽 분석 정보를 조회 가능하다.

3.1.3. 웹 시각화 프로그램

웹 시각화 모듈에서는 네트워크 계층과 사용자 계층에서 전처리 되어 데이터베이스에 저장되어 있는 트래픽 데이터를 활용하여 통계 정보를 작성하고 시각화 한다.

3.2 시스템의 구현

3.2.1 네트워크 계층 트래픽 분석 프로그램

패킷의 실시간 캡쳐와 정적 파일의 분석을 모두 지원하기 위해 pcap 라이브러리를 기반으로 개발한다. 입력 받은 패킷으로부터 출발지 IP, 목적지 IP, 출발지 포트와 목적지 포트 정보를 통해 흐름 정보를 생성하고, 패킷이 발생한 시간 정보를 함께 기록한다.

입력 받은 패킷이 DNS 응답 패킷일 경우 DNS 를 요청한 호스트의 IP 와 포트 정보 및 서비스 IP 와 도메인 이름 정보, TTL 정보, DNS 응답 패킷의 시간 정보를 저장한다. 패킷 흐름 추출 모듈을 통해 생성한 흐름 정보와 시간 정보 및 DNS 정보의 TTL 을 비교하고 유효성을 확인하여 도메인 이름 정보와 매칭시킨다.

하나의 IP 가 여러 개의 도메인을 가진 경우, IP 만을 키로 활용하여 도메인 네임과 매핑 시키기 어렵다는 문제점이 발생한다. DNS 응답 패킷이 발생한 직후 동일한 IP 에 패킷을 보내는 디바이스 IP 와 포트를 해당 DNS 응답 패킷과 매칭시키는 방법으로 해결한다.

3.2.2 사용자 계층 트래픽 분석 프로그램

DNS 패킷을 발생시키지 않고 직접적으로 통신하는 연결 패킷은 네트워크 계층에서 분류가 어렵다. 때문에 디바이스 계층에 프로그램을 설치하여 패킷의 정보를 제공하고 프로세스 별 사용 패킷 분류 및 사용자의 네트워크 행위정보를 수집한다. 이 정보들을 데이터베이스에 업로드 한 후 네트워크 계층에서 수집한 패킷 데이터와 연관 지어 추가적인 분석을 진행한다.

어떤 프로세스에서 발생한 패킷인지 확인하기 위한 방법은 다음과 같다. 먼저 디바이스 내에서 송수신 되고 있는 패킷에서 포트 번호를 확인한다. 확인된 포트 정보를 바탕으로, 연결이 유지되는 동안 포트 번호를 확인한다. 이를 통하여 포트 번호가 동일한 프로세스를 해당 패킷이 발생한 프로세스로 간주한다.

3.2.3. 데이터베이스 및 웹 시각화 프로그램

본 연구에서 제안한 시스템에서는 대량의 네트워크 트래픽 데이터를 흐름 정보 테이블과 DNS 정보 테이블을 이용하여 관리한다. 이러한 데이터의 특징을 고려하여 NoSQL 구조를 가지고 있는 아파치 카산드라를 데이터베이스 시스템으로 제안한다. 본 연구에서 제안한 시스템에서는 최적화된 대량의 트래픽 처리를 위하여 네트워크 계층 트래픽 분석 프로그램과 사용자 계층 트래픽 분석 프로그램에서 1초 주기로 데이터베이스를 업데이트한다.

웹 시각화 모듈에서는 파이썬 기반의 웹 어셈블리 프레임워크로 웹 사이트를 개발한다. 카산드라 엔진과 호환성이 높은 웹 프레임워크 Django를 이용할 것을 제안한다.

사용자 계층 트래픽 분석 프로그램을 설치하지 않은 디바이스는 네트워크 계층 트래픽 분석 프로그램에서 제공하는 데이터를 기준으로 서비스 별 또는 네트워크 흐름정보 별 트래픽 사용에 대한 실시간 모니터링 및 네트워크 트래픽 히스토리를 제공하고, 이를 기반으로 통계 정보를 작성하고 시각화 한다. 사용자 계층 트래픽 분석 프로그램을 설치한 경우, 네트워크 계층에서 분석한 통계 정보와 사용자 계층에서 분석한 통계 정보를 병합하기 위해 각 프로그램에서 패킷의 해시 정보들을 추출하여 비교한다. 사용자 계층 트래픽 분석 시스템을 통해 웹에서는 프로세스 별 네트워크 트래픽 히스토리와 실시간 모니터링을 지원하고 사용자의 네트워크 행위 분석 정보를 시각화하는 기능을 제공한다.

4. 결론

사용자별, 프로세스별 인터넷 접속 기록을 모니터링 하고자 하는 기업의 요구 사항은 항상 존재해 왔지만 현존하는 DNS 기반의 모니터링 제품들은 어떠한 도메인으로의 접속이 어떠한 IP 로부터 출발했는지만을 보여주는 한계점이 있다. 이러한 시스템은 한 사용자의 접속기록을 보는 데는 유용하나 근본적인 접속을 유발하는 사용자 프로그램을 분류화 하는 것은 불가능하였다. 우리는 DNS 패킷 분석과 사용자의 프로그램 설치를 통하여 세분화된 접속 기록을모니터링 하는 방법론을 제시하였고 필요에 따라 확장이 가능하며 수집한 정보를 바탕으로 악성링크를 접속하는 악성 프로그램을 탐지하거나 피싱 사이트 및 파밍 사이트에 접속하는 경로를 세부적으로 모니터링 하는 시스템을 제시하였다. 이러한 시스템은 필요에 따라 가변적으로 확장, 축소되는 방식이기에 고정적인 틀에 갇혀 있는 기존 시스템에 대항하여 유기적인 구조를 가지고 각 기업별, 성능별, 필요성에 따라 구성된다는 장점을 가지고 있다.

참고문헌

[1] 이한우, 최현상, 이희조, DNS-based Botnet Detection and Monitoring System, 한국정보처리학회 추계학술발표대회 논문집, 제 13 권, 제 2 호, 2006

[2] 김지혜, 박준상, 오영석, 김명섭, Header Signature Generator System and Traffic Classification System by Service Using DNS Query, 한국통신학회 학술대회논문집, 한국통신학회, 2010, pp. 695-696

[3] 김지혜, 김명섭, Research on Traffic Classification based on DNS Packet Analysis, KNOM Review, 제 13 권, 제 2 호, pp. 36-44, 2010

[4] David Plonka, Paul Barford, Flexible Traffic and Host Profiling via DNS Rendezvous, Workshop Satin, 2011

[5] Pawel Foremski, Christian Callegari, Michele Pagano, DNS-Class: Immediate classification of IP flows using DNS, International Journal of Network Management, 2014

[6] 이슬기, 김병익, 이태진, 박해룡, A Study on the Analy-sis of Cyber Incidents Using DNS Traffic Information, 한국통신학회 학술대회논문집, 한국통신학회, 2014, pp. 64 -65

[7] 이기룡, 이제현, 권종훈, 이희조, 박해룡, Analysis on .KR TLD DNS Query Log for Malicious Behavior Detection, 사이버 공격의 사전 사후 대응을 위한 사이버 블랙박스 및 통합 사이버보안 상황분석 기술 개발, 미래창조과학부 및 정보통신기술진흥센터 정보통신•방송 연구개발사업, 2014

[8] Milan Cermak, Pavel Celeda, Jan Vykopal Detection of DNS Traffic Anomalies in Large Networks, Masaryk University, 2014

[9] 장상동, Realtime Visualization System for DNS Query, 한국정보과학회 학술발표 논문집, 한국정보과학회, 2015, pp. 441-442

[10] 장상동, A RealTime DNS Query Analysis System based On the Web, 디지털융복합연구, 제 13 권, 제 10 호, pp. 279-285, 20