MapReduce 환경에서 Snort 로그를 이용한 실시간 네트워크 공격패턴 분석 시스템†

강문환[1], 장진수[2], 신영성[1], 장재우[2*]
전북대학교 컴퓨터공학과[1]
전북대학교 IT정보공학과[2]
{calvin, lklin1749, twotoma, jwchang}@jbnu.ac.kr
*Corresponding author

Real-time Network Attack Pattern Analysis System using Snort Log on MapReduce Environment

Moon-Hwan Kang, Jin-Su Jang, Young-Sung Shin, Jae-Woo Chang
Dept of Computer Engineering, Chonbuk University

요 약

최근 급격히 증가하고 있는 네트워크 로그 상에서 보안위협에 신속히 대응하기 위해 기업들은 방화벽, IDS 등의 네트워크 보안 로그를 분석하여 보안 위협을 파악한다. Snort는 이러한 보안 위협에 대응하기 위해 네트워크 로그를 수집하는 도구 중 하나이다. 그러나 보안 관제 담당자는 방대한 양의 보안 관련 로그를 분석하기 위해 많은 시간이 필요하기 때문에, 관제 결과를 보고하고 대응하기까지 시간이 지체되는 문제가 존재한다. 이러한 문제를 해결하기 위해, 본 논문에서는 Snort 로그를 이용한 실시간 네트워크 공격패턴 분석 시스템을 제안한다. 제안하는 시스템은 대용량 데이터 처리에 효과적인 MapReduce 분산 처리를 활용하여 방대한 네트워크 로그를 추출 및 분석하기 때문에 보안 위협 상황 발생 여부를 실시간으로 빠르게 인지할 수 있다.

1. 서론

스마트폰, 사물인터넷 등 IT 기기들의 발달로 네트워크 트래픽의 양이 급격히 증가하고 있다. 트래픽양의 증가는 곧 네트워크 보안 로그의 증가로 이어지기 때문에 기업에서는 의심스러운 네트워크 패킷들을 수집 및 분석하기 위해 네트워크 침입 탐지 시스템(Network Intrusion Detection System)을 활용하고 있다. 하지만 방대한 양의 네트워크 보안 로그에서 보안 위협 정보를 빠르게 탐지하고 분석하여 대응하는 일은 쉽지 않다. 또한 방대한 양의 로그를 분석하는 것은 많은 제약이 따르기 때문에, 최근에는 네트워크 공격 패턴 분석에 필요한 필수적인 로그만을 추출하고 최소한의 연산으로 보안 위협을 인지하는데 초점을 맞추고 있다. 네트워크를 이용해 접근하는 보안 위협들을 효과적으로 방어하기 위해서는 네트워크 침입 탐지 시스템이 필수적이며, 시스템의 활용도를 향상시키기 위한 다수의 연구가 수행되고 있다. 가장 널리 사용되는 침입 탐지 시스템으로는 Snort가 존재한다. Snort 는 침입 탐지 시스템뿐만 아니라 일반적으로 알려진 보안 위협들에 대해 침입 탐지 규칙들을 제공한다. Snort는 실시간으로 발생하는 네트워크 로그들을 텍스트 형태로 기록하기 때문에 분석하는데 많은 시간이 소요된다. 또한 정상적인 로그들이 포함되어 있기 때문에 네트워크 공격 패턴을 분석하는 것은 매우 어렵다. 이러한 문제를 해결하기 위해, 본 논문에서는 Snort 로그를 이용한 실시간 네트워크 공격패턴 분석 시스템을 제안한다. 제안하는 시스템은 텍스트 형태의 네트워크 로그를 빠른 시간 내에 분석하기 위해 스트림 데이터를 분산 처리할 수 있는 Hadoop MapReduce를 사용한다. 이를 통해 Snort를 이용하여 발생한 로그를 실시간으로 수집하고, Hadoop MapReduce를 통해 수많은 네트워크 로그들 중 분석에 필요한 최소한의 로그를 빠르게 추출할 수 있다. 아울러, 추출한 정보를 토대로 네트워크 공격 패턴을 분석하여 사용자에게 네트워크 공격에 대한 분석 결과를 제공한다. 제안하는 시스템은 다음과 같은 특성을 지닌다.

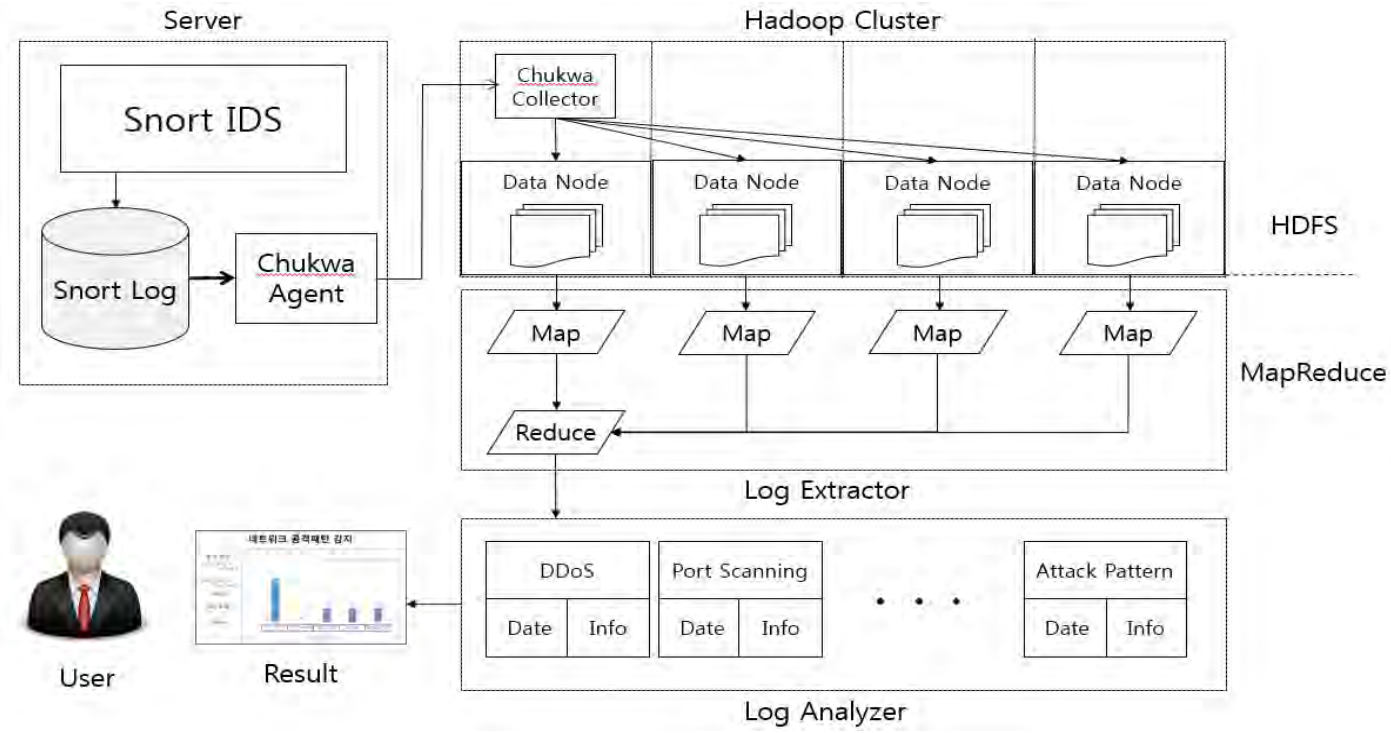
- 대용량 네트워크 보안 로그 추출 : 대용량 네트워크 로그 상에서 보안 위협을 탐지하기 위해 필수적인 요소를 추출한다.
- 실시간 네트워크 공격 패턴 분석 : 실시간 수집되는 보안 로그를 분석하여 보안에 위협이 되는 공격을 탐지한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 기술하고, 3장에서는 실시간 네트워크 공격패턴 분석 시스템을 기술한다. 4장에서는 결론 및 향후 연구에 대해 기술한다.

2. 관련 연구

기존 네트워크 보안 로그 분석에 관한 대표적인 연구

† 이 논문은 2016년 교육부와 한국연구재단의 지역혁신창의인력양성사업의 지원을 받아 수행된 연구임(NRF- 2016H1C1A1065816). 또한 이 논문은 미래창조과학부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음(IITP-2016-R0113-15-0005).



(그림 1) 실시간 네트워크 공격패턴 분석 시스템 구조

는 다음과 같다. Cheon, JeongJin et al.의 연구[1]는 대용량 네트워크 보안 로그에 대해 MapReduce[5] 분산 처리를 활용한 패킷 분석 시스템을 제안하였다. 그러나 해당 연구는 Snort[6]에서 제공하는 보안 Rule의 예외적인 공격 패턴들을 감지하지 못한다. Lee, Yeonhee et al.의 연구[2]는 Tcpdump를 통해 네트워크상에서 송수신되는 패킷들을 수집하여 다양한 분석을 수행하였으나 앞서 언급한 연구와 같이 예외적인 공격 패턴 감지를 수행하지 못하는 문제점이 있다. 한편, Prathibha, P. G. et al.의 연구[3]은 네트워크 보안 로그를 Hive[7]에 삽입하여 다양한 질의처리를 지원한다. 또한 기본적인 공격 패턴 감지를 수행하여 문제가 되는 출발지 정보를 침입 탐지 시스템의 Rule에 블랙리스트로 추가한다. 한편, 보안 로그 상에서 공격패턴 감지를 지원하는 대부분 연구들은 시각화를 중심으로 제안되었으며 대표적인 연구로는 Lee, Dong-Gun et al.의 연구[4]가 존재한다. 해당 기법은 기본적인 보안 위협을 탐지하기 위해 3가지 요소로 Source IP, Destination IP, Destination Port를 유지하며 각각의 비율에 대해 RGB색을 이용한 시각화를 제공한다. 앞서 언급한 모든 연구들은 네트워크 보안 로그 분석을 제공하고 있지만 실시간 네트워크상에서 보안 위협을 탐지하지 못하는 문제를 지닌다.

3. 실시간 네트워크 공격패턴 분석 시스템

본 장에서는 전체적인 시스템 구조 및 실시간 대용량 네트워크 로그의 추출 작업에 대해 제시하고, 이를 기반으로 보안 위협을 감지할 수 있는 공격패턴 분석 알고리즘을 제안한다.

3.1 시스템 구조

제안하는 전체 시스템 구조는 <그림 1>과 같다. 이는 보안 로그를 수집하는 Snort 기반 침입 탐지 시스템, 해당 로그를 HDFS로 삽입시키는 Chukwa 시스템, 분석에 필요한 로그를 추출하는 로그 추출기와 이를 분석하여 보안 위협을 탐지하는 로그 분석기로 구성되어 있다. Snort 로그 수집 모듈에서 실시간으로 수집된 네트워크 로그는 비정형 텍스트 형태로 다양한 속성의 네트워크 정보를 포함하고 있다. Chukwa 시스템은 수집된 로그를 Collector를 통해 10분 단위로 HDFS에 삽입한다. HDFS에 저장된 비정형 데이터는 로그 추출기를 통해 분석에 필요한 데이터로 변환되고 관제 요원의 요청에 의해 로그 분석기가 실행된다. 로그 분석기는 사용자에게 주요 공격 패턴에 대한 탐지결과와 상세정보를 제공한다.

3.2 로그 추출기

로그 추출기는 비정형 형태로 수집된 로그를 정형 데이터로 변환하는 작업을 수행하며, 이 때 분석에 필요한 속성(Time, Source IP, Destination IP, Destination Port) 4가지를 추출한다. 4가지만 추출하는 이유는 방대한 양의 로그를 분석하는 과정에서 불필요한 정보는 연산 부하를 증가시켜 빠른 보안 위협 탐지에 치명적인 결과를 불러오기 때문이다. 로그 추출기는 전체 시스템에서 가장 많은 작업량과 시간을 차지하는 컴포넌트이다. 따라서 대용량 로그의 수집 및 분석결과의 시간 격차를 최소화하기 위해 분산 병렬 컴퓨팅 기술인 MapReduce 활용한다. Chukwa 시스템을 통해 단위시간 10분 간격으로 축적된 보안 로그는 MapReduce를 통하여 Time 속성을 key 값으로 추출된다. 1차 추출이 완료된 로그는 4가지 속성을 유지하는 정형 데이터로 HDFS에 저장된다. 저장된 정형 데이터는 단위시간 당 발생한 Source IP, Destination IP, Destination Port의 종류와 카운트를 유지한다.

3.3 로그 분석기

제안하는 로그 분석기는 방화벽, IDS 로그 분석을 통해 보안 관제원이 위협 상황을 빠르게 인지할 수 있도록 한다. 여기서는 본 연구에서 초점을 맞추고 있는 보안 위협의 유형에 대해 알아본다.

3.3.1 호스트 스캐닝

호스트 스캐닝 공격은 평상시보다 접근하는 목적지가 다양해졌음을 의미하기 때문에, 방화벽 로그 상에서 단위 시간동안 Destination IP의 수가 Source IP와 Destination Port수에 비해 상대적으로 급증하게 된다. 이를 이용하여 추출된 로그를 식(1)에 적용할 경우 호스트 스캐닝에 대한 확률적인 탐지 결과가 도출된다.

$$HS_{prob} = D_{IP}/(S_{IP}+D_{IP}+D_{Port}) \quad \text{-------- (1)}$$

3.3.2 포트 스캐닝

포트 스캐닝 공격은 접근해온 목적지 상에서 열린 포트에 대해 전역적인 탐색을 시도하기 때문에, 방화벽 로그상에서 단위 시간동안 Destination Port의 수가 Source IP 및 Destination IP에 비해 상대적으로 급증하게 된다. 호스트 스캐닝과 마찬가지로 위 조건을 이용하여 추출된 로그를 식(2)에 적용할 경우, 포트 스캐닝에 대한 확률적인 탐지 결과가 도출된다.

$$PS_{prob} = D_{Port}/(S_{IP}+D_{IP}+D_{Port}) \quad \text{-------- (2)}$$

3.3.3 네트워크 장애 발생

네트워크상에서 실시간 로그를 수집하면 일반적으로 보안 위협의 요소를 제외하고도 끊임없이 방대한 양의 로그가 축적된다. 그러나 방화벽 또는 네트워크 스위치 장애가 발생하면 로그 발생량이 급격히 감소한다. 식(3)의 조건은 최소한의 로그 발생을 확인하여 네트워크 장애를 탐지한다. 이 때 공격에 의한 네트워크 장애가 발생할 경우, 장애 발생 직전의 로그 양이 평균치보다 매우 높은 것을 확인할 수 있다. 식(3)의 임계값 Threshold_a 는 추출한 모든 요소 중 단위시간 당 발생한 최소값을 기준으로 한다.

$$\text{if}(\sum Log < Threshold_a) \rightarrow Fail \quad \text{-------- (3)}$$

3.3.4 DDoS 공격

DDoS 가 발생한 경우 방화벽 로그를 관찰해보면 무차별적인 로그 증가가 발생한다. 이때 Destination IP 및 Destination Port에 비해 상대적으로 Source IP가 매우 높은 값을 갖는다. 이는 식(4), 식(5)를 통해 DDoS 공격을 의심할 수 있다.

$$DDoS_{1prob} = (\sum Log - S_{IPavg})/(\sum Log_{avg} - S_{IPavg}) \quad \text{-------- (4)}$$

$$DDoS_{2prob} = S_{IP}/(S_{IP}+D_{IP}+D_{Port}) \quad \text{------- (5)}$$

3.3.5 바이러스 침투 혹은 웜 활동 탐지

웜 또는 바이러스 침투는 목적지의 활동량 증가로 이어진다. 또한 바이러스의 확산 시도 및 웜의 활동량 증가는 주기적으로 일어나는 경우가 많기 때문에 Destination IP 및 Destination Port의 증가가 발생하는 주기를 측정할 필요가 있다. 이를 위해 추가적인 네트워크 조사가 필요하며, 이는 식(6) 을 통해 측정할 수 있다.

$$Virus_{prob} = (D_{IP}+D_{Port})/(S_{IP}+D_{IP}+D_{Port}) \text{ --(6)}$$

4. 결론

시스템 서버 및 네트워크 관제 요원은 빠른 시간 내에 보안 위협상황 발생을 인지시키는 것이 목적이나, 위협적인 출발지에 대해 완벽한 대응을 제공하지 못하는 한계가 존재한다. 이러한 문제점을 해결하기 위하여, 본 논문에서는 Snort 로그를 이용한 실시간 네트워크 공격패턴 분석 시스템을 제안하였다. 제안하는 시스템은 첫째, 대용량 네트워크 보안 로그 추출을 통해 대용량 네트워크 로그 상에서 보안 위협을 탐지하기 위해 필수적인 요소를 추출한다. 둘째, 실시간 네트워크 공격 패턴 분석을 통해 실시간 수집되는 보안 로그를 분석하여 보안에 위협이 되는 공격을 탐지한다. 향후 연구는 제안하는 침입 탐지 시스템을 실제로 구현하여 성능평가를 실시하고, 지능적인 분석을 추가하여 실제로 관제 요원을 대체할 수 있는 시스템을 개발하는 것이다.

참고문헌

[1] Cheon, JeongJin, and Tae-Young Choe. "Distributed processing of snort alert log using hadoop." International Journal of Engineering and Technology 5.3 (2013): 2685-2690.

[2] Lee, Yeonhee, Wonchul Kang, and Youngseok Lee. "A hadoop-based packet trace processing tool." International Workshop on Traffic Monitoring and Analysis. Springer Berlin Heidelberg, 2011.

[3] Prathibha, P. G., and E. D. Dileesh. "Design of a hybrid intrusion detection system using snort and hadoop." International Journal of Computer Applications 73.10 (2013).

[4] Lee, Dong-Gun, Huy Kang Kim, and Eunjin Kim. "Study on security log visualization and security threat detection using RGB Palette." Journal of the Korea Institute of Information Security and Cryptology 25.1 (2015): 61-73.

[5] Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." Communications of the ACM 51.1 (2008): 107-113.

[6] Roesch, Martin. "Snort: Lightweight intrusion detection for networks." Lisa. Vol. 99. No. 1. 1999.

[7] Thusoo, Ashish, et al. "Hive: a warehousing solution over a map-reduce framework." Proceedings of the VLDB Endowment 2.2 (2009): 1626-1629.