온라인 데이터 수집 기반 실시간 비정상 행위 탐지

이명철*, 김창수*, 김익균**
*한국전자통신연구원 빅데이터인텔리전스연구부
**한국전자통신연구원 정보보호연구본부
e-mail : {mclee, cskim7, ikkim21}@etri.re.kr

Real-time Abnormal Behavior Detection by Online Data Collection

Myungcheol Lee*, ChangSoo Kim*, Ikkyun Kim**
*Big Data Intelligence Research Department, ETRI
**Information Security Research Department, ETRI

요 약

APT (Advanced Persistent Threat) 공격 사례가 증가하면서, 이러한 APT 공격을 해결하고자 이상 행위 탐지 기술 관련 연구가 활발히 진행되고 있다. 최근에는 APT 공격의 탐지율을 높이기 위해서 빅데이터 기술을 활용하여 다양한 소스로부터 대규모 데이터를 수집하여 실시간 분석하는 연구들이 시도되고 있다. 본 논문은 빅데이터 기술을 활용하여 기존 시스템들의 실시간 처리 및 분석 한계를 극복하기 위한 실시간 비정상 행위 탐지 시스템에서, 파일 시스템에 수집된 오프라인 데이터 기반이 아닌 온라인 수집 데이터 기반으로 실시간 비정상 행위를 탐지하여 실시간성을 제고하고 입출력 병목 문제로 인한 처리 성능 확장성 문제를 해결하는 방법 및 시스템에 대해서 제안한다.

1. 서론

최근, APT (Advanced Persistent Threat) 공격 사례가 증가하면서, 이러한 APT 공격을 해결하고자 이상 행위 탐지 기술 관련 연구가 활발히 진행되고 있다. 장기간 잠복하며 제로데이 취약점을 이용하고, 새로운 또는 변형된 악성 코드를 일관되게 사용하는 APT 공격의 탐지율을 높이기 위해서는 다양한 소스로부터 장기간에 걸쳐 대규모 데이터를 수집, 처리 및 분석하는 기술과, 데이터를 수집 즉시 실시간 분석하는 기술, 그리고 개별 공격들 간의 상관(correlation) 분석 기술이 동시에 요구되나, 기존 보안 시스템들은 이러한 복잡한 분석 능력이나 컴퓨팅 파워, 신속성 등이 부족한 것이 현실이다.

본 논문에서는 기존 시스템들의 실시간 처리 및 분석 한계를 극복하기 위해, 온라인 수집 데이터 기반 실시간 비정상 행위 탐지 방법 및 시스템을 제안한다.

2. 시스템 구조

본 논문에서 제안하는 온라인 수집 데이터 기반 실시간 비정상 행위 탐지 시스템은 사이버 타겟 공격 대응 시스템인 SINBAPT (Security Intelligence techNology for Blocking APT) 시스템의 일부로 구현되었다[1].

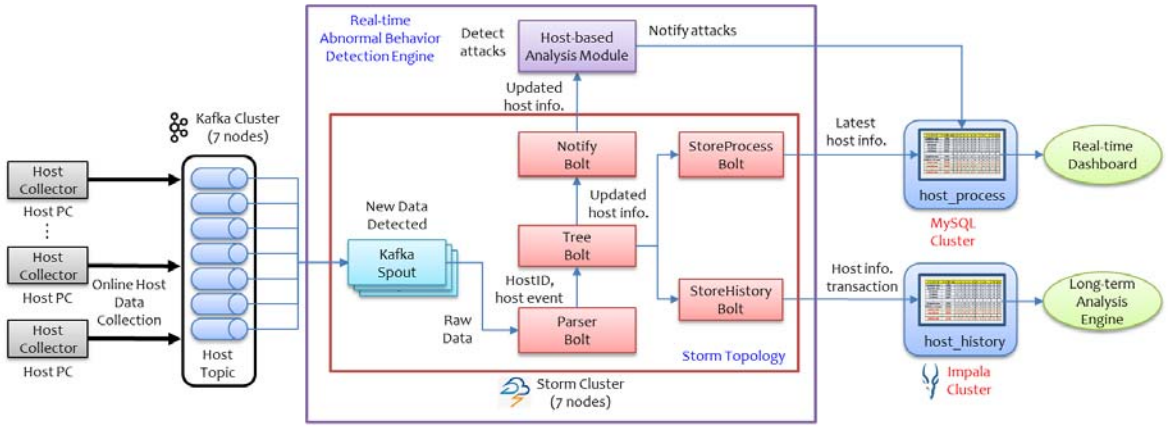
본 논문은 SINBAPT 시스템의 서브시스템으로서 내부 호스트 PC 에서 발생하는 비정상 행위를 탐지하기 위한 실시간 비정상 행위 탐지 시스템[2]을 파일 시스템에 수집된 오프라인 데이터가 아니라 온라인 수집 데이터 기반으로 탐지하여 실시간성을 개선하도록 확장한 내용에 대해서 설명한다.

본 논문의 실시간 비정상 행위 탐지 시스템은 호스트 PC 에 설치된 호스트 이벤트 수집기를 통해 호스트에서 발생하는 모든 프로세스의 행위 이벤트 데이터를 탐지해서 Kafka 클러스터의 호스트 토픽에 수집한다. 수집된 프로세스 행위 이벤트 데이터는 Kafka Spout 를 통해서 Storm 클러스터에 전달되고, Parser Bolt 에 의해 처리가 시작된다.

Parser Bolt 는 바이너리 형태로 표현된 프로세스 행위 이벤트 데이터를 구문 분석하여 프로세스 행위 이벤트 정보를 추출하고, 추출된 프로세스 행위 이벤트 정보는 변경 이력 관리를 위해 Tree Bolt 로 전달된다.

Tree Bolt 는 프로세스 행위 상태 및 이력을 관리하고 추적하기 위해서, 프로세스 행위 특성 인자 벡터를 포함하는 프로세스 행위 정보 트리를 관리한다. 만일 신규로 발생한 이벤트가 이미 트리에 등록된 프로세스의 이벤트라면, 해당 프로세스의 특성 인자 벡터가 신규 이벤트 정보를 포함하도록 갱신된다.

모든 특징 벡터 트리에 대한 변경 사항은 Notify Bolt, StoreProcess Bolt, StoreHistory Bolt 에 통지되며, Notify Bolt 는 TCP 서버로 동작하는 호스트 기반 분석 모듈에 특징 벡터 변경 사항을 통지한다. StoreProcess Bolt 는 최신 프로세스 상태 정보를 MySQL 클러스터에 저장한다. StoreHistory Bolt 는 모든 프로세스 행위 이벤트의 트랜잭션을 Apache Impala 에 저장하여 보다 진보하고 복잡한 이벤트 간의 상호 관계 기반 비정상 행위 탐지를 위한 장기적인 분석에 사용한다.



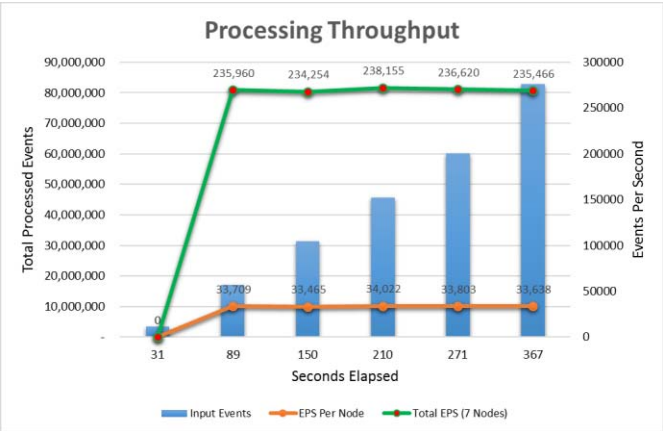
(그림 1) 시스템 구조

호스트 기반 분석 모듈은 Notify Bolt 로부터 통지받은 변경 사항들에 대해 C4.5 결정 트리 알고리즘을 적용한 상관 분석을 수행하며, 공격 여부를 판단하여, 공격인 경우 MySQL 클러스터의 host_process 테이블에 해당 프로세스가 "공격" 상태라고 표시한다.

3. 성능 평가

본 논문에서 제안하는 온라인 데이터 수집 기반 실시간 비정상 행위 탐지 시스템의 성능 시험은 (그림 1)에서 보는 바와 같이 7 대의 노드(Intel Xeon E5-2620 @ 2.00GHz 12-core CPU, 92GB RAM, 1TB HDD)로 구성된 시험 환경에서 수행하였다.

성능 시험을 위한 데이터는 2 개 호스트에서 정상 및 비정상 상황을 가상으로 만든 상태에서 자체 제작한 호스트 이벤트 수집기를 이용하여 프로세스의 모든 행위를 Kafka 클러스터에 수집하였다. 수집한 호스트 프로세스 행위 이벤트 데이터는 (그림 2) 에서 보는 바와 같이 전체 크기는 약 41GB 이고, 약 8,000 만개의 호스트 프로세스 행위 이벤트를 갖고 있다.



(그림 2) 처리 성능

성능 시험은 Storm 토폴로지가 배치되는 약 30 초 정도의 초기 수행 시간은 배제하고 31 초 이후의 처리 이벤트 개수를 약 60 초 간격으로 측정하였다. 성능 시험 결과, 노드당 약 33,000 EPS(Events Per Second), 7 노드 전체로는 약 230,000 EPS 의 성능을 보였으며, 전체 약 8,000 만개의 이벤트 처리에는 초기 시작 시간 30 초를 포함하여 총 367 초의 시간이 소요되었다.

4. 결론

본 논문에서는 APT 공격과 같은 사이버 타겟 공격을 사전에 탐지하기 위한 온라인 수집 데이터 기반의 실시간 비정상 행위 탐지 시스템을 제안한다.

제안 시스템은 기존[2]에 파일 시스템 기반으로 오프라인 수집한 데이터를 주기적으로 처리 및 분석하면서 생기는 분석 지연 문제와 파일 시스템에 저장된 데이터를 여러 파서 및 처리기에서 동시에 읽으며 발생하는 입출력 병목 문제로 인해 처리 노드 수를 증가시켜도 전체 처리 성능이 증대되지 않는 확장성 문제를 해결하기 위하여, 다수의 노드로 구성되는 Kafka 클러스터를 이용하여 온라인 수집된 데이터를 처리 및 분석하는 방법을 제안한다.

성능 시험을 통해 프로세스 행위 이벤트 데이터 초당 처리 성능(throughput)을 측정한 결과, 기존 2 노드로 측정한 클러스터 전체 약 80,000 EPS(프로세스 행위 이벤트/초)의 처리 성능[2]을 넘어서는 클러스터 전체 약 230,000 EPS 의 성능을 확인할 수 있었다.

Acknowledgement

본 논문은 2016 년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. (B0101-15-1293, 다중소스 데이터의 Long-term History 분석 기반 사이버 표적공격 인지 및 추적 기술 개발)

참고문헌

[1] Hyunjoo Kim, Ikkyun Kim, and Tai-Myoung Chung, "Abnormal behavior detection technique based on big data," Lecture Notes in Electrical Engineering, vol. 301, pp. 553-563, Apr. 2014.

[2] 이명철, 문대성, 김익균, "패스트 데이터 기반 실시간 비정상 행위 탐지 시스템," 정보보호학회논문지, 제 25 권 제 5 호, pp. 1027-1041, 2015.10