한국컴퓨터정보학회 동계학술대회 논문집 제24권 제1호 (2016. 1)

Hadoop 기반의 대용량 데이터 보안 시스템에 관한 연구

김효남*
*청강문화산업대학교 게임전공
e-mail: hnkim@ck.ac.kr*

A Study on the Massive Data Security System of the Hadoop Based

Hyo-Nam Kim*
*Dept. of Computer Game, ChungKang College of Culture Industries

● 요 약 ●

현재 스마트 시대에 살고 있는 우리는 매우 복잡하고 거미줄처럼 연결되어 있는 빅 데이터 환경에서 살고 있다. 이런 환경에서는 대용량 데이터를 효율적으로 관리하고 활용하는 것이 개인이나 기업들이 추구하려는 목표이다. 빅 데이터 시대에 데이터의 효율적인 관리와 활용을 위해 다양한 장비에서 수집되고 저장된 대용량 데이터에 대해서 일반적인 데이터 분석을 통한 보안 기술로는 상당한 시간과 자원 낭비가 수반된다. 이를 개선하기 위해 본 논문에서는 하둡을 이용하여 대용량 데이터에 대한 처리 및 분석을 통해 효과적인 보안 시스템을 제안한다.

키워드: Massive Data, Hadoop, Security

I. Introduction

현재 우리가 생활하고 있는 환경은 스마트 기기 기술과 영상·음성·웹 페이지 방문 고객에 대한 클릭 스트림 등 다양한 데이터들을 처리하기 위한 기술이 주류를 이루고 있다. 특히 빅데이터는 대개 전통적인 데이터베이스나 시스템 환경에서 처리하기 힘든 대용량 데이터를 저장·분석·처리를 통해 가치 있는 정보로 만들어낸다. 이러한 대용량 데이터를 처리하는 데에는 바로 '하둡(Hadoop)'을 이용한다. 이에 하둡은 빅데이터를 다루는 개발자들의 많은 관심을 받고 있다[1].

하둡(Hadoop : High Availability Distributed Object-Oriented Platform)은 오픈 소스 기반 분산 컴퓨팅 플랫폼으로 여러 개의 작은 컴퓨터를 마치 하나인 것처럼 묶어 대용량 데이터 처리를 위한 기술이며, 수천대의 분산된 장비에 대용량 파일을 저장할 수 있는 기능을 제공하는 분산파일 시스템과 저장된 파일 데이터를 분산된 서버의 CPU와 메모리 자원을 이용해 쉽고 빠르게 분석할 수 있는 컴퓨팅 시스템이다[2].

빅 데이터 시대에 기업들은 데이터의 효율적인 관리와 활용을 위해 다양한 장비에서 수집되고 저장된 대용량 데이터로서, 일반적인 데이터 분석을 통한 보안 기술로는 상당한 시간과 자원 낭비가 수반된다. 이를 개선하기 위해 본 논문에서는 하둡을 이용하여 대용량 데이터에 대한 처리 및 분석을 통해 효과적인 보안 시스템을 제안한다.

II. The Main Subject

대용량 보안이벤트 수집 및 관리엔진의 핵심 기술은 이기종 장비에 대한 이벤트 및 로그를 실시간으로 수집하고 수집된 정보로부터 공격 특징 정보를 추출하는 것이다. 이러한 기능은 기존의 보안장비나 네트워크 장비로부터 데이터를 얻는 것이므로 이를 물리적인 센서라고 명명한다. 기존의 보안 장비는 보안 관제를 통해 이상징후가 발생할 시 알람을 발생하므로 물리적인 센서라고 칭할 수 있다[3].

대규모 엔터프라이즈 네트웍에 포함된 이기종의 다양한 장비는 그 수가 많고 장비에서 발생되는 로그 및 이벤트의 양 또한 방대하다 일반적으로 보안장비 및 서버로부터 로그를 수집하기 위한 기술은 에이전트 기반 수집기술과 에이전트 없이 직접 수집하는 기술로 분류된다. 에이전트기반 수집기술은 상대적으로 보다 많은 양의 이벤트를 수집하고 독자적인 프로토콜을 사용하여 관리시스템과 통신할 수 있는 이점이 있는 반면 다양한 이기종의 장비를 연동하기 위해서는 장비별 별도의 에이전트개발 및 관리가 필요하며, 기본적으로 수집대상 장비에 탑재되는 기술이기 때문에 장비의 리소스를 일부 점유하는 등 단점이 존재한다[3].

Fig. 1. Realtime Log Collecting Function

따라서 본 논문에서는 그림 1에서처럼 보다 효과적이고 확장성 있는 시스템을 구현하기 위해 에이전트를 사용하지 않고 수집 기술을 적용한다. 에이전트가 없는 기반의 수집기술은 기본적으로 에이전트가 장비에 탑재되지 않고 공개 프로토콜인 SNMP, Syslog 등의 프로토콜을 이용하여 로그를 수집하게 된다. 또한 본 논문의 핵심요소인 대용량의 데이터를 처리하기 위해 분산처리기술인 하둡(Hadoop) 기술을 접목하여 대규모 네트워크의 대용량 이벤트를 실시간으로 수집하게 된다. 단, 방화벽, IDS/IPS등과 같은 보안 장비들은 기본적으로 발생된 보안 이벤트들에 대한 로그 기록을 저장 및 배포가 가능하다. 이 경우 에이전트를 설치하여 저장된 로그를 특정 시스템에 전송하도록 프로그램을 개발하지 않더라도, 설정 값 변경만으로 저장된 로그를 특정 시스템에 배포할 수 있다. 본 기술에서는 이러한 시스템에서 수집된 데이터 또한 수집하여 상관분석을 통해 공격 징후 탐지를 수행한다. 이와 같이 수집된 데이터는 RAW DATA의 원형을 그대로 유지하면서 1차로 저장되고, 저장된 데이터에 대한상관분석을 위해 다양한 이기종 장비에 대해 장비유형별(기존 보안장비인 방화벽, IDS, IPS, 서버 및 네트워크 장비 등) 공통포맷으로 정규화 한 후 정규화된 데이터는 별도로 저장하게 된다.

이렇게 저장된 데이터는 대용량 데이터로서, 일반적인 데이터 분석 기술로는 분석하는데 상당한 시간과 자원 낭비가 수반된다. 이를 개선하기 위해 본 기술에서는 하둡을 이용하여 대용량 데이터 분석을 수행하게 된다.

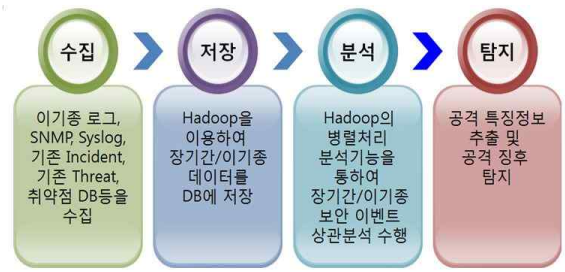

Fig. 2. 4-step Process for the Security of the Massive Data

그림 2는 대용량 데이터 보안 처리를 수행하기 위한 하둡 시스템 기반의 4단계(수집, 저장, 분석 및 결과보고)로 처리에 대한 내용이다. 본 논문에서는 하둡의 4가지 단계에 따라 이기종 시스템의 로그, SNMP/Syslog, 기존 Incident들, 기존 Threat들, 취약점 DB를 대용량 데이터베이스에 저장하고, 저장된 장기간/이기종 보안 이벤트의 상관관계 분석을 통해 공격특징 정보 추출 및 공격 징후를 탐지하게 된다.

III. Conclusions

대용량 보안이벤트 수집 및 관리엔진의 핵심 기술은 이기종 장비에 대한 이벤트 및 로그를 실시간으로 수집하고 수집된 정보로부터 공격 특징 정보를 추출하는 것이다. 추출된 정보를

본 논문에서는 하둡의 4가지 단계에 따라 이기종 시스템의 로그, SNMP/Syslog, 기존 Incident들, 기존 Threat들, 취약점 DB를 대용량 데이터베이스에 저장하고, 저장된 장기간/이기종 보안 이벤트의 상관관계 분석을 통해 공격특징 정보 추출 및 공격 징후를 탐지할 수 있는 대용량 보안 시스템을 제안한다.

향후 연구는 대용량 보안을 위한 하둡 시스템 기반에서 보안 이벤트 상관분석 수행에 대한 설계와 구현이 추가적으로 진행될 예정이다.

Reference

[1] http://www.boannews.com
[2] http://terms.naver.com
[3] Hyo-Nam Kim, “Realtime hybrid analysis based on multiple profile for prevention of malware”, Hongik Univ. Feb. 2014.