Journal of Knowledge Information Technology and Systems

ISSN 1975-7700

http://www.kkits.or.kr

Security Technology Framework in Bigdata Environment

Nan-Ju Kim, Eui-in Choi*

Department of Computer Engineering, Hannam University

ABSTRACT

The number of data is increasing due to develop of Internet, smart devices and IT sector. In other words, the explosive development of the IT field, the number of structured and unstructured data is increasing in geometrical. In addition, problem of numerous security threats was found. It is becoming intelligent and enhanced. An accident such as a large-scale system failures and large amounts of personal information leakage is caused by external hacking. Especially, the attack such as APT (Advanced Persistent Threat) to aim at a specific target is increasing. There are many companies and organizations to arrange countermeasures to respond to these attacks. Also, according to this trend, importance and analysis technology of Big Data is developing. But, security technology is showing a low level relatively. To study the techniques that can effectively respond to the attack by unknown cyber threats such as APT attack from a clear understanding of existing security with a focus on defense against attacks based on known APT attacks. And service development possible attack response services are desperately needed. In this paper, we list about definition and technique of Big Data in IT field buzzword. Also, we will research about security threat and security response technology.

KEYWORDS: Bigdata, Security technologies, Advanced persistent threat, Hadoop, Bigdata technologies

ARTICLE INFO: Received 3 May 2016, Revised 13 June 2016, Accepted 13 June 2016.

1. 서 론

최근 인터넷 및 소셜 미디어, 스마트 단말이 발

*Corresponding author is with the Department of Computer Engineering, Hannam University, 133 Ojeong-dong, Daedeok-gu, Daejeon, 34430, KOREA.
E-mail address: eichoi@hnu.kr

전함에 따라 사진, 이미지, 동영상과 같은 비정형 데이터들과 기존 정형데이터들의 수가 급격히 증가하고 있다. 하지만 수많은 데이터 중에서 가치 있는 데이터는 소수에 불과하고 이를 발견하기 위한 기술의 필요성이 대두되고 있기 때문에 빅데이터에 대한 관심이 증가하고 있다. 즉, 데이터의 수가 급격히 증가하는 빅데이터 시대로 변화하고 있으며 이에 따라 대용량 데이터를 수집, 저장, 관리, 분석하는 기술이 성장하고 있는데 이러한 기술의 발전은 사용자들 또는 기업의 요구에 따라 다양한 형태와 기능들을 제공할 수 있는 형태가 되어 가고 있다.

데이터의 폭발적인 증가에 따라 개방형 네트워크의 사용으로 인한 보안위협이 증가하고 있으며, 데이터 보호와 자원의 관리, 가용성 확보, 개인정보보호 등 해결되어야 할 여러 분야의 복잡한 보안 문제가 발생하고 있다. 특히, 빅데이터 환경에서는 지능형 보안 위협이 증가하고 있으며, 이런 위협은 금전적인 이득을 노린 개인 또는 기업 내 정보 유출과 장애 유발 및 파괴 등의 서비스 가용성과 관련된 침해사고가 일어나는 심각한 피해를 가져오게 된다. 그렇기 때문에 빅데이터와 정보보안을 분리하여 생각할 수 없는 상황이 되었다.

최근 몇 년에 걸쳐 국내·외에서 많은 사례가 발견되고 있으며, 2010년 7월 이란 원자력 발전시설을 해킹한 스턱스넷(Stuxnet)의 경우 원자력 발전시설의 원심분리기 중 20%가 가동 중간되었으며, 2011년 4월 국내 농협 전산망 자료 손상, 같은 해 7월 네이트 3,500만 명 개인 정보 유출 등의 공격은 피해 규모가 클 뿐만 아니라 공격 탐지가 짧게는 2개월, 길게는 몇 년이 소요된다[1].

이와 같이 보안위협은 알려진 공격인 웜, 바이러스로 시작하여 최근에는 지능화된 공격으로 금전적인 피해뿐만 아니라 사회적 혼란, 국가 안보를 위협하고 있다.

이러한 지능형 보안 위협뿐만 아니라 알려지지 않은 공격들에 대해 대응하기 위하여 현재 알려진 대응 기술들에 대한 장점과 문제점들에 대해 해결할 수 있는 방안에 대해 본 논문은 연구를 진행하였다. 또한 수많은 데이터들에 대한 수집 및 분석이 이루어져야 하므로 현 IT 환경에서 대두되고 있는 빅데이터 기술에 대해 알아본다. 이를 토대로 빅데이터 기술과 함께 현재의 보안 대응 기술을 보완한 보안 기술 프레임워크를 제안한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 관련 연구로 빅데이터에 대한 정의, 기술, 활용사례에 대해 열거하고, 제 3장에서는 보안 이슈와 보안 대책에 대해 설명한다. 또한 제 4장에서는 보안위협에 대응할 수 있는 보안 프레임워크를 제안하고, 제 5장에서는 결론을 기술한다.

2. 빅데이터 개요

2.1 빅데이터 정의

Wiki 사전에 따르는 빅데이터란, 기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석의 역량을 넘어서는 대량의 정형 또는 비정형 데이터 세트 및 이러한 데이트로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다[2]. 또한, Mackinsey의 보고에 따르면 기존 데이터베이스 관리 도구의 데이터 수집, 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터를 의미한다[3]. IDC는 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처라고 정의한다[4]. IBM은 빅데이터의 세 가지 요소를 3V라고 정의하고, 3V는 데이터의 다

양성, 규모, 빠른 속도를 의미하며 세 가지 중에서 두 가지를 충족시킬 수 있으면 빅데이터 기술이라고 정의하였다[5].

2.2 빅데이터 기술

2.1절에서 설명하였듯이 3가지의 빅데이터의 요소기술, 3V를 제공하기 위하여 시스템, 소프트웨어, 저장장치 등 여러 가지 구현 기술들이 필요한데, 이러한 다양한 지원기술들을 크게 다음 5가지의 기술 구성으로 분류할 수 있다[7,8].

2.2.1 데이터 수집/통합 단계

데이터 수집 및 통합단계에서는 새로운 데이터 생성, 네트워크에 산재해 있는 외부데이터 수집, 내외부 이종데이터 통합 등 데이터의 형태와 소재에 무관하게 데이터를 확보하는 기술들을 말한다. 수집로봇, 데이터 가상화, logging station등의 기술이 데이터 수집 및 통합단계에서 사용되는 기술들로, 전처리 단계로 진입하기 전에 다양한 원천의 데이터를 확보하는 일이 첫 번째로 수행되는 기술들이 필요하다.

2.2.2 데이터 전처리 단계

데이터 전처리 단계에서는 첫 번째 단계에서 수집된 정보들에서 센싱정보, SNS 등 지속적으로 발생하는 비정형 스트림 데이터를 정제하여 분석 가능한 형태로 구조화하여 분석의 정확성을 높이고 심층분석을 가능하게 하는 기술들을 말한다. 데이터 통합, 익명화, 정제, 검색, 인증 및 ETL(Extraction, Transformation, Loading) 기술의 개발을 통해 다음 단계로의 진입을 원활하게 해야 한다.

2.2.3 데이터 저장/관리 단계

데이터 저장 및 관리 기술은 웹 데이터, 소셜 미디어, 비즈니스 데이터, 센싱정보 등의 증가하는 다양한 형식의 데이터를 실시간으로 저장 및 관리할 수 있는 분산 컴퓨팅 기술을 말하는 것으로, 빅데이터 플랫폼의 핵심기술을 말한다. 빅데이터 저장기술인 NoSQL(Not Only SQL)은 관계형 데이터베이스(RDB, Relational Database)를 넘는 빅데이터 저장을 위한 새로운 데이터베이스개념으로 다양한 형태의 인터페이스를 제공하고 있으며 비표준화 상태이고 종류로는 MongoDB, Cassandra, Hbase등이 있다.

2.2.4 데이터 분석 단계

데이터 분석기술은 빅데이터에 내재된 가치를 추출하기 위해 필요한 대규모 통계처리, 데이터 마이닝, 그래프 마이닝 등의 분석기술, 기계학습, 인공지능 기술을 활용한 다양한 심층분석 기술로서, 빅데이터 플랫폼의 핵심 중의 핵심기술이다. 처리 복잡도가 높은 빅데이터의 특성상, 자연어 처리에 기반을 둔 텍스트 마이닝, 모든 웹 문서와 댓글 등에서 소비자의 의견을 수집, 분석해 제품이나 서비스에 대한 평판을 추출해 내는 평판 분석, 소셜 네트워크 내 영향력, 관심사, 성향 및 행동 패턴을 추출하는 소셜 네트워크 분석, 데이터 간의 유사도, 데이터 간의 거리에 기반을 둔 클러스터 분석 기술이 포함된다.

2.2.5 데이터 시각화 단계

데이터 분석 가시화 기술은 비전문가가 데이터 분석을 수행할 수 있는 환경을 제공하는 분석도구 기술과 분석결과를 함축적으로 표시하고 직관적인

정보를 제공하는 인포그래픽스 기술을 말한다.

분석한 데이터를 시각화하여 보여주는 기술로 분석된 데이터의 특징이나 의미를 쉽게 알 수 있도록 표현해주는 기술을 의미한다. 통계 계산 및 시각화를 위한 언어 및 개발 환경을 제공하며, 기본적인 통계 기법으로부터 모델링, 최신 데이터 마이닝, 시뮬레이션, 수치해석 기법까지 포함한다. 빅쿼리는 구글의 빅데이터 분석 솔루션 플랫폼으로, 오픈 소스 데이터 분석 툴인 하둡(Hadoop)을 활용해 신속히 분석하여 시각화할 수 있는 기술이다.

2.3 빅데이터의 활용 사례

국외의 경우, 미국 국세청에서는 탈세 및 사기로 인한 국가의 재정 위기 가능성이 증가함에 따라, 대용량의 데이터와 다양한 기술을 결합하여 탈세 및 사기 범죄 예방 시스템을 구축하였다. 일본에서는 센서데이터를 활용한 지능형 교통안내 시스템, 미국 국립보건원에서는 유전자 데이터 공유를 통한 질병치료체계를 마련하였다. 또한, 구글에서는 검색어 분석을 통한 독감예보 서비스를 제공하였고, FBI에서는 유전자 색인 시스템을 활용한 단시간 범인 검거 체계를 마련했다[6].

국내에서는 서울 심야 버스 운영 노선을 위하여 빅데이터 분석을 활용하였다.

3. 보안위협 및 보안기술

3.1 지능형 보안 위협

최근 빅데이터 화경에서 뜨고 있는 지능형 보안 위협공격(APT, Advanced Persistent Threat)은 지능형 방법으로 지속적으로 특정 대상에게 가하는 보안위협을 뜻한다. 특히, 불특정 다수를 노렸던 과거의 보안 위협과 달리 하나의 대상을 정해 성공할때까지 지속적으로 공격한다. 기존의 다양한 보안 수단들을 철저히 우회하여 발생하고 있으므로 방어하기가 매우 어려운 심각한 보안위협이다. 지능형 보안 위협 공격은 USB, 외장하드 등 네트워크를 이용하지 않고 정상적인 트래픽 경로를 사용하는 경우가 대부분이기 때문에 방어가 매우 어렵고, 정상적인 메일이나 웹 사이트 등을 통해 악성코드 배포가 가능하기 때문에 이를 사용하여 표적에 가해지는 공격을 완벽하게 탐지하기는 어렵다. 흔히 사용되는 기술로는 특정인을 대상으로 공격하는 스피어피싱 이메일을 사용하며, 사전에 장악한 다수의 프록시화된 컴퓨터들을 이용하여 명령 및 제어(C&C) 통신을 은닉한다.

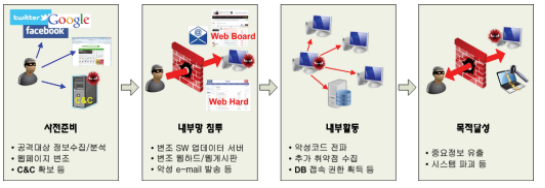

그림 1. 지능형 보안 위협 공격 단계

Figure 1. Attack step of Advanced Persistent Threat

지능형 보안 위협 공격의 단계는 <그림 1>과 같이 사전준비, 내부망 침투, 내부 활동, 목적 달성의 4단계로 정의한다. 사전준비 단계는 공격자가 이후 단계를 성공하기 위해 준비하는 과정으로써, 공격 대상에 대한 정보 수집 및 분석, 웹페이지 변조, C&C 확보 등과 같은 작업이 이루어진다. 내부망 침투 단계는 공격자가 공격 대상의 IT 인프라에 침투하는 단계로써, 악성 E-mail 발송, 변조된 웹 하드 및 웹 게시판 접속, 변조 업데이트 서버 접속 등의 방법을 통해 이루어진다. 내부 활동 단계는 내부망 침투 단계에서 감염시킨 좀비 PC를 거점화하여 최종 공격목적을 달성하기 위해 공격 대상의 내부 IT 인프라에 대한 정보를 수집하는 단계이다. 마지막, 목적달성 단계는 중요정보를 유출하거나

공격 대상 내부 IT 인프라를 파괴하는 등 지능형 보안 위협 공격의 목적을 달성하기 위한 작업이 진행되며, 특정 행위를 실행하기 위한 다양한 악성코드가 사용된다[9].

이 공격은 엔드포인트에 악성코드를 심는 작업으로 시작하여, 공격 전개도 엔드포인트를 기반으로 진행된다. 따라서 엔드포인트를 모니터링하지 않고서는 지능형 보안 위협 공격 대응이 매우 어렵다. 지능형 보안 위협 공격 대응을 위해서는 탐지, 분석, 치료, 방지와 같은 4단계의 대응 시스템을 구축해야 한다[10]. 악성행위를 탐지하고, 공격 전개과정을 파악하고 상황을 분석한 다음, 공격이 발견되었을 때 즉각적으로 멈추게 하고, 알려지지 않은 악성코드에 대한 저항력을 가지게 해야 한다.

3.2 빅데이터 환경에서의 보안 기술

<그림 2>에서 보듯이 첫 번째 과정은 여러 소스를 통해 생산되는 데이터를 수집하는 과정, 두 번째는 분산 처리 및 병렬처리를 위해 데이터의 분산 저장 및 운영 과정, 마지막으로 데이터 분석 및 2차 데이터 생성을 통해 서비스로 재사용되는 과정을 보여준다[7].

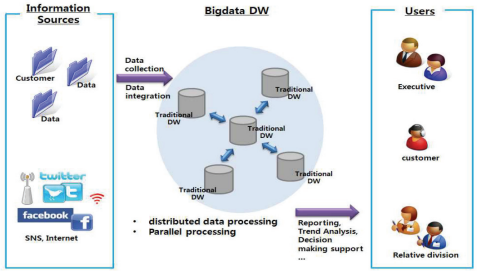

그림 2. 빅데이터 분석 과정

Figure 2. Bigdata analysis process

다음은 이에 따른 보안이슈 및 보안 대책에 대하여 설명한다.

3.2.1 빅데이터 수집 구간

빅데이터는 기존의 정형데이터들을 비롯하여 인터넷 및 소셜네트워크 서비스의 확산으로 인한 비정형 데이터들이 엄청난 양으로 생성되고 있다. 다양한 경로를 통해 생성, 수집되는 많은 데이터들은 많은 보안 위협들을 동반한다.

최근 사이버 표적공격 위협이 심화되어 사회적 국가적 위험으로 야기되고 있으며, 사이버 테러, 사이버전, 핵티비즘 등의 공격방법으로 활용되고 있는 지능형 보안 위협 공격은 목표 대상이 명확한 조직적 공격으로써 주로 정부나 기업을 대상으로 산업기밀이나 군사기밀, 고객정보 등의 정보 탈취를 목적으로 하고 있다[1]. 이에 따라, 데이터 신뢰성 및 무결성에 대한 우려가 높아지고 있다.

또한, PC위주의 주요업무와 개인 테블릿 PC, 스마트폰을 보조적 수단으로 업무에 활용하는 기업이 늘고 있음에 따라 개인 IT단말을 통해 수집되는 데이터들의 수가 증가하고 있다. 따라서 BYOD(Bring Your Own Device) 위협이 증가하고 있는데 BYOD는 개인소유의 IT단말기를 업무에 활용하는 현상을 의미한다. 이로 인한 대표적인 보안 위협에는 기업의 IT 통제권 상실, 단말기의 취약점 및 악성코드로 인한 기업 내부정보 유출위협, 악성코드에 감염된 개인용 기기의 내부 접속으로 인한 IT 자산 위협, 단말기 도난 또는 분실로 인한 데이터 유툴, 보안인식이 낮은 직원에 의한 계정 유출, 개인 정보 유출 등이 있다[11]. 개인 정보 유출로 인하여 개인 데이터가 무분별하게 수집되어 이용되는 프라이버시 문제가 생길 수도 있다.

이 두 가지 문제를 해결하기 위하여 다양한 연구가 진행되고 있는데, 전자서명, 다양한 필터링 기법, 스팸메일방지, 피싱방지, 최소한의 개인정보

를 수집하는 등의 기술들이 적용되고 있다.

3.2.2 빅데이터 분석 구간

다양한 경로를 통해 수집되는 데이터들을 처리, 저장, 운영, 분석하는 구간은 외부뿐만 아니라 내부로부터의 공격 위협에 노출될 수 있다. 다음은 빅데이터 분석 기술을 통한 보안 대책 방법에 대해 설명한다.

3.2.2.1 지능형 보안 관리 시스템

지능형보안관리 개념은 지능형 보안 위협 공격과 같은 알려지지 않은 치명적인 공격에 대응하기 위해 주요 IT기반시설의 네트워크, 시스템, 응용 서비스 등으로부터 발생하는 데이터 및 보안 이벤트의 연관성을 분석하여 보안관리 지능을 향상하는 차세대 보안정보 분석기술을 의미한다[1]. 이 모델은 이 기종 모델 및 다양한 운영체제의 기기로부터 정보를 수집한 다음 빅데이터 분석을 통한 분류, 해석 가능한 표준화 정보로 저장 관리 후, 보안관리 정책에 의해 평가하여 관리자에게 시각화 툴로 제공해 주고 있다[12].

3.2.2.2 단위 보안 시스템

대표적 단위 보안시스템은 침입차단시스템과 침입탐지시스템이 있다.

차세대 침입차단시스템은 기존 보안 솔루션의 한계를 극복하고 포트기반이 아닌 어플리케이션 계층에서 사용자 중심의 탐지분석 및 제어 기능을 할수 있다. 어플리케이션 탐지, 분석, 제어 기능 및 신규 어플리케이션의 식별 등록, 사용자 및 사용자 그룹별 보안정책 수립, 실시간 탐지 및 분석기능의 IPS, AV, URL 필터링 기능 통합, 어플리케이션 단위의 분석으로 처리성능 확보의 요구사항이 존재한다.

차세대 침입탐지시스템은 네트워크 행위 기반의 분석 방법 제공, 사용자 식별 추적 기능, 어플리케이션 모니터링, 자동화된 튜닝 및 워크플로우 기능, 컨텍스트를 인식한 영향평가와 처리 성능을 포함하고 있다[11].

3.2.2.3 통합보안 2.0

일반적으로 통합보안은 침입차단시스템, 침입탐지시스템, 가상사설망 등의 보안 솔루션의 로그와 이벤트 등을 통합한 시스템을 의미한다. 본 논문에서는 빅데이터 이전의 통합보안을 통합보안 1.0이라고 정의하였다. 통합보안 1.0은 네트워크 보안 시스템 중심에서 데이터를 단일 장비에서 수집하여 DBMS에서 처리하는 구조이지만, 통합보안 2.0은 수집성능과 분석성능을 고려하여 병렬처리가 가능하도록 구성하였다. 또한, 통합보안 1.0에서는 네트워크 계층의 규칙기반 연관 분석을 이용하여 알려진 공격과 IP, 포트 임계치 위주로 위협을 탐지하기 때문에 데이터를 실시간 분석하기는 어렵다. 하지만 통합보안 2.0에서는 고도화되고 지능적으로 잠복해 있는 위협을 탐지하기 위하여 사용자 ID, 어플리케이션에서 전후관계를 분석한 어플리케이션 계층 중심의 사이버 상황인지 방법으로 위협을 탐지할 수 있다. 분산 병렬 처리로 인하여 수개월 이상의 장기 데이터를 이용하여 실시간 분석도 가능하다[11].

3.2.2.4 하둡 보안

빅데이터 관리를 위하여 만들어진 하둡 시스템의 초기 버전에는 사용자 인증을 위한 보안을 제공하지 않았다. 알려진 하둡 보안 취약점으로는 하

나의 대칭키 암호키 사용문제와 하둡 생태계 시스템 중 하둡 보안을 지원하지 않는 객체 문제가 있다. 최근 하둡 1.0을 발표하면서 GSSAPI(Generic Services Application Program Interface)를 통하여 SASL(Simple Authentication and Security Layer)을 제공하면서 Kerberos 인증, RPC Digest 방식 등을 제공하기 시작하였다[13].

3.2.2.5 STAP

STAP(Specialized Threat Analysis)기술은 시그니처 기반의 안티악성코드와 침투탐지, 방지시스템을 포함한다. 네트워크 계층이나 엔드포인트 혹은 양쪽 모두에서 운영되는 STAP 제품들은 전형적인 침투상태를 알려주는 봇넷과 C&C트래픽 등을 포함한 들어오고 나가는 트래픽에 대해 스캔하고 이상활동들을 감지한다.

3.2.2.6 SIEM

6개월 또는 1년 이상의 기간동안 각종 보안 장비로부터 수집한 방대한 양의 로그 정보를 기반으로 장기 보안 이벤트에 대한 연관분석을 실시해 지능형 보안 위협 공격등의 보안위협에 대응할 수 있는 SIEM(Security Information and Event Management)은 빅데이터 기술이 접목된 사례로써 로그 정보분석에서 핵심적인 역할을 한다. SIEM은 방화벽, IDS/IPS, 안티바이러스 등의 보안 솔루션과 서버, 네트워크 장비 등으로부터 통계정보, 보안 이벤트 정보를 함께 가져와서 이들 정보들 간의 연관성 분석을 통해 보안 상황인지, 신속한 사건 대응과 로그 관리를 수행하는 기능을 제공한다[14, 15].

3.2.3 빅데이터 2차 데이터생성 및 사용구간

사용자가 원하는 데이터를 추출하기 위해 2차 데이터를 생성 및 사용하는 과정이 필요하다. 2차 데이터 생성시 프라이버시 침해 및 데이터의 기밀성에 노출될 위험이 있기 때문에 반드시 프라이버시 보호를 위해 익명화 및 암호화 기법 등이 도입되어야 한다. 최근 암호화된 상태에서 키워드를 통한 검색을 할 수 있는 키워드 기반 검색기법, 프라이버시를 보호하면서 데이터를 분석하는 PPDM (Privacy Preserving Data Mining) 기법 등이 연구되고 있다[7]. PPDM은 개인정보를 보호할 수 있도록 변환하거나 이를 보호할 수 있는 방법을 사용하여 데이터 마이닝을 수행하고 그 결과를 얻어내는 방법이다[13].

4. 빅데이터 보안 프레임워크

목표가 명확한 공격인 지능형 보안 위협 공격(APT)과 같은 지능형 보안 위협이 증가함에 따라 기존의 보안기술로는 방어가 어려운 것이 실정이다. 앞서 설명했듯이 보안 대응 기술에 대한 연구와 개발이 진행되고 있으며, 빅데이터 시스템, 저장, 분석기술을 이용한 보안 대응기술이 존재한다. 이에 따라 <그림 3>과 같이 기존의 보안 기술을 보완할 수 있는 보안 대응 기술에 대하여 제시한다. 또한, 제시한 프레임워크의 서버는 <그림 4>와 같은 프로세스가 동작하도록 설계한다.

현재 단위보안시스템과 통합보안관리시스템은 IP, 포트중심의 네트워크 계층의 탐지가 이루어지며 알려진 공격위주의 탐지 분석 및 대응, 단시간 내의 데이터 분석만 이루어지고 있다. 그렇기 때문에 알려지지 않은 공격의 탐지가 불가하고 사용자, 어플리케이션 단위 연관성 분석이 불가하며 수개월 범위의 데이터 분석이 불가능하다. 지능화된 보안 위협을 탐지하기 위해서는 침입탐지 로그, DNS 정보 등을 수집해야 한다. 이러한 정보들의 양은

방대하기 때문에 수집, 분석하기 위해서는 빅데이터 기술이 필요하다. 비정형 데이터 및 정형데이터 수집 및 가공, 실시간 분석 및 일괄 분석을 위해서는 하둡과 Storm 기술을 사용한다. 널리 사용되고 있는 하둡의 경우 Batch형 빅데이터 엔진이기 때문에 실시간 처리 및 인덱싱이 가능한 실시간 빅데이터 엔진이 필요하다.

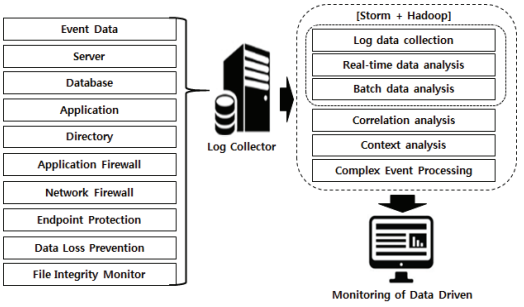

그림 3. 빅데이터 환경에서의 보안 기술 프레임워크
Figure 3. Security Technology Framework of Bigdata Environment

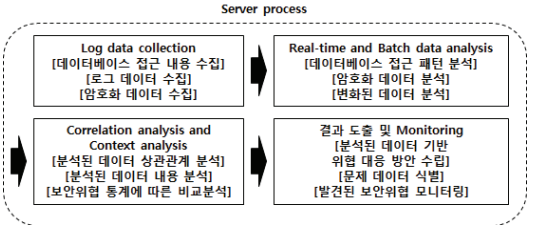

그림 4. 보안 기술 서버 프로세스
Figure 4. Security Technology Server Process

실시간 빅데이터 분석은 하둡에 CEP를 추가하거나, 실시간 분석가능 플랫폼인 Storm을 추가하여 사용한다. CEP는 이벤트에 대한 룰을 결정하고 적용 처리함으로써 빠른 응답시간 및 실시간 분석이 가능하다. Storm은 비정형 데이터에 대한 처리 유연성과 대규모 처리 확장성, 스트림 처리 실시간성을 가지며, 분산 병렬처리가 가능하다. 하둡 엔진에 포함되어 있는 HBase는 오류가 발생하여도 짧은 중단시간을 가지며 데이터 복구에 대한 무결성을 가지기 때문에, 보안 위협에 대해 끊임없이 모니터링이 가능하다. 또한, 지능화된 보안 위협을 탐지하기 위해서는 컨텍스트(Context) 중심의 분석과 상관관계 분석이 필요하다. 이러한 기능을 이용하여 발생하는 보안 위협들에 대해 즉각적으로 대응해야할 위협의 우선순위를 결정하며, 다른 보안 위협에 대해 대응할 수 있는 결과를 모니터링으로 보여준다.

제안한 프레임워크를 기반으로 기술들의 개발이 이루어진다면, 지능적으로 변화하고 있는 다양한 보안 위협들에 대해 대응할 수 있는 보안 대응 기술로 사용될 수 있을 것이라 기대한다.

5. 결 론

IT 기술과 스마트 단말기, 인터넷의 발전으로 인하여 보안위협이 증가하고 있다. 이러한 환경으로 인하여 보안위협의 추세도 단기간, 장난의 목적으로의 공격을 벗어나 장기간, 지능화, 사회혼란을 야기하는 목적으로 변화하였다. 또한, 빅데이터의 활용의 활성화와 함께 빅데이터 보안에 관심이 커지고 있으며 IT 및 보안 환경이 복잡해지면서 보안정보 및 이벤트 관리 솔루션은 필수 요소로 부상하고 있다. 따라서 기존의 보안 기술로는 현재의 보안 위협을 탐지하기 어려운 실정이다.

본 논문에서 제안한 빅데이터 환경에서의 보안 기술 프레임워크는 빅데이터 기술을 사용함으로써 기업이나 관공서에서 기존의 관계형 데이터베이스로 수 많은 데이터들이나 발생하는 보안위협에 대해 분석하기 어려웠던 점을 실시간 또는 일괄적으로 분석함으로써 기존 보안 기술보다 분석측면에서의 성능이 향상될 것으로 예측한다. 또한, 상관관계 분석과 내용기반 분석을 통하여 기존 보안위

협과 발생한 보안위협들을 비교 및 분석하여 효과적으로 대응할 수 있을 것으로 보아, 보안 측면에서도 성능이 향상될 것으로 예측한다.

기존의 보안 기술의 문제점과 현재 개발 중인 보안 기술들의 장단점을 분석하여 발전되는 보안 위협에 대응하는 보안기술 개발이 필요하다. IT 시스템에서 생성되는 대량의 데이터들을 분석하여 대응할 수 있는 기술과 네트워크 및 트래픽 차단, 엔드포인트 보안에 탁월한 기술을 결합한다면, 알려지지 않은 공격뿐만 아니라 지능형 보안 위협 공격에 대응할 수 있는 보안 기술로 사용될 수 있을 것이다. 또한, 알려지지 않은 공격을 탐지하기 위해서는 트래픽과 로그를 전수조사 해야 하는데 수천, 수만 건의 로그와 트래픽을 수초 내에 분석해서 적용하기 어렵기 때문에 빅데이터 플랫폼을 활용해서 문제점을 해결한다. 또한, 현재의 보안 위협이 기업뿐만 아니라 국가적인 차원으로 문제가 발생하고 있기 때문에 정부기관 및 기업 내에서 데이터 보호를 위한 플랫폼 개발이 필요하다.

여러 보안 위협들과 보안대응 기법들에 대하여 문제점을 분석하고, 다양한 보안 대응 기법들에 대한 연구와 개발이 이루어진다면 빅데이터 환경에서의 보안위협을 방지하는데 도움이 될 것으로 기대한다.

References

[1] J-H. Kim, S-H. Lim, I-K. Kim, H-S. Cho, and B-K. No, *Technical trends of cyber security with big data,* Journal of Electronics and Telecommunications Trends. Vol. 28, No. 3. pp. 19-29, 2013.

[2] Bigdata Definition, https://en.wikipedia.org/wiki/Big_data, Feb. 2014.

[3] Mackinsey & Company, *Bigdata research*, Mackinsey & Company, 2011.

[4] IDC, Digital Universes research, IDC, 2011.

[5] Bigdata characteristic, http://www.ibm.com/kr-ko/, Feb. 2014.

[6] NIA and Bigdata Strategic Research Center, *Big data 10 global best practices - It leads the world in big data*, Korea Information Society Agency, Aug. 2011.

[7] K-I. Jeong, H-N. Park, B-G. Jung, J-S. Jang, and M-A. Jung, *Big data and information security*, Journal of Korea Institute of Information Technology, Vol. 3, No. 10 pp. 17-22, 2012.

[8] Y-Y. Joe, *Understanding of big data and major issues,* The Korean Association for Regional Information Society, Vol. 3, No. 16, pp. 43-65, Oct. 2013.

[9] D-S. Moon, H-S. Lee, and I-K Kim, *Feature extraction for host based anomaly detection*, The institute of electronics engineers of korea, Vol. 37, No. 1, pp. 591-594, 2014.

[10] SK C&C, Advanced Persistent Threat, http://skccblog.tistory.com/1801, Feb. 2015.

[11] D-S. Choi, and Y-M. Kim, *Big data and integrated security 2.0*, Journal of Information Science, Vol. 1, No. 1, pp. 65-72, 2012.

[12] S-D. Yu, and D-H. Ryu, *Security response technology in big data environment*, NIPA Focus Week Technology Trends, Vol. 1, No. 1, pp. 1-14, 2013.

[13] D-Y. Lee, *Security issues of bigdata environment*, Bigdata Community, 2012.

[14] IDG IT WORLD, *Big data security analysis,* IDG Tech Focus, Vol. 1, No. 1, pp. 1-19, 2013.

[15] B-C. Kim, *Big data security technology and response study,* The Journal of Digital Policy & Management, Vol. 11, No. 10, pp. 445-451, 2013.

빅데이터 환경에서의 보안 기술 프레임워크

김난주, 최의인
한남대학교 컴퓨터공학과

요 약

인터넷의 발달과 스마트 디바이스, IT 분야들의 발달로 인하여 데이터들의 수가 증가하고 있다. 즉, IT 분야들의 폭발적인 발달로 인하여 정형 및 비정형 데이터들의 수가 기하학적으로 증가하고 있는 추세이다. 또한, 이러한 데이터들을 이용하여 고도화되고 지능화되어 가는 사이버 공격과 수많은 보안위협이 발생하고 있다. 외부 해킹으로 대량의 개인정보 유출, 대규모 시스템 장애 등 사고가 발생하고 있다. 특히, 지능형 보안 위협(APT)공격과 같이 특정 표적을 목표로 하는 공격이 증가하고 있다. 이러한 공격에 대응하기 위하여 많은 기업 및 조직들이 대응 방안을 마련하고 있다. 또한, 이러한 추세에 따라 빅데이터의 중요성과 분석기술들이 발전하고 있다. 하지만 보안에 대한 기술은 상대적으로 낮은 수준을 보이고 있다. 보안 공격에 대한 명확한 이해를 바탕으로 알려진 공격에 대한 방어에 중점을 둔 기존 보안에서 지능형 보안 위협 공격과 같이 알려지지 않은 사이버 위협에 의한 공격에 효과적으로 대응할 수 있는 기법을 연구하고 이를 통해 공격 대응 서비스가 가능한 개발의 필요성이 절실한 상황이다. 본 논문에서는 IT분야의 화두로 떠오르고 있는 빅데이터 정의와 기술들을 열거하고, 그에 따라 늘어가고 있는 보안위협과 대응할 수 있는 기술들에 대해 연구하였다. 그리고 기존의 보안 기법을 보완한 새로운 보안 대응 기술에 대한 프레임워크를 제시하였다.

감사의 글

이 논문은 2015년도 한남대학교 학술연구비조성비 지원에 의하여 연구되었음.

Nan Ju Kim received the bachelor's degree in the Department of Computer Engineering from the Hannam University in 2014. She received the M.S. degree in the Department of Computer Engineering from the Hannam University in 2016. She has been the doctor´s course student in the Department of Computer Engineering at Hannam University since 2016. She current research interests include ontology, cloud computing, big data.

E-mail address: 91knj@naver.com

Eui In Choi received the bachelor's degree in the Department of Calculation Statistics from the Hannam University in 1982. He received the M.S. degree and the Ph.D. degree in the Department of Electronic Calculation from HongIk University in 1984 and 1995, respectively. From 2003 to 2004, he was a visiting professor at UCLA. He has been a professor in the Department of Computer Engineering at Hannam University since 1996. His current research interests include semantic web, ubiquitous computing, mobile, cloud computing, big data.

E-mail address: eichoi@hnu.kr