빅데이터 플랫폼을 이용한 보안로그 분석 시스템 구현 모델 연구

한기형*, 정형종**, 이두식***, 채명희****, 윤철희****, 노규성*****
한전KDN*, 한전KDN**, 아임클라우드***, KT****, 경찰대*****, 선문대학교*****

A Study on implementation model for security log analysis system using Big Data platform

Ki-hyoung Han*, Hyung-Jong Jeong**, Doog-Sik Lee***,
Myung-hui Chae****, Cheol-hee Yoon****, Kyoo-sung Noh*****
KEPCO KDN*, KEPCO KDN**, IMCLOUD Corporation***, KT****,
Korean National Police University*****, Sun Moon University*****

요약 보안 장비에서 발생하는 로그는 그동안 ESM(Enterprise Security Management) 기반으로 통합적으로 데이터를 분석하였으나 데이터 저장 용량의 한계와 ESM자체의 데이터 처리 성능의 한계로 빅데이터 처리에 부적합하기 때문에 빅데이터 플랫폼을 이용한 보안로그 분석 기술이 필요하다. 빅데이터 플랫폼은 Hadoop Ecosystem을 이용하여 대용량의 데이터 수집, 저장, 처리, 검색, 분석, 시각화 기능을 구현할 수 있다. 현재 ESM기술은 SIEM(Security Information & Event Management)방식으로 기술이 발전하고 있으며 SIEM방식의 보안기술을 구현하기 위해서는 현재 보안장비에서 발생하는 방대한 로그 데이터를 처리할 수 있는 빅데이터 플랫폼 기술이 필수적이다. 본 논문은 Hadoop Ecosystem 이 가지고 있는 빅데이터 플랫폼 기술을 활용하여 보안로그를 분석하기 위한 시스템을 어떻게 구현할 수 있는지에 대한 모델을 연구하였다.

주제어 : 빅데이터플랫폼, 하둡, 분산병렬처리, 데이터시각화, 통합보안관리시스템, 보안정보이벤트관리

Abstract The log data generated by security equipment have been synthetically analyzed on the ESM(Enterprise Security Management) base so far, but due to its limitations of the capacity and processing performance, it is not suited for big data processing. Therefore the another way of technology on the big data platform is necessary. Big Data platform can achieve a large amount of data collection, storage, processing, retrieval, analysis, and visualization by using Hadoop Ecosystem. Currently ESM technology has developed in the way of SIEM (Security Information & Event Management) technology, and to implement security technology in SIEM way, Big Data platform technology is essential that can handle large log data which occurs in the current security devices. In this paper, we have a big data platform Hadoop Ecosystem technology for analyzing the security log for sure how to implement the system model is studied.

Key Words : BigData Platform, Hadoop, Distributed Parallel Process, Data Visualization, ESM, SIEM

Received 10 June 2014, Revised 27 July 2014

Accepted 20 August 2014

Corresponding Author: Kyoo-Sung Noh (Sun Moon University)

Email: ksnoh@sunmoon.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

일반적으로 보안 장비에서 발생하는 보안로그는 개별 장비에서 자체 분석하거나 ESM을 통해서 통합적으로 전체 로그 관리를 하고 있다. 이 방법은 데이터 저장 용량과 성능의 한계로 장기간 쌓인 빅데이터에 대한 분석을 할 수가 없으며 빠른 데이터 검색도 불가능하다. 이는 사전적 보안성 강화와 보안 취약성에 대한 사전 진단이 어려운 기능적 제약을 의미한다.

이러한 제약은 빅데이터 기술과 접목되면 해결될 수 있다. 즉 빅데이터 플랫폼에 로그를 저장하고 통합적으로 데이터 수집, 저장, 처리, 검색, 분석, 시각화 하여 기존에 해결이 불가능한 로그 분석을 하게 되면 보안이 강화될 수 있는 것이다.

이에 본 연구는 보안로그 분석을 통해 장비의 보안성을 강화하는 빅데이터 플랫폼 모델을 제안하는 것을 목적으로 하고 있다. 이러한 목적 달성을 위하여 본 연구는 빅데이터 플랫폼을 이용하여 보안장비에서 발생하는 로그를 통합적으로 분석하기 위한 모델을 제시하고, 이 모델 구현을 위한 Hadoop Ecosystem과 Node.js 엔진을 이용한 세부 기술 구성 요소와 분석 UI 구현 방법을 제시한다.

2. 이론적 배경

2.1 빅데이터 플랫폼 정의

[Fig. 1] Concept of Big Data

빅데이터 플랫폼 기술은 대용량의 데이터를 처리하기 위한 데이터 저장공간, 고성능 수집처리, 데이터 분석엔진, 빠르게 검색할 수 있는 검색엔진 그리고 분석된 데이터를 표현하기 위한 시각화 기술이 필요하다.

빅데이터 플랫폼을 적용하는 방법은 플랫폼, Ecosystem, Convergence 등의 Smart Work 전략을 기반으로 서비스와 디바이스 레이어의 기술적인 구현을 통해서 빅데이터를 수집 및 분석 가공하여 보안 기술을 서비스에 접목시키기 위한 방법으로 활용한다.

2.2 빅데이터 플랫폼 사전진단

TMS/ESM/EMS 등 다양한 보안 및 모니터링 솔루션에서 수집하는 데이터의 통합 분석을 통해 End-to-End 위협을 실시간 모니터링하여 서버 및 장비의 상태정보, 보안장비에서 외부 침입정보, 최종 단말에서 트래픽 정보를 수집하여 빅데이터에 대한 분석을 실시한다[7]. 침입자의 경로, 침입패턴 분석, 현재 침입상태의 정확한 알람기능 제공은 위협의 감소 및 사전진단을 가능하게 한다.

[Fig. 2] Pre-Diagnostics of Big Data

3. 빅데이터 플랫폼 아키텍처

3.1 보안로그 빅데이터 S/W 아키텍처

빅데이터 플랫폼 소프트웨어[5]는 데이터의 대용량 수집, 저장, 처리, 검색, 분석, 시각화 요소로 구성되어 있으며 각각의 기능은 독립적인 소프트웨어 엔진으로 동작되고 동시에 다른 기능들과 다양한 상호 작용으로 연동된다.

Hadoop Ecosystem은 이러한 기술을 제공해 줄 수

있는 플랫폼으로 기존의 ESM의 한계를 극복할 수 있는 SIEM기반의 보안 기술을 구현할 수 있는 방법이다. Hadoop Ecosystem이 가지고 있는 Flume(수집), Sqoop(수집), Hive(Map/Reduce), R(분석), Solr(검색), Node.js(시각화), Hadoop(저장) 등의 개별 기술을 활용하여 빅데이터 기반의 보안로그 시스템에 대한 모델 제시가 가능하다.

[Fig. 3] Big Data Platform Software Architecture

3.2 적용 오픈소스 기술

빅데이터 소프트웨어 플랫폼은 Hadoop을 중심으로 다양한 오픈소스의 연동으로 구성되어 있다.[5,8] Hadoop Ecosystem이라고 불리는 오픈소스들은 주로 Hadoop과 데이터 연동을 지원한다. 로그 데이터 수집처리는 Flume엔진을 이용하고, 관계형 데이터베이스에 존재하는 데이터의 수집은 Sqoop을 이용한다.

<Table 1> Open Source Big Data Platform

Division	Open Source	Features and Description
Data collection	Flume Sqoop	Collection process for Log data and RDBMS
Data storage	Hadoop	Save source data on HDFS MAP/REDUCE, Data replication and recovery
Data Process Scheduler	Hive	SQL on hadoop
Search	Oozie	Job scheduling
	Solr	indexing, search process
Analysis	R	Data analysis using statistics technology

Hive는 Hadoop에 저장된 데이터를 MAP/REDUCE방

식으로 데이터를 분산병렬 처리하고 Solr는 Lucene을 기반으로 한 검색엔진으로 Hadoop의 원본 데이터를 이용하여 검색에 필요한 인덱싱 작업을 처리한다. R엔진은 통계기반으로 데이터를 분석하기 위한 다양한 통계적 함수를 이용하여 데이터 패턴과 특성을 구분해 낸다. 빅데이터 시각화는 빅데이터 자료처리에 적합한 방식으로 Node.js 서버 기술을 이용하여 구성한다.

3.3 보안로그 빅데이터 수집

3.3.1 보안 Log 수집 엔진 Flume

데이터 수집 처리는 Flume 엔진[5,9]을 이용하는데, Flume은 보안장비에서 발생하는 대용량의 로그를 실시간으로 수집하기에 적합하다. 보안 장비에서 발생하는 로그 데이터는 일반적으로 Syslog 데이터 타입으로 되어 있으며 Flume은 Syslog 데이터 타입을 지원한다. Flume은 TCP 또는 UDP 표준 프로토콜을 이용하여 데이터를 수집하여 Hadoop에 고속으로 저장한다. 수집 기능이 대용량일 경우에 Flume은 다수의 Flume Agent를 소프트웨어적으로 다수 생성하여 처리한다. Flume Agent는 다수의 물리적인 서버 여러 대로 Scale-out 방식으로 구성할 수도 있다. 내부적으로는 Source, Channel, Sink 3가지로 구성되며 Channel은 메모리 또는 하드디스크에 임시로 데이터를 생성했다가 Sink를 통해서 데이터를 저장하는 버퍼링 역할을 수행한다.

결론적으로 Flume 엔진은 대용량의 보안 로그 데이터를 Hadoop에 원활하게 저장하면서 병렬적으로 확장(Scale-out)가능하여 매우 유연성 있게 데이터를 고성능으로 수집 처리하는 기능을 가지고 있다.

[Fig. 4] Syslog Collection Process

3.3.2 RDBMS 방식 Legacy 연동

보안로그 데이터는 대부분 장비로부터 로그를 수집한다[5,13]. 그러나 빅데이터 기반의 보안로그 데이터 수집은 기존 보안장비가 이미 구축한 통합 보안로그 모니터링 솔루션에 존재하는 시스템의 관계형 데이터베이스로부터 데이터를 수집할 수 있다. 또한 보안로그와 기존의 Legacy시스템과의 상관관계를 분석하기 위해서 업무 데이터를 일부 수집할 수 있다. 이 경우 데이터 수집은 Sqoop엔진을 이용한다. Sqoop은 Oracle, MySQL, MS-SQL 등 기존의 RDBMS 형태의 데이터베이스에 있는 데이터를 SQL문으로 설정하여 Hadoop에 데이터를 저장하는 엔진이며, Sqoop은 RDBMS에 데이터베이스 연결 기술의 표준인 JDBC 연결을 통해서 고속으로 데이터를 가져오는 Command-Line 인터페이스 애플리케이션이다.

3.4 보안로그 빅데이터 저장

3.4.1 Hadoop의 데이터 복제 기능

대량의 보안로그 파일 저장은 파일시스템 기반의 Hadoop을 사용한다. Hadoop은 데이터를 저장할 때 물리적으로 64MB단위의 데이터 블록 단위로 저장하며 데이터가 저장된 DataNode 서버는 논리적으로 Ring-Node를 구성하여 물리적인 DataNode 서버가 마치 하나의 데이터 저장소처럼 동작한다. 특히 Hadoop은 데이터를 저장할 때 자동 데이터 복제 기능을 제공한다. 장애가 발생하면 이미 복제되어 있는 데이터 블록을 이용하여 또 다른 DataNode 서버에 자동으로 데이터를 복구한다. 복제 기능은 디폴트로 3개의 동일한 데이터 블록을 가지지만 복제 숫자는 다르게 설정이 가능하다. Hadoop의 자동 복제 기능은 백업이 불필요한 방식으로 이론적으로 무한대의 데이터 저장이 가능하다.

3.4.2 MAP/REDUCE 방식 분산병렬 처리

하둡은 MAP/REDUCE[5,10,11] 기술을 통해서 데이터 저장 및 조회 등의 프로세스 처리과정에서 분산 병렬 처리 기능을 제공한다. 매우 큰 보안로그 데이터 사이즈 구축시 하둡의 DataNode 서버 대수가 매우 커질 수 있으며 이 경우 하둡의 데이터 처리 방식을 순차적으로 처리하게 되면 너무 많은 시간이 걸리게 된다. 이때

MAP/REDUCE 기술은 각각의 데이터 노드가 병렬적으로 처리하고 처리한 결과를 최종적으로 연산처리 함으로써, 순차적인 처리 방식보다 더 빠른 결과를 내게 된다. 이 기술은 데이터 사이즈가 작을 때 보다 대용량의 보안로그 빅데이터를 처리할 때는 매우 큰 효과를 낼 수 있다.

[Fig. 5] Parallel Distributed Processing and Storage System

3.5 보안로그 데이터구조화 및 Job 스케줄링

3.5.1 Hive를 이용한 데이터 구조화 처리

빅데이터를 저장하는 Hadoop은 HDFS(Hadoop Distributed File System)에 데이터를 저장한다[4,5]. 일종의 자체적인 파일시스템이다. Hive는 Hadoop이 설치된 상태에서 사용되는 소프트웨어 엔진으로 Hadoop에 저장된 원본 데이터에 대한 스키마를 자체적으로 관리한다. 데이터베이스, 테이블, 컬럼 등 스키마 정보를 관리하고 SQL문법을 이용해서 데이터를 처리한다. Hive는 자체적으로 MAP/REDUCE기능을 제공하기 때문에 Hadoop에 데이터를 저장할 때 자동적으로 분산 병렬처리 기능을 제공한다.

3.5.2 Oozie를 이용한 데이터 처리 스케줄링

보안로그 데이터를 수집하고 저장하고 처리하는 모든 과정들은 한번만 발생하지 않고 Job으로 구분한 시나리오 방식의 스케줄링이 필요하다[5,12]. Oozie는 빅데이터 처리에 있어서 중요한 스케줄러 기능을 담당한다. Oozie는 데이터를 수집하고 저장하고 조회하는 Job(작업 단위)이 있다면 반드시 순차적으로 Job이 발생하도록 스케줄링하는 것이 가능하다 일종의 워크플로우를 구성하는 엔진으로 볼 수 있다. 빅데이터 처리는 매우 복잡하고 다양한 Job들로 구성 되어 있기 때문에 Job의 순서를 스케줄링하는 일은 매우 중요한 기술 중 하나이다. 이 기술

을 이용하면 보안로그 데이터를 분석하기 위해서 처리해야 하는 일의 순서를 정함으로서 시나리오 기법의 분석 Rule을 정의 할 수 있다.

3.6 보안로그 빅데이터 검색

보안로그 빅데이터는 대용량의 데이터이기 때문에 데이터를 실시간으로 조회하는 것은 속도를 보장할 수 없다. 검색 엔진으로 이 문제를 해결할 수 있다. Solr를 이용한 데이터 검색 엔진은 원본 데이터에 대한 인덱싱 처리를 한다[5]. 인덱싱은 배치처리 방식으로 미리 지정한 Key를 이용하여 데이터를 빠르게 검색할 수 있도록 해준다. 인덱싱한 데이터는 Solr에 검색 키워드 또는 조건을 입력하면 Solr엔진이 인덱싱 데이터를 찾아내는 방식으로 동작된다. 검색엔진은 사실 실시간 데이터 검색은 아니며 미리 인덱싱하기 때문에 Near Realtime 이라고 한다. 이 방식은 구글 같은 검색 서비스가 동작하는 원리와 동일하며 보안로그가 가지고 있는 컬럼의 정보를 기준으로 빠르게 검색해낼 수 있는 인덱싱 작업을 해낼 수 있다.

[Fig. 6] Solr System Big Data Searching Structure

3.7 보안로그 데이터 상관 분석 기술

보안로그[1,2,3] 빅데이터를 분석하기 위해서 R엔진을 이용한 통계적인 기법을 사용한다[14]. R의 통계 함수를 이용하여 분석에 필요한 Rule을 사용자가 만들어야 한다. 이 Rule은 데이터끼리 비교하기 위한 패턴이 될 수도 있고 시계열 방식의 히스토리 비교가 될 수도 있다. 빅데이터는 동일한 데이터를 비교하기 보다는 서로 연관성이 높은 데이터를 이용한 연관분석 또는 융합분석이 가능하다. R엔진은 이런 분석 기법에 매우 효과적으로 사용이 가능하다.

보안 장비가 발생시키는 로그 데이터는 Source IP, Port, Target IP, Event 등 다양한 컬럼이 존재하는데 데이터가 발생한 시간을 기준으로 특정 시간대에 정상적인 사용자가 평균적으로 트래픽을 발생시키는 것과 비교해서 특이한 데이터 발생 패턴을 분석하는 방식으로 데이터를 분석하며 해당 데이터와 연관성이 높은 다른 보안 장비의 동일 IP기준 로그를 동시에 연관분석하여 종합적으로 상관분석 작업을 할 수 있다.

3.8 보안로그 빅데이터 시각화

3.8.1 Node.js 방식의 시각화 엔진

빅데이터 분석은 사용자 관점에서 데이터를 인지할 수 있는 시각화 서비스가 필요하다[5]. 빅데이터 시각화는 여러 가지 기법이 있으나 HTML5, JQuery, JSON, Java Script, ECMA Script 등의 표준적인 웹스크립트 기법으로 구글이 개발한 크롬의 V8엔진을 이용하여 인터프리터 방식으로 서버에서 동작되는 Node.js 엔진을 이용한 시각화 방법을 제시한다.

[Fig. 7] Node.js System Big Data Visualization Technology

과거에는 데이터 시각화 서비스를 구성하려면 클라이언트에 사용하는 스크립트 기술과 달리 서버는 별도의 무거운 서버 프로그래밍 스크립트 기술을 이용해야 했다. 이것은 다양한 서버의 미들웨어 프로그래밍 기술로 발전되어 왔다 빅데이터를 매우 쉽게 처리하고 시각화된 서비스로 구성하기 위해서 서버에서도 클라이언트에서 사용하던 스크립트 기술을 그대로 사용할 수 있는 Node.js를 이용한 웹 서비스 방법을 제시한다.

3.8.2 보안로그 시각화 군집화 분석

보안로그 시각화는 로그에 발생하는 데이터에 대한 군집화 작업이 필요하다. 발생한 로그 데이터에 대한 분포, 크기를 그룹핑해서 전체적인 데이터에 대한 분포와 크기를 판단할 필요가 있다. 군집화 방법은 다양한 시각화 방식이 있으나 원으로 표현하여 사이즈가 큰 데이터의 원을 더욱 크게 표현하는 방식으로 표현할 수 있으며 해당 원을 클릭하여 메뉴의 깊이를 만들고 추적해 들어가는 방식으로 구현한다. 이 표현 방식은 보안로그 데이터에 대한 분석을 위한 시작점으로 사용한다.

[Fig 8] Clustering Visualization UI

3.8.3 보안로그 시각화 추적 분석

보안로그는 IP, Port 에 대한 Key를 기준으로 서로 다른 장비에서 이벤트가 어떻게 얼마나 발생하는지 측정하는 것이 중요하다 또한 IP가 어떤 보안장비 순서대로 발생되었는지와 어떤 장비까지 탐지가 되었는지를 측정함으로써 보안로그 데이터에 대한 추적 분석이 가능하다.

[Fig.9] Log Tracking Visualization UI

4. 빅데이터 기술의 활용

4.1 빅데이터 기술적용

기존의 보안장비와 ESM을 이용한 로그관리 방식은 수개월 정도의 데이터만 저장이 가능하며 통합 관리되는 데이터의 다양한 조건에 대한 검색이 불가능하다.

[Fig. 10] Big Data Technology Utilization

이는 사이버공격에 대한 사전탐지 및 사후 분석 시간의 지연을 초래한다. 이런 문제를 해결하기 위한 빅데이터 방식의 보안로그 관리가 필요하다. 빅데이터 방식의 기술을 보안 Rule과 매칭하여 데이터를 분석, 가공 및 생성하고 보안관계 업무에 쉽게 적용하기 위한 Task 관점으로 빅데이터 기술을 적용한다.

4.2 단계별 적용 방안

F/W, IDS, IPS와 같은 단일장비와 ESM, TMS와 같은 통합 솔루션 등 보안관계 로그 기반으로 3단계 빅데이터 플랫폼 구축이 필요하다[6]. 1단계에서는 기본적인 빅데이터 플랫폼과 빠른 검색에 비중을 두고, 2,3단계에서는 분석엔진 구축, 시각화 서비스, 분석 고도화 작업이 이루어진다.

[Fig. 11] Big Data Development Phase Type

4.3 빅데이터 플랫폼 처리

[Fig. 12] Big Data Processing Platform

다양한 보안장비로부터 고속으로 데이터를 수집, 처리하여 Hadoop에 저장하게 된다. 저장된 데이터는 MAP/REDUCE 방식으로 처리되고, 검색 엔진화된 플랫폼은 Script 처리된 환경에서 통합조회, 검색이 제공된다. 첫째, 저장은 대용량 분산 처리 기반 Hadoop으로 구현되며, 둘째, 검색은 다양한 보안장비 데이터를 Map/Reduce 방식으로 처리하여 검색엔진을 구현한다. 마지막으로 No-컴파일 방식의 ECMA Script 기반의 결과 값을 도출하게 된다.

4.4 플랫폼의 자동화 구현

4.4.1 보안로그 검색 시나리오

보안로그의 실시간 수집 건수는 매초 단위로 실시간 수집이 가능하여야 하며 일일 수집건수는 날짜별로 Reporting이 되어야 즉시성과 자동화가 보장된다. 처리된 데이터는 보안장비별 수집과 검색대상 장비별 통합 검색이 가능하여야 한다. 검색조건 컬럼 선택을 통해 타켓 IP와 소스IP에 대한 다양한 대비가 가능하다. 특히 IP에 대한 추적 검색이 매우 중요한 요소로 침입자의 IP를 추적하기 위해서는 NAT변환 또는 IP를 변경하면서 공격하는 방식에 대한 조건 검색 기능이 필수적이다.

4.4.2 시각화를 통한 지역별 노드 클러스터링

국내 지점별 노드화를 통해 지역별 로그 수집 현황을 검색할 수 있으며 화면의 경우 경기도를 타겟으로 하는 IP를 검색한 결과를 제시하고 있다. 이 영역을 확대하면 국가별 소스IP 검색이 가능하다. 각 지점 관제센터간의 노드확대는 빅데이터의 수집, 저장, 분석, 처리의 품질을

[Fig. 13] Automation processing Dashboard

높여주며 보안취약점의 손쉬운 control을 지원하게 된다. 그동안 일반 EMS 기반의 데이터 분석의 경우 저장 용량의 한계와 ESM 자체의 데이터 처리 성능의 한계로 빅데이터 처리에 부족하였으나, 빅데이터 플랫폼의 적용으로 현재 보안장비에서 발생하는 방대한 로그 데이터를 실시간, 자동화, 시각화, 군집화 적용하여 처리할 수 있음과 동시에 대용량의 데이터를 수집 및 저장하는 과정에서 성능 이슈가 없음을 확인 할 수 있다.

[Fig. 14] Regional Search and Automated Environment

5. 결론

빅데이터 플랫폼을 이용하여 보안로그 분석을 위한 시스템 구성은 Hadoop Ecosystem이 가지고 있는 수집,

저장, 처리, 분석, 검색, 시각화 기술을 채택하여 구현할 수 있다. 보안장비에서 발생 로그를 장기간 저장하고 빠르게 검색 및 분석 가능한 방식으로 시각화 방법을 구현하면 기존에 불가능했던 분석을 가능하도록 시스템을 구성할 수 있다. 특히, 시각화 부분은 Node.js방식을 이용하여 무거운 JVM방식의 웹서비스 구축 방식이 아닌 손쉬운 스크립트 방식으로 구현하였다. 이 방식은 다양한 보안로그 분석에 필요한 시나리오 방식의 Rule을 생성해서 시각화 서비스와 쉽게 매칭할 수 있는 방법으로 군집화 방식과 추적방식의 분석 방법을 제시하면 효과적으로 시각화된 보안로그 분석이 가능하다.

본 연구는 실증적 구현에도 불구하고 통계적 검증을 하지 못한 한계를 가지고 있다. 향후에 장기간 저장된 로그 데이터에 대한 더욱 다양한 보안로그 분석 기법과 Rule에 대한 연구와 실증적 분석 연구가 추가적으로 필요하다.

REFERENCES

- [1] Cardenas, A. A. , Manadhata, P. K. , Rajan, S. P. : "Big Data Analytics for Security", IEEE SECURITY AND PRIVACY MAGAZINE Vol.11 No.6, 2013.
- [2] Dae-Soo Choi ,Gil-Jong Mun ,Yong-Min Kim, Bong-Nam Noh : "An Analysis of Large-Scale Security Log using MapReduce", Korean Institute Of Information Technology, Vol.9 No.8, 2011.
- [3] Forte, D. : "The importance of log files in security incident prevention", NETWORK SECURITY Vol.2009 No.7, 2009.
- [4] Jiaqi Zhaoa, Lizhe Wangb, Jie Taoc, Jinjun Chend, Weiye Sunc, Rajiv Ranjane, Joanna Kolodziejf, Achim Streite, Dimitrios Georgakopoulouse: "A security framework in G-Hadoop for big data computing across distributed Cloud data centres", Journal of Computer and System Sciences, Vol.80 No.5, 2014.
- [5] Kyoo-Sung Noh, Doosik Lee : "A study on implementation model for Big Data Platform", The

Society of Digital Policy & Management, Spring Conference Paper, 2014.

- [6] Ok Hyun Ha: "A Study on Conversion Security Control System for Industrial Security", Korea Information Assurance Society, Journal of Information and Security, Vol.9 No.4, 2009.
- [7] Tankard, Colin : "Big data security", Network Security. Jul2012, Vol. 2012 Issue 7, 2012.
- [8] Xin Liang Li , Jian De Zheng : "Improvement of Hadoop Security Mechanism", Applied Mechanics and Materials (Volumes 484 - 485), Green Power, Materials and Manufacturing Technology and Applications III Chapter 7: Computer and Numerical Technologies 912-915, 2014.
- [9] <http://flume.apache.org/>
- [10] <http://hadoop.apache.org/>
- [11] <http://hive.apache.org/>
- [12] <https://oozie.apache.org/>
- [13] <http://sqoop.apache.org/>
- [14] <http://www.r-project.org/>
- [15] <http://www.nodejs.org/>

한 기 형(Han, Ki-hyoung)

- 2010년 2월 : 건국대 정보통신대학원 석사
- 1995년 7월 ~ 현재 : 한전KDN
- 2002년 1월 ~ 2005년 12월 : 한수원 SAP BC담당
- 2014년 1월 ~ 현재 : 한수원 사이버보안관제센터 PM
- 관심분야 : 정보보안, 빅데이터
- E-Mail : hankdn1@naver.com

정 형 중(Jeong, Hyung-Jong)

- 1981년 8월 : 중앙대학교 전기공학과(공학석사)
- 2013년 8월 ~ 현재 : 숭실대학교 IT 정책경영학과(박사과정 재학중)
- 2013년 4월 ~ 현재 : 한전KDN(주) E-ICT사업단장
- 관심분야 : 프로젝트관리, 빅데이터
- E-Mail : jhj0613@naver.com

이 두 식(Lee, Doo-sik)

- 2007년 8월 : 고려대학교 미디어공학 석사
- 1995년 6월 ~ 2000년 3월 : Microsoft
- 2000년 3월 ~ 2001년 3월 : 옥션 시스템운영 팀장
- 2008년 6월 ~ 2010년 12월 : NHN 수석PM
- 2014년 4월 ~ 현재 : (주)아임클라우드 대표
- 관심분야 : 빅데이터, 오픈소스, 보안
- E-Mail : dslee@imcloud.co.kr

채 명 희(Chae, Myung-hui)

- 2000년 2월 : 경북대학교 컴퓨터공학 학사
- 2000년 3월 ~ 현재 : KT 연구원
- 2010년 11월 : 컴퓨터시스템응용기술사
- 관심분야 : 네트워크보안, 보안관계
- E-Mail : lovevanil@naver.com

윤 철 희(Yoon, Cheol-hee)

- 2004년 2월 : 한성대학교 정보시스템 학사
- 2012년 2월 ~ 현재 : 경찰대학교
- 2012년 2월 ~ 현재 : 정보통신운용 정보 보안담당
- 2012년 2월 ~ 현재 : 국제사이버범죄센터 연구원
- 관심분야 : 네트워크보안, 보안관계
- E-Mail : bertter@police.ac.kr

노 규 성(Noh, Kyoo-sung)

- 1984년 2월 : 한국외대 경영학과 (경영학사)
- 1995년 8월 : 한국외대 대학원 경영정보학과(경영정보학 박사)
- 2003년 7월 ~ 2010년 12월 : 中國延邊科學技術大學 兼職教授
- 1997년 9월 ~ 현재 : 선문대 경영학부 교수
- 2004년 4월 ~ 현재 : 한국디지털정책학회 회장
- 2012년 1월 ~ 현재 : 스마트융합학술전국연합 의장
- 관심분야 : 디지털정책&스마트융합, 디지털경제민주화 행정 혁신, 빅데이터
- E-Mail : ksnoh@sunmoon.ac.kr