# 빅데이터 보안 강화 및 활용 방안

강정구*, 박석천**, 김종현***
*가천대학교 일반대학원 모바일소프트웨어학과
**가천대학교 컴퓨터공학과 정교수(교신저자)
***위세아이텍 대표이사
e-mail : webclub@nate.com

# Reenforcement and Utilization Method of BigData Security

Jung-Ku Kang*, Seok-Cheon Park**, Jong-Hyun Kim***
*Dept of Mobile Software, Gachon University
**Dept of Computer Engineering, Gachon University(Corresponding Author)
***Representative Director, WISEITECH co., ltd

요 약

최근 데이터의 양이 기하급수적으로 증가하면서 빅데이터가 이슈가 되어 많은 관심을 받고 있는 현실이다. 현재 빅데이터의 기술은 데이터 추출과 이용에만 초점이 맞춰져 있어 보안에 취약한 시스템에 해킹시도가 있을 경우 개인과 기업에 막대한 피해가 발생될 수 있다. 따라서 본 논문에서는 빅데이터 보안에 초점을 두어 외부로부터 피해를 방지하고 안전하게 빅데이터 서버를 운영하는 방법을 제시한다. 즉, Iptable을 이용한 IP나 포트 허용 여부를 지정하고 가상사설망(VPN)을 이용하여 외부 접속을 방지하며 패스워드 강화를 통해 빅데이터 서버의 보안 강화 및 활용 방안을 제시하였다.

## Ⅰ. 서 론

컴퓨터와 인터넷, 그리고 스마트폰을 포함한 스마트 기기의 보급으로 데이터가 급격히 증가하고 있다. 최근 소셜 네트워크 서비스(SNS)의 이용자가 늘어남에 따라 매달 300억 건의 새로운 콘텐츠가 페이스북에 추가되며, 매일 1.4억 개의 트윗이 전송되고, 매분 35시간 분량의 비디오가 유튜브에 업로드 되고 있다.

이러한 많은 양의 데이터를 관리하기 위해 빅데이터 추출 방법을 이용하고 있다. 빅데이터는 크게 추출, 분류, 통계, 저장 등이 있는데 사용자는 이 기술에만 집중하고 있는 상황이다.

하지만, 이렇게 기술에만 집중하다가 빅데이터를 운영하는 서버에 해킹이나 악성코드와 같은 공격에서는 취약할 수밖에 없다. 특히, 사이버 테러의 위험성은 우리가 생각하는 것보다 훨씬 더 심각하며, 최근에는 조직적인 해커 그룹이 특정 표적을 치밀하게 계획적으로 해킹함으로써 주요 정보 유출, 제어시스템 공격, 사이버 무기화 등을 통해 사회적 혼란을 야기하고, 나아가서 국가 안보를 위협하는 수준에까지 이르고 있다[1].

이처럼, 빅데이터 보안은 매우 중요하다. 현재 사용하는 기술에만 초점을 두었다가 취약한 보안으로 인해 정보를 탈취하기 위한 공격이 발생했을 경우 서버가 노출 되어 피해를 볼 수밖에 없다.

따라서, 빅데이터 서버에 보안을 실시함으로써 공격에 대비하여 안전하게 데이터를 관리 할 수 있도록 보안에 힘써야 한다.

본 논문에서는 Iptable을 이용한 IP나 포트 허용 여부를 지정하고 가상사설망(VPN)을 이용해 외부 접속을 방지하며 패스워드 강화를 통해 빅데이터 서버 보안을 강화하는 방법 및 활용 방안을 제시한다.

## Ⅱ. 관 련 연 구

### 2.1 빅데이터 개념

최근 몇 년간 빅데이터가 IT업계의 화두가 되어 왔으나, 아직 빅데이터에 대한 단일한 개념이 정립된 것은 아니다. 기존 논의에서는 대략 세 가지 정도의 개념이 주로 언급되어 왔다. Gartner (2012. 6)는 빅데이터를 '향상된 시사점(Insight)과 더 나은 의사 결정을 위해 사용되는 비용 효율이 높고, 혁신적이며, 대용량, 고속 및 다양성의 특성을 가진 정보 자산'이라고 정의하였다[2].

McKinsey는 데이터베이스의 규모에 초점을 맞추어, '일반적인 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터'라고 정의하였다. 마지막으로 IDC는 데이터베이스가 아닌 업무수행에 초점을 맞추어, '다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지

원하도록 고안된 차세대 기술 및 아키텍처'로 개념화하였다[2].

빅데이터는 데이터의 규모가 방대하고(Volume), 데이터의 종류가 다양하며(Variety), 데이터 처리 및 분석을 적시에 해결해야 하는(Velocity) 특성을 가지고 있으며, 그 결과로 새로운 가치를 창출해 낼 수 있어야 한다. 빅데이터는 일반적으로 데이터베이스로 저장, 관리, 분석할 수 있는 한계를 넘어서며, 기업정보, 웹, 이미지/동영상, SNS, 센서 스트림 등 정형/비정형 데이터를 모두 포함하고, 분석과 예측에 있어서 실시간 처리 등 적시성을 요구한다[4].

아래의 표 2.1은 빅데이터 정의를 표로 정리하였다.

<표 2.1> 빅데이터 정의 표

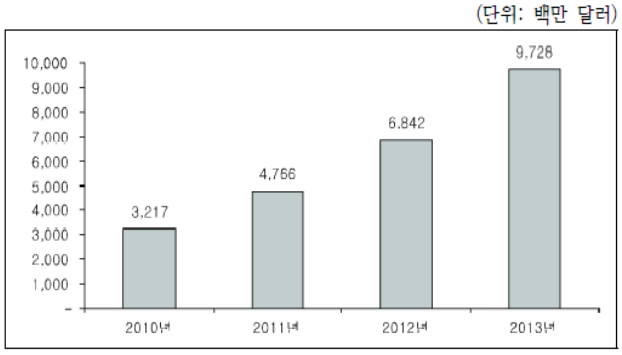
| 기관 | 빅데이터 정의 |
|---|---|
| Gartner | 향상된 시사점(Insight)과 더 나은 의사결정을 위해 사용되는 비용 효율이 높고, 혁신적이며, 대용량, 고속 및 다양성의 특성을 가진 정보 자산 |
| McKinsey | 일반적 데이터베이스SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터 |
| IDC | 다양한 종류의 대규모 데이터로부터 낮은 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처 |

## 2.2 빅데이터 시장 현황 및 전망

최근 다수의 글로벌 ICT 리서치업체와 소프트웨어 기업들이 2013년 ICT의 핵심기술로 빅데이터를 선정하였다. Gartner는 '2013년 10대 전략 기술 트렌드' 중 하나로 '전략적 빅데이터'를 선정했고, IDC는 빅데이터를 2013년에 IT의 주류가 될 기술로 예상했으며, 국내 삼성 SDS도 빅데이터를 통한 가치창출을 2013년 IT 메가 트렌드 중 하나로 선정했다[3].

이와 같이 국내외 글로벌 ICT 기업들의 빅데이터에 대한 관심이 확대되면서 관련 시장도 대폭 성장할 것으로 예상한다. IDC에 따르면 전 세계 빅데이터 시장 규모는 2012년 68억 달러에서 2013년에는 전년 대비 42% 증가한 97억 달러에 이를 것으로 전망된다[3].

아래의 그림 2.1은 2013년 전 세계 빅데이터 시장 전망한 그림이다.




(그림 2.1) 2013년 전 세계 빅데이터 시장 전망

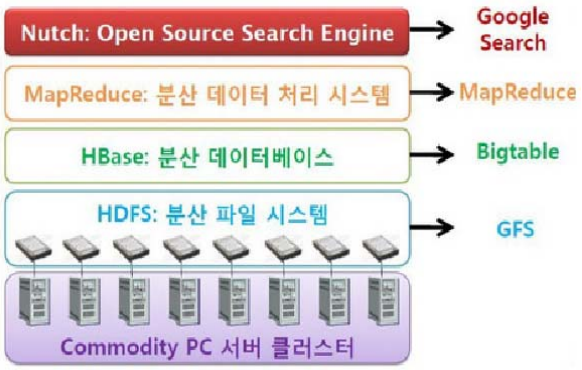
## 2.3 빅데이터 기술

빅 데이터를 데이터 용량에 따른 분류가 아닌 기존 데이터베이스 처리방식으로 해결할 수 없는 데이터의 집합으로 정의하고 이를 처리할 수 있는 기술이나 역량을 보유한 기업이나 국가가 미래의 경쟁력을 갖게 될 것이다[5].

매킨지의 분석에 따르면 전 세계 인구의 60%에 달하는 40억 명이 모바일 폰을 사용하고 있으며 인구의 12% 수준이 보유한 스마트폰은 수년 내에 모든 모바일 폰을 대체할 것이다[5][6].

빅데이터 분석 기법들은 테라바이트 규모의 데이터에 적용되고 있다. 그렇다면 엄청난 규모의 빅 데이터 분석을 수행할 수 있는 인프라 기술에는 하둡(Hadoop), 오픈소스 프로젝트 R, NoSQL 등이 있다[5].

그림 2.2는 하둡 구조에 대응하는 구글 분산처리기술이다.




(그림 2.2) 하둡 구조 대응 구글 분산처리기술

## 2.4 빅데이터 보안

다양한 IT서비스와 플랫폼이 등장하면서 엄청난 양의 데이터가 쏟아지고 있다. 이른바 '빅 데이터(Big Data)' 시대가 도래 하면서 기업들은 빅 데이터 솔루션 도입을 고려하고 있는 추세다. 빅 데이터 도입이 적극 고려되기 시작한 이유는 모바일 기기의 진화와 트위터, 페이스북 등과 같은 소셜 네트워크 서비스의 출현으로 기업 내 데이터를 폭발적으로 증가시켰기 때문이다[7].

다음에서는 빅 데이터의 생성에서부터 서비스에 이르기까지 세 단계로 나누어 보안이슈를 살펴보도록 한다[8].

첫째, 데이터 생성단계는 다양한 경로를 통해 생성, 수집되는 많은 양의 데이터들은 곧 다양한 경로의 보안위협을 의미한다. 최근 장시간에 걸쳐 목적을 가지고 공격하는 지능형 지속 위협(APT, Advanced Persistent Threat) 등이 발생하면서 빅 데이터 생성 및 수집 과정에서 데이터 신뢰성 및 무결성에 대한 우려가 높아지고 있다[8].

또한 빅 데이터들은 많은 수가 개인 IT단말을 통해 생성되어 수집되는데 이때 의도하지 않게 개인정보가 노출되

거나 개인 데이터가 무분별하게 상업적으로 이용되는 등 프라이버시를 침해할 수 있다[8].

둘째, 데이터 저장·운영 단계는 빅 데이터가 생성되어 저장, 분석되어 서비스로 제공되기까지의 일련의 과정 중 가장 보안에 주의해야 하는 구간이 바로 빅 데이터의 저장 및 운영구간이다[8].

여기서는 다양한 사용자를 수용하는 클라우드 컴퓨팅을 활용하는 빅 데이터가 내·외부의 다양한 공격자에게 노출될 수 있어 사용자 인증 및 접근제어, 데이터 기밀성·무결성, 프라이버시 침해, 재해·물리적 침입, 네트워크 보안 등의 문제가 발생할 수 있다[8].

셋째, 서비스 단계는 1차적으로 모여진 많은 양의 데이터를 산업별, 이용자별 각 필요와 요구에 따라 분석하는 과정은 빅 데이터 서비스를 위해 반드시 거쳐야 하는 절차이다[8].

이 과정에서 이전의 암호화 등을 통해 데이터의 기밀성과 익명화 과정을 거쳤다고 해도 사용자가 원하는 데이터를 추출하기 위해 데이터의 복호화 등 데이터 복구 과정을 수행하여야 한다. 따라서 분석 및 2차 데이터에서도 프라이버시 침해 및 데이터의 기밀성이 노출될 위험이 있다[8].

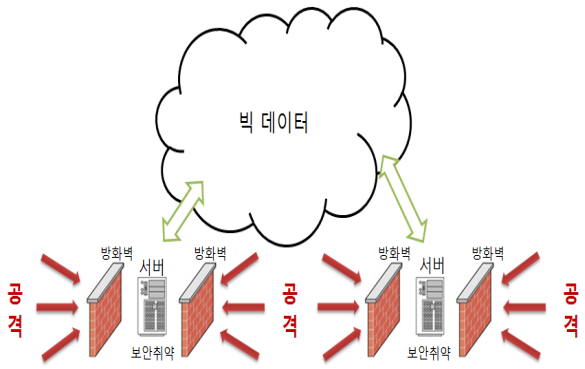
## III. 빅데이터 보안 강화 및 활용 방안

### 3.1 빅데이터 보안 환경

최근에 빅데이터가 이슈가 되어 관심을 받고 있다. 빅데이터를 추출하는 방법, 이용하는 방법 등 많은 분야에서 관심이 있으며 사용하려고 준비하고 있다.

하지만, 빅데이터의 보안에 관해서는 연구가 많이 부족한 상황이다. 어떻게 보안을 해야 하는지 또한 문제가 발생했을 때 해결책 없이 당황하는 경우가 생긴다. 아래의 그림 3.1은 빅데이터의 보안 환경 그림이다. 그림과 같이 빅데이터를 추출하는 서버들은 외부로부터의 공격에 매우 취약하다.

따라서, 본 논문에서는 빅데이터의 구성환경인 리눅스의 방화벽에 보안을 강화하여 외부로부터의 침입을 방지함으로써 사전에 보안 관리를 실시하는 것을 제안한다.




(그림 3.1) 빅데이터 보안 환경

### 3.3 빅데이터 보안 강화 방법

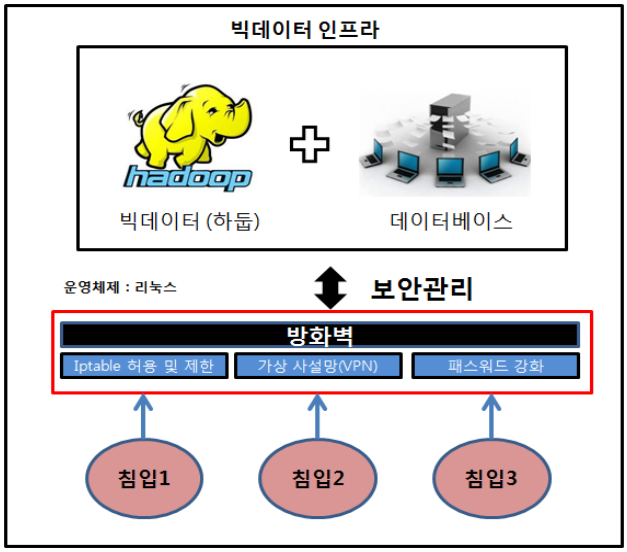
빅데이터 하둡을 이용하기 위해서는 리눅스 환경으로 구성을 해야 한다. 그림 3.2는 빅데이터 보안 방법 구조도이다. 아래의 그림 3.2처럼 빅데이터에 취약한 부분을 보완하기 위해서 방화벽을 구성하는데 여기에 크게 3가지 방법을 사용한다.

첫째, Iptable을 이용하여 IP, 포트 등 제한을 두어 침입을 방지한다. 미리 허용되어 있는 IP나 포트에 관해서만 빅데이터 서버에 접근할 수 있도록 하고 다른 IP나 포트가 접속을 하면 제한을 두어 접근하지 못하도록 구성한다.

둘째, 가상사설망(VPN)을 통하여 보안을 하는 방법이다. 사내 내부에서만 서버에 접속을 할 수 있도록 하여 외부에서는 서버를 찾지 못하게 함으로써 보안을 강화 할 수 있다.

셋째, 패스워드를 통한 빅데이터 서버를 보완한다. 패스워드 규칙을 숫자, 영문, 특수문자 모두 조합 및 사용하도록 기준 한다. 또한, 최소 10개 이상의 패스워드 길이를 지정을 하여 복잡한 패스워드를 통해 침입을 방지함으로 보안이 강화된다.

따라서, 본 논문에서 제안하는 보안 방법을 통하여 외부의 침입을 사전에 방지하여 보안을 강화할 수 있다.




(그림 3.2) 빅데이터 보안 방법 구조도

### 3.3 빅데이터 보안 활용 방안

빅데이터의 활용은 주로 데이터를 추출하는 방법과 통계하는 방법, 이용하는 기술에만 초점이 맞추어져 있다. 즉, 빅데이터의 발전 초기 단계라서 기술을 이용하는 것에 집중되어 있는 것이 현실이다.

하지만, 이렇게 기술을 활용하는 것에만 신경을 써 보안에 신경을 쓰지 않다가 해킹을 당했을 때 심각한 피해를 볼 수가 있다.

따라서, 본 논문에서 제안한 빅데이터 보안 강화 방법은 Iptable을 이용한 IP나 포트 제한을 두어 침입을 방지하고, 가상사설망(VPN)을 통하여 사내 내부에서만 서버에 접속할 수 있도록 하며, 패스워드 규칙을 강화하여 침입을 미연에 방지하도록 하였다. 이와 같이 빅데이터를 이용하는 많은 기업이나 사용자에게 본 논문에서 제안하는 보안 강화 방법을 활용함으로써 빅데이터를 안전하게 사용할 수 있으며 효율적으로 빅데이터를 관리 할 수 있다.

## Ⅳ. 결 론

오늘날 데이터의 양이 기하급수적으로 증가 하면서 빅데이터가 이슈가 되어 많은 관심을 받고 있다. 현재 빅데이터의 기술은 데이터 추출과 이용에만 초점이 맞춰져 있는 것이 현실이다.

하지만 취약한 보안 시스템에 해킹 시도가 있을 경우 개인과 기업에 막대한 피해를 발생 할 수 있다. 이러한 피해가 발생함으로써 앞으로 빅데이터 보안에 더욱 많은 연구가 필요하다.

따라서, 본 논문에서 제안하는 빅데이터 보안 강화 및 활용 방안은 빅데이터 서버에 방화벽을 구성하여 Iptable을 이용한 IP나 포트 허용 여부를 지정하고 가상사설망(VPN)을 이용 외부 접속을 방지한다. 또한 패스워드 강화를 통해 빅데이터 서버 보안을 강화하였다.

향후, 제안한 방법을 기반으로 빅데이터 보안 시스템을 구현하고 테스트를 진행하여 생성된 결과를 바탕으로 제안 시스템을 수정 보완할 예정이다.

## 사사의 글

본 연구는 2013년도 지식 경제부의 SW전문인력양성사업의 재원으로 정보통신산업진흥원의 고용계약형 SW석사과정 지원사업(HB301-13-1003)으로부터 지원받아 수행되었습니다.

## 참고문헌

[1] 김종현, 임선희, 김익균, 조현숙, 노병규 “빅데이터 활용한 사이버 보안 기술 동향”, 2013

[2] 배동민, 박현수, 오기환 ″빅데이터 동향 및 정책 시사점″, 2013

[3] 정부연 ″빅데이터 시장의 현황 및 전망″, 2013

[4] 안창원, 황승구 “빅 데이터 기술과 주요 이슈”, 2012

[5] 김정숙 “빅 데이터 활용과 관련기술 고찰”, 2012

[6] 이만재 “빅 데이터와 공공 데이터 활용”, 2011

[7] 이기주 “스마트 사회의 보안위협과 정보보호 정책추진에 관한 제언”, 2013

[8] 정교일, 박한나, 정부금, 장종수, 정명애 “빅 데이터와 정보보안”, 2012