빅데이터 분석 기술과 사이버 보안

김익균 ETRI 네트워크보안연구실 실장

1. 머리말

과거 2000년 이후 2010년대 초반까지의 네트워크 보안기술은 방화벽, IPS, WAF 등 네트워크 경계 영역 보안장비들의 성능향상에 집중되어 기술 성장을 거듭하였다. 네트워크 패킷 처리기술의 성능을 고도화하기 위하여 네트워크 프로세서, FPGA, 멀티코어 프로세서, 전용 ASIC 구현 기술을 도입하였고, 현재 10G급 이상의 고성능 처리 플랫폼이 개발되어 네트워크 경계영역에서 운영되고 있다. 이는 네트워크 전달 기술의 급속한 성장에 따라 관련 보안기술이 그 성능을 만족시켜야 할 상황에서 필연적으로 전개된 부분이고, 앞으로도 40G, 100G 등 초고속급의 성능 향상이 지속적으로 요구될 것으로 예상한다. 하지만 초고속 네트워크 인터페이스의 성능을 만족시키는 고속화 기술이 아무리 발전하더라도, 경계망보안(Perimeter Security) 기술로는 지능형 표적 공격[1](APT:Advanced Persistent Threat) 등의 이슈를 근본적으로 해결하지 못하기 때문에 네트워크 보안장비에 대한 성능 향상 노력과는 별개로 또 다른 대안을 찾아야 할 상황에 직면해 있다.

2010년대 초반부터 최근까지 발생하는 사이버 위협들은 내부자의 위협, 정상적인 네트워크 서비스를 통한 악성코드 감염, E-메일 및 USB 등을 통한 감염 등 내부망 자체는 언제, 어떤 경로를 통해서든지 다양한 방법으로 침투될 수 있다고 할 수 있다. 이에, 내부 네트워크에서 발생하는 다양한 시스템 로그 정보와 동적 행위 정보를 수집하여 분석하는 내부망 행위 분석 기술이 주목을 받고 있다.

1) 지능형 표적 공격(APT: Advanced Persistent Threat): 사이버 테러의 목적으로 공격 표적으로 한 기업이나 기관 등 조직의 네트워크에 다양한 방법으로 은밀하게 침투해 오랫동안 잠복하면서 기밀정보를 유출하거나 주요 시설의 제어 능력을 확보하는 것을 목표로 하는 전문적인 해킹기법의 지능형 사이버 공격

이와 더불어 최근, 사이버 위협에 대한 새로운 개념으로서 가트너 그룹이 정의한 시큐리티 인텔리전스(Security Intelligence) 분야가 주요 대안으로 주목받고 있으며, 그 기반 기술로서 장기간 누적되어 있는 다양한 종류의 대용량 데이터에 대하여 빅데이터 처리/분석기술을 활용하고 있다. 이는 APT 공격과 같이 알려지지 않은 보안위협을 사전에 예측하고 방어하는 데 초점을 맞추고 있고 향후 5년에서 10년간 사이버 방어기술의 핵심 개념으로 위치할 것으로 평가되고 있다. 본 고에서는 빅데이터 분석 기술을 근간으로 하는 시큐리티 인텔리전스 분야의 동향과 주요기술에 대하여 살펴보고 향후 발전 방향에 대하여 논하고자 한다.

2. 사이버 위협 동향과 방어기술의 수준

2.1 사이버 위협 동향

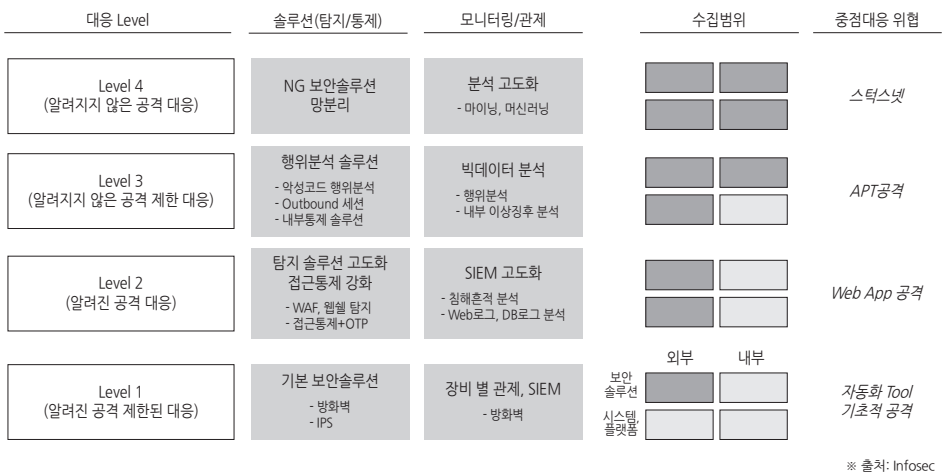
최근 사이버테러, 사이버戰, 핵티비즘 등의 공격방법으로 활용되고 있는 APT 공격은 '목표대상이 명확한 조직적 공격'으로써 주로 정부나 기업을 대상으로 산업기밀이나 군사기밀, 고객정보 등의 정보탈취를 목적으로 하고 있다. 지난해 3월 20일 국내 금융권(농협, 신한은행, 제주은행) 및 방송사(KBS, MBC, YTN)를 대상으로 전산 장애가 동시 다발적으로 발생하였다. 이른바 3.20 사이버 테러로 인하여 총 3만 2천 대의 PC가 피해를 입은 것으로 보고되고 있다. 이러한 사이버 테러 발생은 어제오늘의 일이 아니며, 그 위험성은 우리가 생각하는 것보다 훨씬 더 심각하다. 2010년 세상에 알려진 악성코드인 스틱스넷(Stuxnet)은 이란 원자력 발전 시설에 대한 해킹에 활용되어 원심분리기 중 20%가 가동이 중단되는 등 큰 사회적 파장을 일으킨 바 있고, 이듬해인 2011년에 발견된 나이트 드래곤(Night Dragon)은 미국 글로벌 에너지 기업들에 대한 공격에 활용되어 가스와 석유 분야의 중요 정보들이 유출된 바 있다. 또한, 같은 해 3월에는 150여 명의 외교관 컴퓨터가 공격받아, 프랑스 정부가 보유한 파리 G20 관련 파일들이 유출된 사고가 있었고, 방위산업 업체인 록히드 마틴(Lockheed Martin)을 공격한 사건을 분석하던 중에 국내 통신 업체를 포함한 총 760여 개의 세계적으로 유명한 기업들을 대상으로 공격이 진행되고 있음이 밝혀지기도 하였다.

국내에서는 2009년 발생한 7.7 DDoS 공격 이후, 농협 전산망 장애 공격이 크게 사회적 이슈화된 바 있고, 2011년 7월에 네이트 및 싸이월드가 사이버테러 공격을 받아 회원 3천 5백만 명의 개인정보가 유출된 SK컴즈 사건과 2013년도에 발생한 3.20 방송금융 전산망 마비 사건은 전문적인 해커집단이 개입된 가장 대표적인 APT 공격으로 분류된다.

APT 공격은 기존의 개인적인 목표로 이루어진 공격들에 비해서 사전조사, 알려지지 않은 제로데이 공격, 사회 공학적 기법, 정상 사용자로의 은닉, 권한상승, 탐지 기능에 대비하는 적응, 지속적 공격관리 등의 기법들을 복합적으로 사용하므로 기존 방어기술로 대응하기에 한계가 있다.

2.2 현 보안기술의 수준

현재 사용되고 있는 네트워크 기반 침입탐지 시스템의 대부분은 공격 특징을 분석하여 자동으로 감지할 수 있는 패턴 형태로 만들고, 네트워크 패킷을 DPI(Deep Packet Inspection) 기술로 실시간으로 검색하여 동일한 패턴을 탐지하는 오용탐지(Misuse Detection) 기법이 대부분이다. 이러한 방식은 비교적 높은 탐지율을 보여주면서 실제 사용 측면에서 효율적이라고 할 수 있다. 그러나 전문 분석가 그룹에 의하여 생산되는 탐지용 패턴(시그니처)을



※ 출처: Infosec

[그림 1] 사이버 위협 대응기술 수준

필요로 하기에 패턴의 생성과 관리 측면에서 높은 비용이 발생되고 탐지용 패턴을 우회하는 공격이나 새로운 유형의 공격에 대해서는 효력을 볼 수 없는 것이다. 최근까지 이러한 오용탐지 기법의 정해진 탐지 규칙을 고성능으로 찾아내는 기술에 집중 투자되었고 네트워크 인터페이스 규격 측면에서는 10Gbps 이상 트래픽을 선로 속도로 패턴 매칭을 처리하는 수준에 이르고 있다.

[그림 1]과 같이 사이버 위협 대응기술의 수준을 4단계로 나누어 보았을 때, 현재의 대응기술의 수준은 단계 3의 초반에 있는 것으로 판단된다. 단계 1과 2 수준에서는 앞서 언급되었듯이 IPS 같은 보안 솔루션들이 네트워크 경계 영역 보안을 위하여 오용탐지기법을 고성능화하는데 집중 개발되었으나, 최근에는 APT 등 전문 해킹 공격의 위협이 증가함에 따라 기존의 보안 제품들이 활용하고 있는 패턴 기반의 공격 제어 기법의 한계를 넘어서 내부 네트워크의 다양한 특성 인자들(시스템 프로세스, 활동성, 네트워크 트랜잭션 등)의 관계성 분석을 통하여 알려지지 않은 새로운 공격을 탐지하는 기술 개발이 주목을 받고 있는 상태이다. APT 공격과 같은 알려지지 않은 치명적인 공격에 대응하기 위해서는 주요 IT 기반 주요시설의 네트워크, 시스템, 응용서비스 등으로부터 발생하는 데이터 및 보안이벤트의 연관성을 분석(Security Analytics)하여 보안 지능을 향상하는 차세대 보안정보 분석 기술이 필요하며 이는 빅데이터 분석 기술을 활용한 시큐리티 인텔리전스 제품군으로 주목을 받고 있다.

IT 시장조사 전문업체 가트너 그룹은 시큐리티 인텔리전스 제품군을 새롭게 정의하고 있고 이를 향후 5년에서 10년간 지속될 보안기술로 평가하고 있다[1]. 시큐리티 인텔리전스란 빅데이터 분석 방법을 활용해 주요 IT 기반 시설의 네트워크, 시스템, 응용서비스 등으로부터 발생하는 데이터 및 보안 이벤트 간의 연관성을 분석함으로써 지능적으로 보안 위협에 대응하는 보안 기법을 말한다. 시큐리티 인텔리전스 개념의 차세대 보안기술의 큰 방향성은 외부위협 탐지 중심에서 내부 영역까지 확대, 패턴 중심 탐지에서 다양한 행위 분석으로 확대, 실시간 탐지 중심에서 사후 분석까지 확대, 보안 이벤트 탐지/분석 중심에서 포렌식 분석으로 확대되는 경향으로 기술개발이 이루어지고 있다. 다음 장은

시큐리티 인텔리전스에서 활용하는 빅데이터 처리 및 분석의 기반기술에 대하여 살펴본다.

3. 빅데이터와 사이버 보안

3.1 빅데이터와 그 활용

빅데이터는 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처로 정의된다. 2011년도 매켄지에서 인터넷 데이터의 증가가 전 세계에 미치는 영향에 대한 보고서를 빅데이터라는 이름으로 발표한 이래[2], 다양한 분야에서 빅데이터 분석을 이용한 응용을 내놓고 있다. 기업에서 데이터를 수집, 정리, 분석하고 활용하여 효율적인 의사결정을 할 수 있는 방법에 대해 연구하는 학문인 비즈니스 인텔리전스 분야, 기업 마케팅을 위한 고객정보 분석 이외에도 국가 안전, 재난재해 대비를 위한 예측기술로도 확대되고 있다.

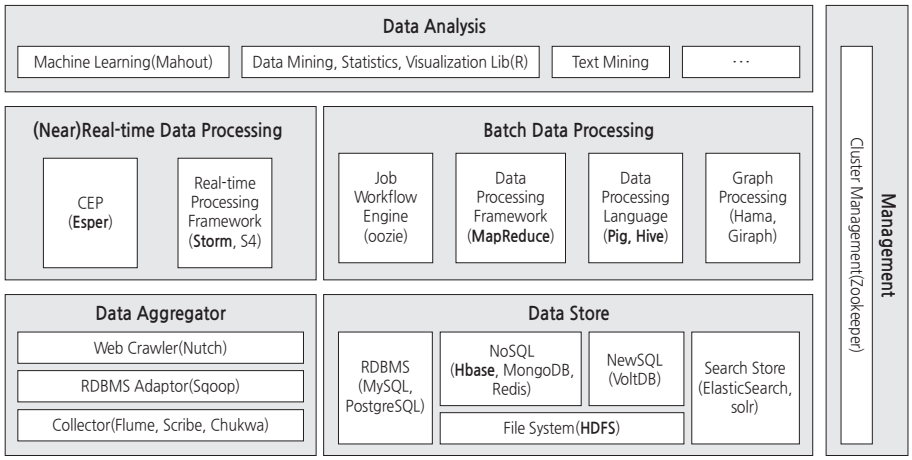
국가 안전을 위협하는 글로벌 요인이나 질병, 위기 등을 분석하여 선제 대응하기 위해 기후변화, 해상오염방지, 방사능 유출탐지 분야 등 광범위한 지역에 걸쳐 생성되는 대용량 정보를 실시간 처리하는 기술이 적용되고 있고 미국, 영국, 일본 등 방재 선진국들은 첨단장비나 센서를 이용한 다양한 감시체제를 연계 구축하여 재난 예방 상호 협력에 활용하고 있다. 예를 들면, 국가 재난을 대비하기 위해 태풍, 강수량, 지진 등의 정보를 분석하여 쓰나미, 홍수 등의 가능성을 예측하는 서비스의 일환으로 IBM은 물 보유량과 기상 데이터, 현재 강수량 등을 종합하여 홍수, 가뭄 대책 수립에 활용하고 있다[4]. 또한, '테라데이타'와 SAS는 데이터 웨어하우징 기술과 분석기술을 결합하여 '자금세탁방지', '신용위험관리' 등 위험관리와 금융사기 예방을 위한 솔루션 공급하기도 했다[3]. 한편, 세계적으로 국가안전(Homeland Security)에 대한 관심 및 투자 증대에 따라, 미국은 홈랜드 안보를 위해 소셜네트워크, 신문, 잡지, 기사 등으로부터 수집된 정보를 분석하여 테러 동향을 파악, 테러 징후 사전 예측하는 기술을 개발하고 있고 이와 더불어, 전통적인 전쟁에 정치, 경제, 사회문화적 요소와 사이버공간 등이 추가되어 전쟁과 범죄가 융합되는 현상인 5세대 전쟁으로 정의되는 안보위협을 네트워크 분석을 통해 해결하고자 노력 중이다.

특히, 사이버 보안영역에서의 빅데이터 분석 활용도는 대용량 로그 분석, 비정상 트랜잭션 및 행위탐지, 악성코드 탐지 등에 활용된다. 대표적인 사례로써, Sourcefire에서는 200만 개의 단말을 모니터링하여 악성코드가 있는지를 탐지하기 위해 Hadoop 플랫폼을 이용하고 있고[5], EMC RSA에서는 빅데이터 분석으로 지능형 리스크 관리를 통해 기민한 대응을 위해 문맥 또는 상황을 인지하는 기술을 적용하고 있다[6]. 아트 코비엘로 EMC RSA 정보보안 사업부문 사장은 현재 100% 완벽한 사이버 보안은 없으며, 과거의 기술과 사고의 연장선상에서 대응해서는 안 되고, 보다 창의적인 보안 대응방법이 필요하며, 이를 위해서는 빅데이터 분석을 중심으로 한 지능형 보안 시스템 구축이 반드시 필요하다고 밝힌 바 있다[7].

다음 절에서는 앞서 언급된 다양한 빅데이터 응용분야에서 공통적으로 사용되는 빅데이터 처리 플랫폼에 대하여 간략히 살펴보도록 한다.

3.2. 빅데이터 처리 플랫폼 및 분석기술

빅데이터 응용 프로그램은 기존의 데이터 웨어하우스 작업이 특정 시간 간격에 대한 데이터를 유지하는



[그림 2] 오픈소스 기반 빅데이터 처리 플랫폼

반면 장기간 이력을 분석하기 위해 아주 긴 기간 동안의 데이터를 유지한다. 이 같은 하둡(Hadoop) 생태계의 NoSQL 데이터베이스와 같은 데이터 처리 도구는 복잡한 쿼리 및 분석의 처리 속도를 증가시키는 기술을 제공하게 되었다. 기존의 데이터웨어하우스 처리/분석 과정에서 새로운 스키마를 통합하는 것은 어려웠던 반면 빅데이터 툴을 통해 사용자는 미리 정의 된 형식 이외에도 다양한 형식의 구조적 및 비구조적 데이터를 로드할 수 있고 데이터를 사용하는 방법에 유용한 선택을 할 수 있다.

[그림 2]에서 보는 바와 같이 빅데이터 처리 플랫폼은 저장 데이터에 대한 배치 처리와 실시간 스트림 데이터에 대한 분석이 가능하다. 하둡은 배치 처리를 위한 가장 유명한 기술 중 하나이다. 하둡 프레임워크는 대용량 파일에 대하여 분산, 병렬 처리할 수 있는 대규모 데이터 처리 문제를 조정하는 맵-리듀스 프로그래밍 모델을 개발자에게 제공한다. 구글은 초기에 맵-리듀스라고하는 프로그래밍 모델과 대용량 데이터 분산처리 프레임워크와 대용량 데이터를 효과적으로 저장하고 확장할 수 있는 GFS(구글 파일시스템) 기술을 확보하고 이를 적극적으로 활용하고 있었고, 이를 바탕으로 구글만의 검색기술과 검색서비스를 가능하게 한 것이다.

구글이 가진 기술을 참고해서 등장한 다양한 맵-리듀스 프레임워크 중에서 가장 주목을 받으면서 그 기반으로 에코시스템을 갖추게 된 것이 자바 기반의 아파치 하둡(Apache Hadoop)이다. 하둡은 크게 두 개의 요소로 나뉘어 있다. 하나는 맵-리듀스 프레임워크(MapReduce Framework)와 하둡 분산 파일시스템(HDFS)이다. 초기에는 하둡을 이용해서 대용량 데이터 분석을 위해서는 자바언어를 이용해서 직접 프로그래밍을 해야 했지만 하둡으로 데이터 분석 로직을 손쉽게 구현할 수 있는 프로세싱언어인 PIG와 SQL과 같은 언어를 제공하는 HIVE가 등장하였다. 최근엔 오픈 소스 통계 툴로 유명한 R이 하둡과 연동되면서 하둡을 중심으로 대용량 데이터 분석에 필요한 다양한 기술들이 통합되고 응용되면서 하나의 에코시스템을 이루어가고 있고 관련 솔루션 업체, 스타트업들이 많이들 등장하고 있다.

대용량 데이터의 분석을 위해서는 앞서 언급한 분산처리 프레임워크와 분산파일 시스템도 중요하지만 이러한 컴퓨팅 환경에서 데이터 분석을 효율

적으로 할 수 있는 처리할 수 있는 확장성 있는 분석 기법과 알고리즘의 확보가 매우 중요하다. 예를 들면, 아파치 마하웃(Apache Mahout) 프로젝트는 다양한 주요 마이닝 알고리즘들을 하둡 프레임워크 상에서 구현하여 오픈 소스로 공유하자는 차원에서 만들어졌다.

3.3 사이버보안과 마이닝 기술

사이버보안 분야에서 기존 네트워크 침입 탐지 방법들의 문제를 해결하기 위해서는 각종 인공지능(Artificial Intelligence)기법들이 도입되고 있다. 즉, 정상적인 행위와 각종 비정상적인 행위를 수집하고, 이 행위에 다양한 기계학습(Machine Learning) 알고리즘을 적용하여 지식을 자동으로 생성한다. 그리고 이렇게 학습된 지식을 기반으로 실시간으로 발생하는 이벤트들에 대하여 정상 및 비정상 여부를 판단하는 문제로 볼 수 있다. 기계학습은 그 학습 방법에 따라 크게 감독 학습(Supervised Learning)과 비감독 학습(Unsupervised Learning)으로 나누어지는데, 감독학습은 해당 자료에 대한 지식을 기반으로 학습하는 방법이며, 비감독 학습은 이러한 지식의 도움 없이 스스로 학습하는 방법이다. 감독 학습 방법에 의해 학습된 지식은 학습 알고리즘의 특성에 따라 트리, 인스턴스, 가중치, IF-THEN 규칙 등 다양한 형태로 표현된다. 또한, 비감독 학습에 의해 만들어진 지식은 동일한 특성이 있는 자료들의 그룹(Cluster) 형태로 지식이 표현된다. 비정상 행위 탐지 기반 침입탐지 분야에서는 감독 학습법과 비감독 학습법이 모두 사용되고 있는데, 감독 학습법이 비감독 학습법에 비해 정확도가 높다. 그러나 비감독 학습은 비록 정확도는 낮지만, 새로운 유형의 공격 감지에서는 효과적이라고 알려져 있다.

4. 빅데이터 분석 기술을 활용한 사이버 보안 연구동향

4.1 봇넷 식별을 위한 NetFlow 모니터링: BotCloud 프로젝트

BotCloud 연구 프로젝트[8]는 봇넷에 감염된 호스트를 식별하기 위하여 대용량의 NetFlow 데이터를 분석할 수 있는 맵-리듀스 패러다임을 연구하였다. BotCloud는 봇넷의 명령 및 제어(C&C) 채널을 추적하는 페이지 랭크(PageRank) 및 클러스터링 알고리즘의 조합을 사용하여 호스트 관계를 조사하는 BotTrack에 의존한다. 종속성 그래프 생성, 페이지 랭크 알고리즘 및 DBScan 클러스터링 등의 방법을 이용하여 봇넷 탐지를 수행한다.

4.2 BEEHIVE[9]: APT 탐지를 위한 행동 프로파일

APT 공격에 대한 관찰 측면에서 볼 때, 미국 RSA 연구소에서는 감염된 시스템의 행동은 민감한 정보를 훔치거나 시스템 작업을 파괴하기에 시스템의 평소 행위 패턴과 차별화가 될 수 있다고 가정한다. APT 공격은 여러 단계로 구성되어 있기 때문에 공격자에 의해 각 작업은 표준 행위에서 편차를 감지할 수 있다고 판단된다. 겉보기에 독립적인 이벤트의 상관관계는 이전의 방법으로 식별할 수 없었던 은밀한 공격에 대하여 증거를 확보할 수 있다. 행위 편차의 탐지는 호스트 또는 기업의 네트워크 내에서 사용자의 활동의 한 측면을 검사하게 된다. APT 공격 감지뿐만 아니라, 행동 프로파일은 상호 작용을 기반으로 사용자를 인증하고, IT 인프라의 중요한 서비스 및 사용 패턴을 검사하여 조직 내에서 권한이 없는 IT 인프라의 식별, 관리 및 동작 기반 인증 등이 가능하다. 따라서 이런 기술은 조직 환경 운용 관리에 대한 통찰력을 제공하게 된다고 볼 수 있다.

4.3 보안 빅 데이터 분석 실험을 위한 WINE 플랫폼[10]

세계 정보 네트워크 환경(WINE)은 시만텍에서 수집된 필드 데이터(안티바이러스 원격 측정 및 파일 다운로드 등)를 사용하여, 대규모의 데이터 분석을 수행하기 위한 플랫폼을 제공하고 실험 방법을 추진하고 있다. WINE은 전 세계의 호스트 수백만에서 발생하는 데이터를 수집하고 샘플링 및 집계한다. 예를 들어, 실제 데이터에 대한 새로운 아이디어를 검증하는 실험 연구를 수행하고, WINE에 저장 및 참조되는 데이터 집합에 대해 서로 다른 알고리즘의 성능을 비교하기 위해 개방된 형태로 재현 실험을 수행할 수 있다. 하지만 현재 WINE 프로젝트 환경은 시만텍의 엔지니어 및 학술 연구자에 의해서만 사용되고 있다.

5. 맺음말

최근의 사이버 위협은 사회적 혼란을 야기하고 국가 안보를 위협하며 개인에게 금전적으로 피해를 주는 등 다양한 목적으로 자행되고 있다. 다양한 사이버 테러 기법 분석과 연구 등을 통해 피해를 줄이거나 이를 원천 봉쇄할 수 있는 지능형 보안 기술로써 빅데이터 분석 기술의 통합을 시도하고, 공격 로그 이벤트 및 내부 상황정보 등을 수집하여 상관관계를 분석함으로써 공격자의 의도를 사전에 인지하고 차단할 수 있는 보안 기술이 요구되고 있다. 현재 사이버 보안 영역에서 빅데이터 분석 기술의 활용이 중요한 도전 중의 하나이지만, 빅데이터 분석 기술을 기반으로 하는 시큐리티 인텐리전스 분야의 기술적 성공을 위해서는 몇 가지 해결해야 할 부분이 있다. 이 중 가장 중요한 부분이 실시간으로 실행 가능한 정보를 획득하는 것입니다. 이와 더불어 분석에 사용되는 데이터의 신뢰성 및 무결성이 보장되어야 하고, 또한 수집된 정보에 대하여 개인 정보 보호 정책이 반영되어야 한다.

전 세계적으로 빅데이터 활용도를 극대화하는 연구 경향에 맞추어, 현재 정교해지는 사이버 위협에 대응하는 전통적인 대응 방법들의 한계를 벗어나 빅데이터 분석기법과 같은 새로운 방법론과 접목을 통하여 차세대 보안 기술의 비약적인 발전을 기대해 본다.

[참고문헌]

[1] Lawrence Pingree, Ruggerro Contu, Eric Ahlm, 'Context-Aware Security and Intelligence-Sharing Concepts Merge to Create Intelligence-Aware Security Controls', Gartner Group, March, 2014.

[2] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs R., Roxburgh, C., & Hung Byers, A. (May 2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute (MGI).

[3] Sam Harris, 'Teradata Enterprise Risk Intelligence for Fraud and Financial Crimes Prevention'

[4] 'IBM's flod prediction technology', Analytics and visualization of Big Data, March, 2013 (http://auburnbigdata.blogspot.kr/2013/03/ibms-flood-prediction-technology.html)

[5] Sourcefire FireAMP Brings Big Data Analytics to Enterprise Security, Jan, 2012

[6] EMC RSA Unveils Big Data Enhanced Risk Management Tool, Nov, 2013

[7] RSA's Art Coviello Points to Big Data as Transformative Solution to Security Challenges, Feb, 2013

[8] Jeróme Franc¸ois, Shaonan Wang, Walter Bronzi, Radu State, Thomas Engel, 'BotCloud: Detecting Botnets Using MapReduce', Nov, 2011, Information Forensics and Security (WIFS), 2011 IEEE International Workshop

[9] Ting-Fang Yen, et al 'BEEHIVE: Large-Scale Log Analysis for Detecting suspicious activity in enterprise networks', ACSAC '13, Dec, 2013

[10] Tudor Dumitras, Darren Shou, 'Toward a Standard Benchmark for Computer Security Research', BADGERS'11 10 April 2011

[11] Alvaro A. Cárdenas, Pratyusa K. Manadhata, Sree Rajan, Big Data Analyics for Security Intelligence, Cloud security alliance