빅데이터를 활용한 보안로그시스템

전경식[*] · 이현경[**] · 전삼현[***] · 김종배[****]

[*,**,***,****]숭실대학교

E-mail : postwin@gmail.com, ketia89@naver.com, shchun@ssu.ac.kr, kjb123@ssu.ac.kr

요 약

최근 사이버 공격이사회, 국가적 위협으로 대두되고 있다. 최근 신종 악성코드에 의한 A.P.T 공격이 사회적으로 큰 혼란을 야기하고 있다. 이에 따라 기업 내에서 방화벽, IPS, VPN 등의 네트워크 보안 시스템의 통합 관리를 목적으로 하는 통합관제시스템(ESM)의 필요성이 제기되었다. 그러나 기존의 ESM의 방식은 외부에서 내부로 유입되는 트래픽만을 모니터링하는 네트워크 기반 공격 탐지 기법을 사용하기 때문에, 외부 사이버 공격만을 차단할 수 있다는 한계점을 가지고 있다. 따라서 본 연구는 주요 IT 기반시설의 네트워크, 시스템, 응용 서비스 등으로부터 발생하는 데이터 및 보안 이벤트 간의 연관성을 분석하여 보안 지능을 향상시키는 빅데이터를 활용한 보안로그시스템을 제안한다. 본 연구에서 제안한 빅데이터를 활용한 보안로그시스템을 통해 분산 기반의 저장/처리 기술 적용하고자 한다.본 기술을 적용한 지능형 정보 분석 플랫폼 구성을 통해, 가용성과 확장성을 확보하여 통합적 보안 관제가 가능하도록 한다. 뿐만 아니라 기업 내로의 악성코드 유입, 감염(전파) 그리고 실시간 모니터링이 가능하여 고객 서비스 만족도가 향상되는 파급효과가 기대된다.

키워드

빅데이터, 보안로그시스템, 데이터보안

Ⅰ. 서 론

최근 신종 악성코드에 의한 큰 사회적 혼란을 야기하고 있는 A.P.T공격은 과거 표적형 공격의 한 유형으로 Target 대상 정보를 수집 확보한 후 사용 가능한 모든 방법을 동원하여 목적을 달성하기 때문에 탐지 및 차단이 매우 어려운 상황이다. 이처럼 지능적인 A.P.T공격의 위협요소들로부터 기업의 자산이 외부로 유출되지 않도록 하기 위해서는 기업 내에서 방화벽, IPS, VPN 등의 네트워크 보안 시스템의 통합 관리를 목적으로 하는 통합관제시스템(ESM)이 필요하다. 그러나 기존의 ESM은 외부에서 내부로 유입되는 트래픽만을 모니터링하는 네트워크 기반 공격 탐지기법을 사용하기 때문에, 외부 사이버 공격만을 차단할 수 있다는 한계점을 가지고 있다. 따라서 공격의 형태보다는 시스템, 응용 서비스 등으로부터 발생하는 데이터 및 보안 이벤트 간의 연관성을 분석하여 관리자와 관리자 권한을 도용한 공격자를 구별할 수 있어야 하며, 보안 지능을 향상시키는 빅데이터를 활용한 보안로그시스템을 제안한다.

본 연구에서 제안한 빅데이터를 활용한 보안로그시스템을 통해 분산 기반의 저장/처리 기술을 적용하여 정형/비정형의 방대한 데이터를 신속하게 수집, 분석하고자 한다. 이를 위해 특히, 분석 기반의 다중 검색으로 실시간 분석 성능을 제공하고, 가시적 분석을 위해 유연한 대시보드, 모든 요소 상관분석 기법을 적용한다. 이와 같이 빅데이터를 활용한 보안로그시스템을 적용한다면 가용성과 확장성을 확보하여 통합적 보안 관제가 가능 할 것으로 본다.

Ⅱ. 본 론

기존의 ESM(통합보안관리시스템, Enterprise Security Management)관련 연구는 방화벽, IDS, VPN 및 각종 보안제품 뿐만 아니라 서버, 라우터 등의 네트워크 장비들을 연결하여 실시간으로 전자적 침해에 대해 연구하고 있다. 이들 연구들은 IT 단위시스템 로그통합 및 장애처리 모니터링을 활용하고, IT 시스템과 보안시스템의 연계분석을 통한 보안관제용으로 활용한다. 그러나 이들 연구에서 제시하고 있는 방법들은 수많은 데이터를 저장하는 DB의 성능과 비용의 문제점이 있어 최근 A.P.T공격에 대비하기에는 어려움을 겪고 있다[1]. SIEM은 ESM의 기능을 포함하고, Application 연계를 통한 종합분석으로 활용하여 알려지지 않은 보안 위협 상관분석이 가능한 시스템 모델이다. 즉, 대량의 정형/비정형 데이터를 수집하여 통합적 보안 관제가 가능하도록 하는

것이다[2][3][4].

Ⅲ. 보안로그시스템 아키텍처

지능형 정보 분석 플랫폼은 데이터의 대용량 수집, 저장, 분석, 최적화 요소로 구성되어 있으며 각각의 기능은 고유 체계를 구축하고 있다. 다양한 데이터 Source로부터 안정적으로 데이터를 수집하고, 다중 병렬 구조로 균등하게 데이터를 저장하며, 고속 검색을 기반으로 지능형 분석이 가능한 시스템 구성을 제공한다.

본 연구에서 제시하는 통합수집기는 내부망의 비정상 트래픽 분석 시스템과 웹 이상행위 탐지, DNS싱크홀을 통해 유해사이트 접속을 통제하고 웹셸을 이용하여 웹 해킹을 조기에 탐지할 수 있다. 정보를 수집하기 위해 Agent/Agentless의 두 가지 방식을 사용하고, 실시간 및 안정성을 고려하며 수집방식 선택에 유연성을 더해주고자 한다.

대량의 보안로그 파일 저장은 분산 아키텍처를 적용한다. 분산 아키텍처는 대용량 데이터를 수용하기 위해 병렬처리하고, 분산 기반 다중 인덱서로 실시간 인덱싱 및 저장을 수행한다. 이러한 분산 아키텍처를 적용하여 일 단위 수 TB(Tera byte)에 달하는 용량을 처리할 수 있고, 각 수집기마다 200,000 EPS의 처리성능을 가진다. 특히 각각의 수집기는 데이터를 저장할 때 자동으로 무결성을 체크하고, 데이터의 원본을 압축, 암호화 하여 보관한다. 또한 Multi 시스템에서 발생할 수 있는 결함으로부터 원본 데이터를 자동으로 보호하기 위해 백업/Hot Spare 수집기를 따로 구성하여 데이터를 백업하고, 자동으로 복구한다.

데이터 저장 단계에서 수집기에 저장된 인덱싱 데이터는 검색 키워드 또는 조건을 입력하여 인덱싱 데이터를 찾아내는 방식으로 동작이 가능하다. 검색된 대량의 보안로그 데이터 분석은 문제를 여럿으로 세분화하면서 분석하는 기법인 데이터 드릴다운으로 다중검색을 용이하게 한다. 또한 보안 장비가 발생시키는 데이터는 베이스라인 및 임계값 기반으로 데이터의 급격한 변화를 감지하는 방식과 통계를 기반으로 데이터를 예측하는 추이(Trending)분석을 사용하여 분산기반의 다중 검색으로 실시간 분석 성능을 보장한다.

데이터의 가시적 분석을 위해서 사용자별 다양한 동적 Dashboard를 구성하고, 장비별/로그유형별, 실시간 모니터링하여 모든 이벤트들의 상관관계를 분석하여 직관적으로 도식화한다. 분석 과정을 실시간으로 모니터링하여 분석 데이터에서 이상을 발견했을 시, 경보를 울려 보안 위협을 사용자에게 시각화하여 보여준다. 최대 20억 건까지의 단일 검색을 수행하며, 검색 속도는 단순 검색 조건 기준 1일 200G~400GB 로그 정보에서 1분 이내에 검색이 가능하다.

데이터 수집 및 중계를 위하여 모든 데이터 수집 기술 방식, 대용량 데이터 전송, 운영 안정성 및 고 가용성을 고려한 데이터 수집 체계를 구축해야 한다. 이를 위해 보안 장비에서 발생되는 모든 소스, 모든 포맷 데이터, 정형/비정형 원본로그, 로그 원본을 실시간 수집하여 데이터 전송기를 통해 수집기에 저장하게 된다. 비정형 데이터의 경우 각각의 단말 Adaptor에서 수집된 로그 데이터를 원본의 변형없이 Collector에 전송하는 데이터로, Adaptor의 IP주소와 Security Log, 데이터 수집 날짜와 시간 등의 내용을 Collector에 저장한다. 그 후 통합로그서버의 정규화 과정을 거쳐 특정 인자값 Seperater기술 및 정규표현식을 지원하는 라이브러리인 PCRE(Perl Compatible Regular Expressions) 기술을 통해 정형데이터로 만든다. 이러한 방법을 통해 정형화된 데이터는 서버에 저장된 데이터에 접근한 사용자의 정보와 데이터의 조회, 문서 출력 등의 데이터 조작의 흔적들을 수집하고, 조회, 출력된 데이터의 정보를 출력한다. 구조를 단일화 함으로써 에러 메시지가 출력된 데이터로의 접근이 보다 원활해지고, 사용자에게 데이터 정보를 보다 가시적으로 보여줄 수 있다.

Ⅳ. 결론

빅데이터를 활용한 보안로그 분석을 위한 시스템 구성을 위해서 지능형 정보 분석 플랫폼을 바탕으로 수집, 저장, 처리, 분석 기술을 적용하였다. Agent/Agentless 방식으로 수집한 보안 로그를 통해 가용성과 확장을 확보하여 대량의 데이터를 저장한다. 또한 빠르게 검색 및 분석 가능한 방식으로 시각화 방법을 구현하여 기존에 불가능했던 분석을 가능하도록 시스템을 구성하였다. 뿐만 아니라 기업 내로의 악성코드 유입, 감염(전파) 그리고 실시간 모니터링이 가능하여 고객 서비스 만족도가 향상되는 파급효과가 기대된다.

참고문헌

[1] Lee Seung Ha, Kang Seung Won, Kim KiHong, Pang Sechung, " Design of Big Data ETL Model for Aggregating of Security Log/Event ", KICS, 2014.06

[2] Kelly M. Kavanagh, Mark Nicolett, Oliver Rochford, “magic quadrant for security information and event management”, Gartner Group, 2014.06

[3] M. Nicolett and J. Feiman, “SIEM Enables Enterprise Security Intelligence,” Gartner Group, 2011.01

[4] N. MacDonald, “Information Security Is Becoming a Big Data Analytics Problem,” Gartner Group, 2012.05