Journal of The Korea Institute of Information Security & Cryptology(JKIISC) ISSN 1598-3986(Print)
VOL.23, NO.4, August 2013 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2013.23.4.721

빅데이터 기반의 실시간 네트워크 트래픽 분석 플랫폼 설계

이 동 환,†‡ 박 정 찬, 유 찬 곤, 윤 호 상
국방과학연구소

On the Design of a Big Data based Real-Time Network Traffic Analysis Platform

Donghwan Lee,†‡ Jeong Chan Park, Changon Yu, Hosang Yun
Agency for Defense Development

요 약

빅데이터는 오늘날 가장 각광받고 있는 데이터 수집 및 분석기술의 경향으로, 대량의 비정형 데이터 분석을 요구하는 다양한 분야에 접목되어 효용성을 인정받고 있다. 네트워크 트래픽 분석 역시 대량의 비정형 데이터를 다루는 분야로, 빅데이터 접목시 그 효과가 극대화될 수 있다. 따라서 본 논문에서는 고도의 보안이 요구되는 군 C4I망과 같은 내부망 환경의 침해사고 및 이상행위를 실시간으로 탐지하기 위한 빅데이터 기반의 네트워크 트래픽 분석 플랫폼(RENTAP)을 소개한다. 빅데이터 분석 지원을 위해 최근 각광받고 있는 오픈소스 솔루션들을 대상으로 비교·분석을 수행하였으며, 선정된 솔루션을 기반으로 고안된 최종 설계에 대해서 설명한다.

ABSTRACT

Big data is one of the most spotlighted technological trends in these days, enabling new methods to handle huge volume of complicated data for a broad range of applications. Real-time network traffic analysis essentially deals with big data, which is comprised of different types of log data from various sensors. To tackle this problem, in this paper, we devise a big data based platform, RENTAP, to detect and analyse malicious network traffic. Focused on military network environment such as closed network for C4I systems, leading big data based solutions are evaluated to verify which combination of the solutions is the best design for network traffic analysis platform. Based on the selected solutions, we provide detailed functional design of the suggested platform.

Keywords: Network Traffic Analysis, Malicious Network Traffic, Insider Threat, Big Data

I. 서 론

빅데이터는 최근 많은 기술 및 학문 분야에서 각광받고 있는 분석 기술 경향의 하나로, 대용량의 비정형 데이터를 고속으로 처리할 수 있는 기술 또는 이러한 기술이 소개됨에 따라 기존에는 분석하기 어려웠던 여러 가지 데이터들에 대한 새로운 분석 및 해석이 가능하게 되었음을 의미한다. 이에 따라 다양한 분야에서 빅데이터 수집 및 분석기법을 적용하여 유의미한 분석 결과를 내놓고 있어 앞으로 활용가능성이 매우 높을 것으로 보고 있다. 네트워크 트래픽 수집 및 분석 분야의 경우에도 대용량 로그분석의 효용성에 대한 관심이 점점 높아지고 있는 분야로서[1], 빅데이터 기술 적용 시 정보보호 측면에서 높은 시너지 효과가 예상된다.

한편, 우리 군에서는 최근 적의 C4I 체계, 무기체계 등을 대상으로 하는 사이버전 개념이 대두되면서 군의 내부 네트워크에 대한 외부 침입 및 악의적인 내부자에 의한 사이버 공격의 가능성이 대두되고 있는 실정이다[2]. 특히, C4I 개념의 발달로 많은 무기체

접수일(2012년 12월 6일), 수정일(1차: 2013년 3월 28일, 2차: 2013년 5월 8일), 게재확정일(2013년 5월 22일)
† 주저자, dlee@add.re.kr
‡ 교신저자, dlee@add.re.kr(Corresponding author)

계가 C4I 네트워크에 편입되는 등, 군 작전수행의 많은 부분이 정보체계에 의존하게 됨에 따라, 군 네트워크에 대한 침입이 발생할 경우, 그 피해는 아군의 전력을 심각하게 손상할 수 있을 정도라고 예상할 수 있다.

따라서, 이러한 사이버 공격에 대응하기 위해 우리 군에서는 끊임없이 군 내부 트래픽의 감시 및 대응 수준을 높이려는 노력을 진행 중이다. 군 네트워크의 경우 민간과는 달리, 대부분의 경우 인터넷이나 공공망과는 분리되어 있으며, 라우터 및 통신선로 등의 기반장비를 관련 부대에서 직접 관리하고 있다. 이러한 군 네트워크의 특수성 때문에, 군 내부의 모든 단말의 상태 및 발생 트래픽의 수집이 가능하다. 따라서 이러한 트래픽을 전수 수집하여 대용량 로그를 만들고 이에 대해 실시간으로 처리 및 분석할 수 있는 기술이 있다면, 군 네트워크 내부에 침입한 사이버 공격에 대한 탐지율을 극대화할 수 있을 것이다.

트래픽 수집 및 분석에 관한 기존의 연구들[3-5]은 대체로 침입방지 혹은 탐지시스템에 적용된 기술로, 대용량 트래픽 로그 데이터에 대한 처리 기술의 한계로 인하여 소용량 혹은 단위기간 동안의 데이터에 대한 분석기법에만 집중하기 때문에 APT 공격과 같이 장기간에 걸친 지능적인 공격에 대응하기에는 제한이 있었다. [6]의 경우, 클라우드화와 이에 따른 대용량 로그에 대한 관심이 높아짐에 따라 빅데이터 솔루션(Hadoop)을 적용한 대용량 로그수집 시스템의 구조를 처음으로 제안하였다. 하지만, 해당 연구의 경우, 로그수집구조 제시에 중점을 두어 로그 분석을 위한 검색 기능에 대한 고려가 없는 점, 적용 가능한 다른 빅데이터 솔루션과의 비교분석이 이루어지지 않았다는 점이 한계로 지적될 수 있다. 이후 역시 Hadoop 솔루션을 활용한 시스템 상태 정보 수집 구조에 관한 연구[7]는 있었으나, 트래픽 로그에 대한 전수수집 및 분석 구조에 대하서는 이제까지 소개된 바가 거의 없다.

본 논문에서는 이러한 한계점을 보완하여 군 네트워크 환경에 맞게 개발된 실시간 네트워크 트래픽 분석 플랫폼(RENTAP) 및 그 구조에 대하여 소개한다. 구체적으로는, 먼저 주요 빅데이터 처리 솔루션에 대하여 소개하고, 플랫폼 개발에 적용된 주요 요구사항에 대하여 기술한다. 다음으로, 빅데이터 솔루션들 간의 성능비교 및 선정과정에 대하여 기술한다. 그리고 마지막으로 RENTAP의 최종설계 및 개발결과를 소개한다.

II. 빅데이터 솔루션

위와 같이 요구사항을 종합한 결과, 구축 시스템은 다양한 형식을 지원하고 대용량 데이터에 대한 실시간 처리 기능이 탑재된 빅데이터 기반 플랫폼의 형태로 구축하는 것이 타당하다는 결론을 얻게 되었다. 특히 최근에는 빅데이터 데이터 저장 및 분석을 지원하는 다양한 오픈소스 솔루션이 출시됨에 따라 이를 활용하도록 하였다. 본 절에서는 RENTAP 설계 시 고려된 다양한 빅데이터 기반 솔루션을 소개하고 각각의 특성 및 장점을 소개한다.

2.1 MongoDB

MongoDB [8]는 미국 10gen사(社)의 지원을 받아 개발된 오픈소스 NoSQL(Not Only SQL) 솔루션이다. 현재 빠른 시장 선점을 통해 NoSQL 분야 점유율 1위에 올라있다. 자체 개발 파일 포멧인 BSON(Binary JSON)을 통해 데이터를 저장하고 관리하는 것이 특징이며, 이를 통해 자유로운 스키마 설정을 지원한다. 자체적으로 Sharding을 통한 분산 파일 처리 및 클러스터링을 지원하고 분산-클러스터링 구조에 적합한 정합 및 정렬 기능인 Map-Reduce 기능 역시 지원한다. C++로 개발되었으며 Windows, Linux, Solaris 등 다양한 운영체제를 지원한다.

2.2 HBase/Hadoop

HBase [9]는 ASF(Apache Software Foundation)의 지원을 받아 개발된 오픈소스 NoSQL 솔루션이다. HBase는 구글의 분산-클러스터링 구조 데이터베이스인 BigTable을 모델로 하여 개발되었으며, ASF의 지원을 받아 개발된 분산파일 처리 솔루션인 Hadoop [10]을 기반으로 하여 구동된다. 따라서 HBase는 기본적으로 분산-클러스터링 구조 및 Map-Reduce 기능을 지원하며 이러한 구조에 가장 최적화된 솔루션이라고 볼 수 있다. Java를 기반으로 개발되어 운영체제에 구애받지 않고 설치·사용할 수 있다.

2.3 Lucene/Solr

Lucene [11]은 Doug Cutting에 의해 개발된

오픈소스 검색엔진 솔루션으로서, 최근에는 ASF의 지원하에 개발되고 있다. 검색엔진 용도로 개발된 만큼, Full-Text Indexing, 다양한 문서파일에 대한 파싱 지원 등 다른 솔루션에서 찾기 힘든 강력한 기능들을 제공한다. Solr [12]는 ASF에서 개발하는 Lucene 기반의 통합 검색 솔루션으로, Lucene에는 없는 데이터베이스 기능 등 다양한 기능을 통합·지원한다. 특히, Lucene에는 없던 NoSQL 기능들을 완벽하게 지원하고 자체적으로 Sharding을 통한 분산-클러스터링 구조를 지원한다. 이러한 이유로 Lucene은 주로 Solr 솔루션과 함께 활용되고 있다. Lucene/Solr 역시 Java를 기반으로 개발되어 대부분의 운영체제를 지원한다.

III. 시스템 요구사항

RENTAP의 설계를 위해 다양한 요구사항들이 고려되었으며, 최종적으로 군 네트워크 관리자 및 정보보호 실무자 등의 의견을 수렴하여 개발에 적용될 주요 요구사항을 도출하였다.

3.1 대용량 로그 증적 및 분석

가장 중요한 요구사항으로서, 네트워크 침입의 효과적인 탐지를 위해 군내 네트워크에서 발생하는 모든 통신 사실 및 전 계층의 트래픽에 대한 로그를 수집할 수 있어야 한다. 하지만, 군 작전 및 업무에서의 네트워크 활용도가 높아지면서 군 네트워크에서 매 순간 발생하고 있는 트래픽의 용량은 상상을 초월하는 실정이다. 군내 주요부대의 내부망 및 로그 데이터로 저장 대상을 한정한다고 하더라도, 수일 만에 기가바이트급의 로그가 저장될 수도 있을 것이다. 따라서 구축 시스템은 적게는 기가바이트에서 많게는 테라바이트급에 이르는 데이터를 지속적으로 저장하고 분석할 수 있어야 한다. 또한, 최대한 다양한 데이터를 분석에 활용할 수 있게끔 다양한 데이터 형식을 지원하는 유연한 파일 포맷을 사용해야 할 것이다.

3.2 실시간 로그 조회 및 탐색

대용량의 데이터를 저장하고 있더라도 자료의 대용량성으로 인해 시스템의 가용성이 떨어지면 전·평시간 시스템의 활용도가 떨어질 수 밖에 없다. 따라서 구축 시스템은 데이터 수집·탐색 간 인덱싱 및 분산처리 등을 통해 실시간성을 지원해야 할 것이다. 로그 탐색을 위해 정규표현식 등을 통한 다양한 조건문 형태의 질의(Query)을 지원해야 한다.

3.3 기타 요구사항

기존의 보안장비 및 소프트웨어에 대한 지원을 통해 호환성 및 확장성을 지원해야 한다. 특히 지원 로그형태 및 프로토콜 사용 간 표준 기술의 적극적인 활용을 통해 호환성을 확보해야 할 것이다. 저장되는 로그 정보의 상세를 정보보호와 관련된 항목으로 제한하여 과도한 사생활 침해의 여지가 없어야 한다.

IV. 솔루션별 비교·분석

위에서 알아본 빅데이터 솔루션들 중, 구축 시스템에 가장 적합한 솔루션을 선정하기 위하여 비교·분석을 수행하였다.

4.1 비교 환경

비교를 위한 환경은 다음과 같이 구축하였다. 우선, 하드웨어의 경우, 빅데이터 솔루션 코어를 설치하기 위한 메인서버를 두고 데이터의 분산처리·저장을 수행하기 위한 보조서버 두 대를 100Base-T 이더넷 스위치를 통해 연동하였다. 전체 하드웨어 설정은 [표 1]과 같다. 솔루션별 버전은 MongoDB는 2.3.1,

[표 1] 솔루션별 비교에 사용된 하드웨어 사양

구분	사양
메인서버	CPU : Intel Core i7-2600 CPU / 3.4GHz / 8 Core 메모리 : 16GB 하드디스크 : 2TB OS : CentOS 6.2(Kernel V. 2.6.32-220)
보조서버1	CPU : Intel Xeon CPU E5606 / 2.13 GHz / 8 Core 메모리 : 16GB 하드디스크 : 1TB OS : CentOS 6.2(Kernel V. 2.6.32-220)
보조서버2	CPU : Intel Xeon CPU E5606 / 2.13 GHz / 8 Core 메모리 : 16GB 하드디스크 : 1TB OS : CentOS 6.2(Kernel V. 2.6.32-220)

〔표 2〕 솔루션별 비교에 사용된 JSON 파일 포멧

필드(Field)	설명	값(Value) 예시
connect_ip	접속 아이피	192.168.1.22
in_out	인/아웃바운드	1
attack_nm	공격명	HTTP DDOS
src_port	소스 포트번호	49855
dstn_port	목적지 포트	80
mac_addr	접속 맥 주소	00:88:65:B2:3E:FA
recv_time	접속 시간	20120805144428
country_name	접속 국가	S. Korea
city	접속 도시	Seoul
longitude	접속지점 위도	122.057
latitutde	접속지점 경도	37.419
agent_ip	수집에이전트 아이피	192.168.1.23
cnt	접속 카운트	122
isp	접속 ISP	KT
log_type	로그 타입	1

HBase/Hadoop은 0.92.2/1.1.0, Lucene/Solr는 4.0/4.0 버전을 각각 사용하였다. JAVA JDK는 Oracle의 1.6.0.31 버전을 사용하였다. 저장 데이터는 JSON 형식으로 〔표 2〕의 예시와 같이 필드를 지정하고 임의의 값으로 설정한 로그 파일(CSV 형식 텍스트 파일)로부터 입력을 받도록 하였다.

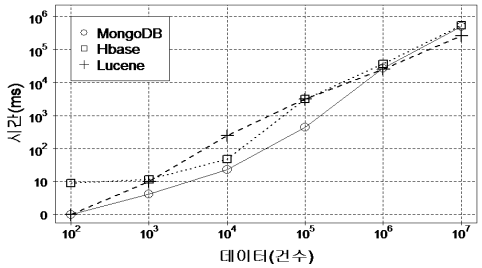
4.2 성능 비교

4.2.1 데이터 입력(Data Insertion)

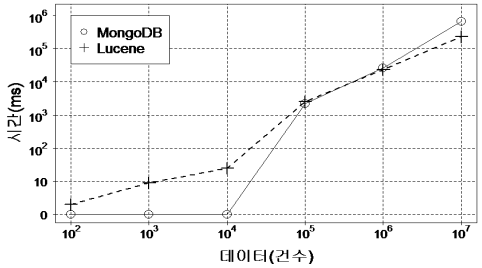
데이터 입력 비교는 대용량 로그를 지속적으로 저장해야 하는 구축 시스템의 특성상 성능 비교시 가장 중요한 평가요소라고 볼 수 있다. 데이터 입력 비교는 100건의 로그 데이터부터 10,000,000건에 이르는 로그 데이터를 입력받아 처리하도록 하였다. 〔그림 1.(가)〕을 통해 세 솔루션의 데이터 입력 성능을 확인할 수 있다. 우선 100,000건 이하에서는 MongoDB의 성능이 가장 뛰어남을 확인할 수 있으나 대용량이라고 볼 수 있는 1,000,000건 부터는 Lucene/Solr의 성능이 더 우수함을 확인할 수 있다. 10,000건을 제외하면 MongoDB와 Lucene에 비교하여 HBase/Hadoop의 입력성능이 가장 떨어짐을 확인할 수 있다. 특히, 저용량에서의 성능이 다른 두 솔루션에 비해 떨어지며 이는 시간축을 고려할 때, 평시(지속적인 저용량 입력) 환경에서 타 솔루션에 비해 매우 불리할 것으로 예상된다. 그래프의 Y축이 Log-Scale임을 고려하면 고용량에서의 성능도 타 솔루션에 비해 많이 떨어지는 편이다(예 : 10,000,000건에서 Lucene/Solr-4분 20.23초, Hbase/Hadoop-8분 55.45초). 이에 따라 중요 평가요소인 입력성능에서 낮은 성능을 보여준 HBase는 선정 솔루션에서 제외하고 남은 두 솔루션으로 나머지 성능비교를 진행하였다.

4.2.2 데이터 삭제(Data Deletion)

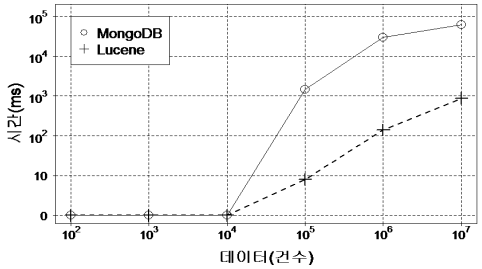
데이터 삭제 성능의 경우 평가 요소 중 상대적으로 중요도는 떨어지지만 유지보수를 위한 필수기능이기 때문에 참고 수준에서 성능비교를 수행하였다. 〔그림 1.(가)〕의 시험결과를 보면 100,000건 이하의 삭제 성능에서 MongoDB가 Lucene에 비해 더 좋은 성



(가) 데이터 건수별 데이터 입력 지연시간



(나) 데이터 건수별 데이터 삭제 지연시간



(다) 데이터 건수별 데이터 출력 지연시간

〔그림 1〕 솔루션별 비교시험 결과

능을 보여주는 것을 확인할 수 있다. 반면, 상대적으로 대용량인 1,000,000건 이상의 삭제성능에서는 Lucene의 성능이 더 우수함을 확인할 수 있다.

4.2.3 데이터 출력(Data Fetch)

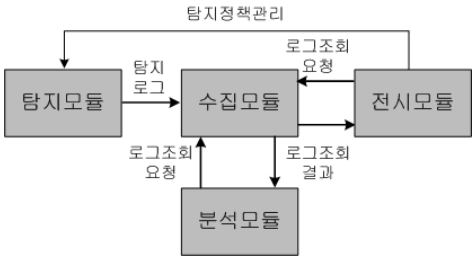
데이터 출력 역시 데이터 입력과 함께 구축 시스템의 주요 수행 작업 중 하나로서 중요한 평가요소라고 할 수 있다. 데이터 출력 비교는 구축 시스템 특성상 빈번하게 사용되는 범위(Range) 검색 질의를 통해 전체 입력 데이터를 출력(Fetch)하는 형태의 작업을 수행하는 방식으로 수행하였다. 단, 100,000건 이상의 경우, 소요시간을 감안하여 출력 데이터를 100,000건으로 한정하여 진행하였다. 〔그림 1.(다)〕에 나타난 시험결과를 보면 MongoDB와 Lucene/Solr 모두 10,000건 이하에서는 매우 우수한 성능을 보여주었으나, 100,000건 이상에서는 Lucene/Solr가 압도적으로 뛰어난 성능을 보여준다. 이는 Lucene/Solr이 원래 검색엔진 용도로 개발된 만큼, 데이터 출력에 최적화된 인덱싱 기능에 기인한 성능차이라고 볼 수 있다.

4.3 비교결과 분석

세 솔루션을 선정하여 비교한 결과, 전반적으로 구축 시스템에 적용하기에 가장 적합한 성능을 보여주는 것은 Lucene/Solr라고 볼 수 있다. 특히, Lucene/Solr는 1,000,000건 이상의 성능 비교에서 모두 수위를 차지하여 대용량 데이터를 빠르게 처리하는 데 강점이 있음을 보여준다. 따라서 RENTAP 시스템 설계에는 Lucene/Solr를 적용하기로 결정하였으며 RENTAP의 시스템 설계는 해당 솔루션을 중심으로 수행하였다.

V. 시스템 설계 및 구현결과

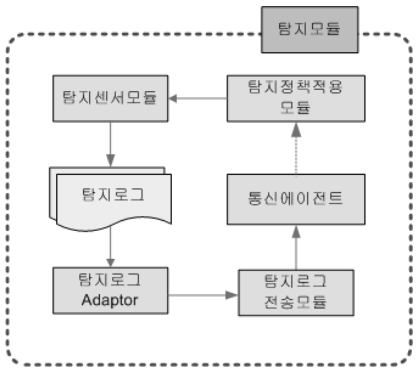
선정된 솔루션을 기반으로 시스템 설계를 수행하였다. 앞서 분석한 요구사항을 모두 반영할 수 있도록 설계를 수행하였으며, 담당 기능을 기준으로 소프트웨어 모듈별로 나누어 설계를 수행하였다. 소프트웨어 모듈은 크게 탐지 모듈, 수집모듈, 분석모듈, 전시모듈로 구분되며, 각각의 모듈 간 데이터 흐름은 다음 다이어그램 〔그림 2〕와 같다. 다음은 각 모듈에 대한 세부 모듈 구성 및 데이터 흐름에 대한 설명이다.



〔그림 2〕 RENTAP 전체모듈에 대한 FBD(Function Block Diagram)

5.1 탐지 모듈(Sensor Module)

탐지 모듈은 네트워크/보안장비를 통해 네트워크 트래픽 발생을 감지하고 1차적으로 탐지로그(원시로그)를 생성·보고하는 역할을 수행한다. 탐지 모듈의 세부 모듈 구성 및 데이터 흐름은 다음 다이어그램 〔그림 3〕과 같다.

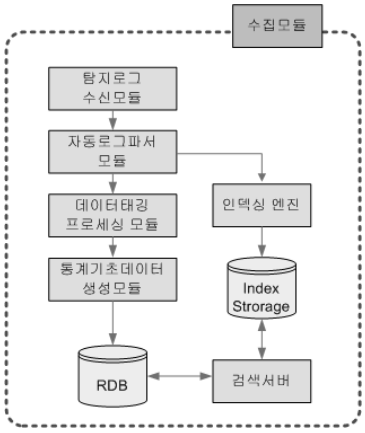


〔그림 3〕 탐지 모듈에 대한 세부 FBD

탐지센서 모듈은 네트워크 스위치 장비의 SPAN (Switched Port ANalyzer) 포트 등을 통해 획득한 패킷의 헤더정보를 통해 전 계층에 대한 통신사실을 텍스트 로그형태로 생산한다. 탐지로그 Adaptor는 자체 탐지로그 뿐만 아니라 타 보안장비나 호스트에서 생산된 로그를 수신하는 역할을 수행한다. 생산된 로그는 통신에이전트를 통해 수집모듈로 송신된다.

5.2 수집 모듈(Collector Module)

수집 모듈은 탐지 모듈로부터 수집된 로그를 수신하고 정규화하여 약속된 로그 포멧(JSON) 형태로 변환한다. 또한 Lucene/Solr를 통한 분산 인덱싱 및 적재를 수행한다. 다음 〔그림 4〕는 수집 모듈의 세부



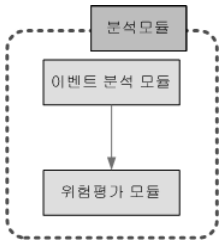
〔그림 4〕 수집 모듈에 대한 세부 FBD

구성을 나타낸 다이어그램이다.

수집된 로그들은 Lucene/Solr를 통해 구현된 인덱싱 엔진을 통해 처리되어 저장된다. 단, 로그 수집 현황(호스트별 수집로그 건수, 이벤트별 건수 등)에 대한 통계정보 저장을 위해 RDB(MySQL)를 사용한다. 검색서버 모듈은 다른 모듈로부터의 로그 및 통계정보 조회 요청을 Lucene/Solr를 통해 처리하여 결과를 해당 모듈에 전송한다.

5.3 분석 모듈(Analyzer Module)

분석의 세부 모듈은 다음 〔그림 5〕와 같이 크게 수집된 로그를 기반으로 침해사고 이벤트 발생 여부 등에 대한 분석을 수행하는 이벤트 분석 모듈과 이를 기반으로 내부네트워크에 대한 위험평가를 수행하는 위험평가 모듈로 구성된다. 이벤트 분석 모듈은 필요시 수집모듈을 통해 로그 조회를 수행하여 미리 정의된 침해사고 이벤트 발생여부를 체크한다. 이벤트 분석 모듈은 미리 정의된 룰 이외에도 정규표현식 등을 통해 사용자가 구체적인 룰을 추가할 수 있도록 하였으며, 위험평가 모듈 역시 이벤트별 위험도를 사용자가

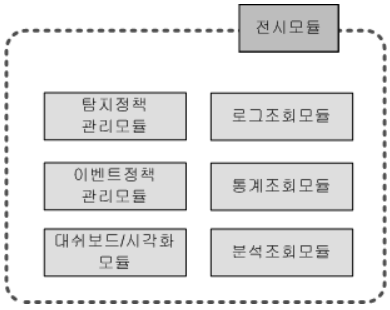


〔그림 5〕 분석 모듈에 대한 세부 FBD

정의할 수 있게 하여 보호대상 네트워크 상황에 맞는 분석이 가능하게끔 설계하였다.

5.4 전시 모듈(Viewer Module)

전시 모듈은 관리자가 로그 및 위험분석 현황 등을 모니터할 수 있도록 콘솔화면에 전시하는 역할을 수행한다. 각 세부 모듈은 다음 다이어그램 〔그림 6〕을 참고한다.



〔그림 6〕 전시 모듈에 대한 세부 FBD

5.5 시스템 구현 결과

상기에 소개한 설계를 바탕으로 RENTAP 시스템을 구현하였으며, 시범운용결과, 실제 기존 방화벽 및 침입탐지시스템만 적용한 경우와 비교하여 다양한 공격행위에 대한 효과적인 탐지가 가능함을 확인하였다. 특히, 다음과 같은 시나리오의 공격에 대한 탐지의 경우, RENTAP을 통한 탐지가 유리하다.

- Fastflux 도메인 접속 탐지
  (간헐적으로 IP를 변경하는 도메인 접속)
- 테더링에 의한 정보유출 탐지
  (단일 IP에 대한 다수 TTL값 탐지)
- 내부 호스트에 대한 하프오픈 스캔 시도 탐지
  (스캔 딜레이를 이용한 간헐적인 스캔 시도)
- 기타 악성코드에 의한 C2채널 및 Covert Channel 탐지 등

상기에 언급한 공격징후들은 대부분 내부망 내부에서 외부로 향하는(아웃바운드) 트래픽들로서, 기존의 정보보호장비의 경우, 상대적으로 인바운드 트래픽에 비해 필터링 수준이 낮은 부분이다. 또한, 대부분의 경우, 기존의 정보보호체계에서는 분석에 있어 윈도우

사이즈를 두어 해당 기간 내에 설정한 임계값 초과 여부만을 검사한다. 이러한 경우, 이를 노린 간헐적인 공격이 일어나면 탐지할 수 없는 경우가 대부분이다. 하지만, RENTAP의 경우, 인/아웃바운드를 포함, 장기간에 걸쳐 증적된 모든 통신 로그에 대한 분석을 실시하므로 위와 같은 행위 탐지에 매우 유리하다.

VI. 결 론

본 논문에서는 군 내부 네트워크를 실시간으로 점검하고 감시하기 위한 빅데이터 기반의 실시간 네트워크 트래픽 분석 플랫폼, RENTAP의 특성 및 설계과정에 대해 소개 하였다. 플랫폼 설계 결과는 시제 구축에 적용되어 현재 시범운용을 실시중이다. 시범운용 결과를 토대로 볼 때, 본 플랫폼의 실제 적용 시 다음과 같은 기대효과가 예상된다.

- 군 내부망 트래픽 현황에 대한 세밀한 실시간 감시가 가능함에 따라 보안관제 간 빠르고 정확한 대응을 할 수 있다.
- 대용량 로그 검색을 통해 장기간에 대한 트래픽 사용 현황분석이 가능함에 따라 이를 통한 내부 네트워크상 이상 현상(Anomaly) 탐지가 가능하다.
- 내부망 침해사고 발생시, 통신사실에 대한 로그의 전수 확보가 가능함에 따라 빠르고 정확한 분석이 가능하다.
- 기존 군 정보보호장비를 그대로 활용하면서도 정보보호 능력을 강화할 수 있다.

군 정보보호 분야에서 빅데이터 활용은 이제 시작에 불과한 단계라고 볼 수 있다. 따라서 RENTAP의 사례를 참고하여 향후 다른 국방정보체계 구축 시 빅데이터 기술을 적용한다면 다양한 시너지 효과가 있을 것으로 기대된다.

참고문헌

〔1〕 A, Oliner, A. Ganapathi, and W Xu, "Advances and challenges in log analysis," ACM Queue, vol. 9, no. 12, pp. 30, Dec. 2011.

〔2〕 장희진, 이동환, 박찬일, 윤호상, "베이지안 네트워크를 이용한 내부자 사이버 위협 예보 시스템," 한국군사과학기술학회 종합학술대회 논문집, pp. 470~473, 6월, 2012.

〔3〕 C. Taylor and J. Alves-Foss, "NATE: Network analysis of anomalous traffic events, a low-cost approach," Proceedings of the 2001 ACM Workshop on New Security Paradigms, pp. 89-96 Sep. 2001.

〔4〕 P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement. pp. 71-82, Nov. 2002.

〔5〕 C. Livadas, B. Walsh, D. Lapsely, and W. T. Strayer, "Using machine learning techniques to identify botnet traffic," Proceedings of the 31st IEEE Conference on Local Computer Networks, pp. 967-974, Nov. 2006.

〔6〕 A. Rabkin and R. Katz, "Chukwa: A systam for reliable large-scale log collection," Proceedings of the 24th USENIX International Conference on Large Installation System Administration, pp.1-15, Nov. 2010.

〔7〕 M. S. Rehman, M. Hammoud, and M. F. Sakr, "VOtus: A flexible and scalable monitoring framework for virtualized clusters," (Poster Paper) Proceedings of the 3rd IEEE International Conference on Cloud Computing and Science, Dec. 2011.

〔8〕 Introduction to MongoDB, http://www.mongodb.org/about/introduction

〔9〕 Apache HBase Architecture Overview, http://hbase.apache.org/book/architecture.html#arch.overview

〔10〕 What is Apache Hadoop?, http://hadoop.apache.org/index.html#What+Is+Apache+Hadoop%3F

〔11〕 Apache Lucene Core Features, http://lucene.apache.org/core/features.html

〔12〕 Apache Solr Features, http://lucene.apache.org/core/features.html

〈저 자 소 개〉

이 동 환 (Donghwan Lee) 정회원
2006년 2월: 고려대학교 산업시스템정보공학과 졸업
2008년 2월: 고려대학교 컴퓨터통신공학과 석사
2008년 2월~현재: 국방과학연구소 연구원
〈관심분야〉 정보보호, 빅데이터 분석, 무선네트워크 보안

박 정 찬 (Jeongchan Park) 정회원
1994년 2월: 광운대학교 컴퓨터공학과 졸업
1996년 2월: 광운대학교 컴퓨터공학 석사
1996년~현재: 국방과학연구소 선임연구원
〈관심분야〉 정보보호, 분산협업기술

유 찬 곤 (Chan-gon Yoo) 정회원
1997년 2월: 충남대학교 컴퓨터과학과 졸업
2003년 8월: South Dakota State University 석사
2003년 8월~현재: 국방과학연구소 선임연구원
〈관심분야〉 정보보호

윤 호 상 (Hosang Yun) 정회원
1987년 2월: 고려대학교 수학과 졸업
1990년 8월: 고려대학교 전산학과 석사
2003년 8월: KAIST 전산학과 박사
1990년 9월~현재: 국방과학연구소 책임연구원
〈관심분야〉 정보보호, 운영체계, 네트워크