Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.27, NO.5, Oct. 2017 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2017.27.5.1117

빅데이터 분석 기술(Hadoop/Hive) 기반 네트워크 정상행위 규정 방법*

김 성 진,† 김 강 석‡
아주대학교

A Normal Network Behavior Profiling Method Based on Big Data Analysis Techniques (Hadoop/Hive)*

SungJin Kim,† Kangseok Kim‡
Ajou University

요 약

사물인터넷 시대의 도래로 인터넷에 연결된 다양한 기기들의 사용은 급성장 하였으나 사물인터넷 보안은 아직 취약한 상태이다. 사물인터넷은 목적에 따라 다양한 기기들이 사용되고 또한 저 전력 환경에서 동작할 수 있도록 각기 다른 프로토콜들을 사용하고 있으며, 많은 양의 트래픽을 발생시켜 기존 보안 기술들을 접목시키기 어렵다. 그러므로 본 논문에서는 이러한 문제점을 해결하기 위한 방안중의 하나로 Hadoop/Hive를 이용한 빅데이터 분석 기술 및 통계 분석 도구인 R을 활용하여 네트워크 정상행위 규정 방법을 제시하며 시뮬레이션을 통해 제안한 방법의 유효성을 검증한다.

ABSTRACT

With the advent of Internet of Things (IoT), the number of devices connected to Internet has rapidly increased, but the security for IoT is still vulnerable. It is difficult to integrate existing security technologies due to generating a large amount of traffic by using different protocols to use various IoT devices according to purposes and to operate in a low power environment. Therefore, in this paper, we propose a normal network behavior profiling method based on big data analysis techniques. The proposed method utilizes a Hadoop/Hive for Big Data analytics and an R for statistical computing. Also we verify the effectiveness of the proposed method through a simulation.

Keywords: Big Data, Intrusion Detection, Simulation, Security Data Analysis, Normal Behavior Profiling

I. 서 론

최근 국가 주요 시설들을 대상으로 하는 사이버 공격의 발생이 증가하고 있다. 2010년 이란의 원자력 발전소를 대상으로 한 스턱스넷(Stuxnet) 이후 발생한 국가 주요 시설들을 대상으로 한 공격들은 모두 물리적인 피해와 경제적인 피해를 동시에 일으킬 수 있는 치명적인 공격들로 사이버 테러에 대한 인식을 새롭게 하였다[1].

2016년 말 우크라이나는 전력분야를 포함한 국가 주요 시설들이 6,500건의 공격을 받았고, 그 중 일부는 심각한 경제적/물리적 피해를 입혔다[2]. 우크라이나 금융 부처(Ministry of Finance)에서는 사이버 공격으로 인해 네트워크 장비들이 피해를 입었고, 3TB 가량의 데이터 손실, 2일간 약 300,000건의 거래 불가능 등의 피해를 받은 것으로 파악되었다. 전력분야에서는 우크라이나 수도인 키에프(Kiev)시에 정전이 발생하는 피해가 있었다. 정전은

Received(05. 02. 2017), Modified(1st: 08. 01. 2017,2nd: 09. 05. 2017), Accepted(09. 05. 2017)
* 이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2015R1D1A1A01060236).
† 주저자, ksjskyblue@ajou.ac.kr
‡ 교신저자, kangskim@ajou.ac.kr(Corresponding author)

1시간 15분 만에 복구되었으나, 2015년 우크라이나 정전 사고에 이어 또 다시 사이버테러로 정전이 발생하여 사이버테러의 위험성을 다시 상기시켰다.

이러한 공격들의 대상은 주요 기반시설에서 점차 개인으로 변화될 것으로 예상된다. 사물인터넷 시대의 도래로 인터넷에 연결된 다양한 기기들의 사용은 급성장 하였으나 사물인터넷 보안은 아직 부족한 상태이다. 사물인터넷은 다양한 목적에 따라 여러 기기들이 사용되기 때문에 각기 다른 프로토콜들이 사용되고 있어 기존 보안 기술들을 접목하기 어렵고, 낮은 기계의 성능에 비해 많은 양의 데이터가 발생하기 때문에 새로운 보안 기술을 필요로 한다.

현재 사물인터넷 시대의 도입부임에도 불구하고 사물인터넷 기기들에 대한 공격이 이미 발생하였다. 미국 보안 업체 프루프포인트가 2013년 12월부터 2014년 1월까지 세계 전역에 발송된 스팸메일 중 25%가 좀비화된 스마트 가전기기에서 발생하였음을 발견하여 발표한 스마트 냉장고 해킹과 스마트 TV 공격 등 다양한 공격들과 공격 가능성이 확인되었다[3]. 이 외에도 Fig. 1과 같이 해커가 교통 표지판을 해킹한 사례도 존재한다[4].

위 Fig. 1은 장난스러운 문구를 올려 실제 큰 피해는 발생하지 않았지만, 악의를 가지고 교통에 문제되는 경고문을 사용할 경우 교통사고의 발생도 가능한 상황이기 때문에 그 위험성에 대해 충분히 고려할 필요가 있다.

이러한 사물인터넷 기기들의 보안 문제점 해결을 위해 본 논문에서는 사이버 공격을 감지하기 위해 빅데이터 분석 방법을 적용한 네트워크 정상행위 규정 방법을 제안한다. 침입 탐지를 위해 패킷과 트래픽의 특성을 빅데이터 분석 방법을 이용하여 더욱 빠르고 효과적으로 파악한다. 제안한 기법은 다양한 프로토콜이 사용되고 있는 사물인터넷 환경에서 네트워크 침입 탐지를 위한 기초적인 패킷 및 트래픽의 특성 분석과 정상행위의 정의에 많은 영향을 미칠 것으로 사료된다. 특히 정상행위를 기반으로 침입 탐지 규칙을 생성할 경우 사전에 발생 가능성을 식별할 수 있는 요소(known-unknown)와 발생 가능성을 사전에 식별할 수 없는 요소(unknown-unknown) 둘 모두에 적합한 기법이기 때문에 고도화된 공격에도 적합한 방법으로 판단된다.

Fig. 1. Hacked direction sign

본 논문의 2장에서는 빅데이터 분석 기술과 보안을 접목한 다양한 연구들에 대해 살펴보고, 3장에서는 빅데이터 분석 기술들에 대해 논의한다. 이후 4장에서는 살펴본 빅데이터 분석 기술들을 바탕으로 침입 탐지 방법을 제안하고, 5장에서 실제 환경과 유사하게 모델링 하여 제안한 방법의 유효성을 검증한다. 마지막 6장에서는 결론과 향후 연구에 대해 논의한다.

II. 관련연구

빅데이터 분석 기술은 다양한 IT 분야에서 활용되고 있으며, 기존에 처리가 불가능하였던 대용량의 데이터를 분석하여 새로운 유의미한 데이터를 발굴하여 변화의 중심에 있다. 사이버 보안에서도 기존 보안 기법에 빅데이터 기술을 접목하려는 시도들이 많이 존재하고 있으며, 이미 일부는 제품으로도 출시된 상황이다.

빅데이터 기술이 가장 빠르게 접목된 것은 SIEM(Security Information and Event Management)이며, IBM, Splunk등 업체들도 다양한 제품들을 출시하고 있다[5]. 학계에서는 APT(Advanced Persistent Threat) 공격 대응을 위한 방법으로 빅데이터 기술을 이용하는 SIEM을 제안하고 있다. 김도근 등의 연구에서는 APT 공격 방어를 위해 여러 소스를 통해 데이터를 수집하고 이를 분석하는 지능형 보안 시스템을 제안하였다[6]. 설계한 시스템에 대한 상세 내용들에 대해 논의되지는 않았으나 기존의 보안 기법으로는 방어가 불가능한 APT 공격을 막기 위한 새로운 시도라 판단된다.

기존 시스템에 추가하는 사례는 로그 분석 및 실시간 예측 시스템에서도 나타난다. 이상준 등은 수집되는 로그의 실시간 예측분석 시스템을 설계하여 사

이버 공격에 빠르게 대응할 수 있는 시스템을 제안하였고, 실시간 로그의 분석을 통해 사이버 공격에 대한 예측이 가능함을 보여준 사례를 보여 주어 실시간 예측분석 시스템에 빅데이터 기술의 접목이 가능함을 보여주었다[7]. 이 외에도 2010년부터 진행되고 있는 DARPA(Defense Advanced Research Projects Agency)의 CINDER(Cyber Insider Threat) 프로그램에서는 군사관련 네트워크의 첩보자를 찾기 위해 내부자의 행위 분석에 빅데이터 분석 기술을 적용하고 있다[8].

이와 같이 빅데이터 기술을 접목하는 연구들이 활발히 진행되고 있으며, 빅데이터 기술을 접목 시 프라이버시 등 여러 보안 이슈들이 발생 할 수 있는데, 정교일 등은 이러한 이슈들을 빅데이터 기술 접목을 위한 데이터 생산 및 수집, 저장 및 운영, 분석 등 각 구간별 보안 이슈들에 대해 논의하였다[9]. 이러한 연구들은 빅데이터 분석 기술을 접목시킬 때 바탕이 되는 연구라 할 수 있다.

이렇게 살펴본 빅데이터 분석 기술들은 사이버 보안 기술에 다양하게 적용되고 있는 상황이다. 본 논문에서는 분석 기술들과 함께 시뮬레이션 기법도 활용하여 제안하는 기법의 유효성 검증을 수행하고자 한다. 다음 장에서는 본 논문에서 제안하는 기법을 사용하기 위한 빅데이터 분석 도구들에 대해 간략히 논의한다.

III. 빅데이터 분석 기술 및 도구

3.1 빅데이터 분석 플랫폼 : Hadoop/Hive

빅데이터 분석에 사용되는 아파치 Hadoop[10]은 병렬처리를 통해 큰 규모의 데이터를 빠르게 처리할 수 있도록 하는 프레임워크로 빅데이터 분석에서 가장 널리 사용되는 도구이다. Hive[11]는 Hadoop 상위에서 동작하는 응용프로그램으로 사용자에게 익숙한 SQL Query를 이용하여 빅데이터 분석이 가능하도록 해 주는 도구이다. 빅데이터 분석의 가장 기본이 되는 도구들로 데이터의 세부적인 특징들을 상세히 분석하여 대규모의 데이터에서 비교적 쉽게 확인되는 특징 점들을 도출하고, 추가 분석이 필요한 필드들을 선별하는 데 사용 할 수 있다.

3.2 통계적 분석 : R

R[12]은 통계 계산을 위한 프로그래밍 언어로 세부적인 데이터 분석에 활용 될 수 있다. 다양한 통계 계산들을 라이브러리로 제공하고 있고, 데이터를 그래프로 시각화하기 위한 라이브러리 함수들도 제공되고 있어 기본 데이터만으로는 확인되지 않은 특징 점들에 대한 분석이 가능하다. 하지만 큰 규모의 데이터를 처리하기엔 부적합하기 때문에 빅데이터 분석에서는 앞에서 소개한 Hadoop/Hive를 이용하여 필드들을 도출하고, 이를 분석하는 용도로 사용하는 것이 바람직하다.

3.3 이산 사건 모델링 및 시뮬레이션 : Arena

Arena[13]는 이산사건(discrete event) 시뮬레이션 도구로 비교적 간단한 인터페이스를 기반으로 다양한 이산사건 모델링의 시뮬레이션이 가능하다는 장점이 있다. 더욱 널리 사용되는 다양한 시뮬레이션이 존재하지만, Arena는 이보다 간단한 인터페이스를 바탕으로 작동하기 때문에 비교적 진입 장벽이 높지 않고, 시뮬레이션 동작 내용을 시각적으로 확인할 수 있으며, 빅데이터 분석에서 Arena는 데이터 분석 내용을 토대로 이를 검증하는 시뮬레이션 용도로 활용이 가능하다.

IV. 빅데이터 분석 기술을 이용한 네트워크 정상행위 규정 방법

사물인터넷 시대의 도래로 다양한 기기들이 특정 목적에 따라 통신하며 사용자에게 편리한 서비스를 제공하기 시작하였다. 이에 따라 전체 네트워크 트래픽은 기하급수적으로 증가하였다. 이렇게 증가한 트래픽은 현실로 다가왔지만, 기존의 보안 기술은 이를 처리하지 못하고 있기 때문에 새로운 사물인터넷을 위한 보안 기술의 필요성이 요구되고 있다.

폭발적으로 증가한 네트워크 트래픽 외에도 사용되는 프로토콜의 수도 증가하였다. 기존 환경에서 사용되던 프로토콜들과는 달리 저 전력 환경에서도 동작이 적절하도록 새로운 하위 계층의 프로토콜들이 개발되었고, 응용계층에는 각 업체별로 다른 프로토콜을 사용하고 있다[14]. 따라서 기존 네트워크 트래픽에 비해 다양성이 증가하여 새로운 침입 탐지 방법을 필요로 하고 있다.

본 논문에서는 빅데이터 분석 기술을 네트워크 트래픽의 정상행위 규정에 활용하는 기법에 대해 제안한다. 기존과는 달리 빅데이터 기술을 필요로 할 정도로 대규모의 트래픽이 생성 및 사용되고 있고, 정형 데이터라 할지라도 프로토콜의 다양성으로 인해 복잡도가 증가하여 빅데이터 기반의 분석 기술의 도입이 적합하다고 판단된다.

빅데이터 분석 기술은 많은 양의 비정형 데이터를 정형화하고, 특정 요소를 추출하여 분석하는 것에 초점을 맞추고 있다. 이를 활용하여 본 논문에서는 기존 트래픽 레벨의 정상행위 규정과 함께 패킷의 세부 필드까지 분석한 정상행위도 규정하고, 이를 토대로 침입탐지를 수행하는 기법을 제안한다. Fig. 2는 제안하는 빅데이터 분석기법을 활용한 침입탐지 기법이다.

제안하는 기법은 Fig. 2와 같이 총 4가지 단계로 구성된다. 첫 번째 단계인 대상 네트워크 트래픽 수집은 정상적으로 동작하고 있는 네트워크의 패킷을 수집하는 것이다. 이때 수집되는 데이터의 양은 이를 바탕으로 정상행위를 규정하여도 될 정도로 많은 양의 데이터를 수집하는 것이 필수적이다.

이후 진행하는 단계인 주요 필드 추출단계는 빅데이터 분석 플랫폼을 이용하여 다양한 각도로 데이터를 살펴보는 과정이다. 이때 트래픽 단위와 패킷 단위의 분석이 병행될 필요가 있다. 일례로 시간당 발생하는 트래픽의 수가 일정하다는 특성을 네트워크가 가지고 있을 때, 패킷 단위의 분석을 수행할 경우 파악하기 힘들지만 트래픽 레벨에서 데이터를 분석할 경우 손쉽게 파악 할 수 있다. 따라서 트래픽과 패킷 단위의 수준에서 데이터를 분석할 필요가 있다. 두 관점으로 데이터를 분석하여 유의미한 패킷의 필드와 트래픽 특성들을 이 단계에서 추출한다.

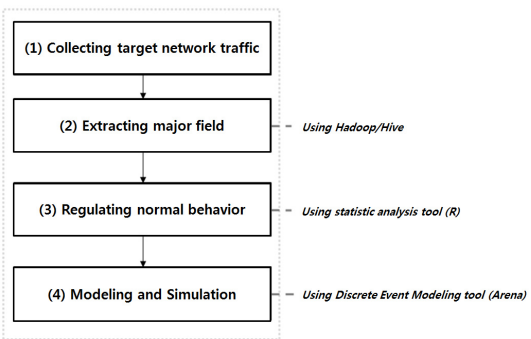

Fig. 2. Intrusion detection process based on big data analysis technique

이후 정상행위 규정 단계에서 추출한 특징들에 통계적 분석을 수행하여 구체적인 정상행위를 규정한다. 일례로 패킷이 5초로 오는 경향이 있다는 사실은 빅데이터 분석 플랫폼에서도 확인 할 수 있다. 하지만 이 특징을 기준으로 데이터를 추출하여 통계적 분석 도구를 활용하여 분석을 수행 할 경우 정확한 주기와 오차 범위를 계산 할 수 있기 때문에 통계적인 분석을 이용하여 정상행위를 규정한다. 이를 통해 단순히 빅데이터 플랫폼을 통해 정상행위를 규정하는 것에 비해 상세한 정상행위 규정이 가능하여 해당 시스템에 적용 시 낮은 오탐율을 기대할 수 있다.

마지막 단계는 규정한 정상행위의 입증을 수행한다. 실제 적용에 앞서 실제 환경과 유사하게 모델링하여 규정한 정상행위를 사용했을 때 발생 가능한 문제점들을 사전에 조사한다. 패킷의 송수신은 각각 별개의 사건이기 때문에 이산사건모델링 기법이 활용된다.

총 4단계로 수행되는 침입탐지를 위한 정상행위 규정에는 3장에서 언급한 Hadoop/Hive, R, Arena가 각 단계에 활용된다. 다음 5장에서는 사물인터넷 환경 중 하나인 산업 사물인터넷을 대상으로 본 기법을 적용하여 침입탐지 방법을 개발한다.

V. 실 험

빅데이터 분석 기술을 비정상 트래픽의 침입탐지 기술에 접목한 방법의 유효성을 검증하기 위해 전력 시스템에서 주로 사용되는 DNP3(Distributed Network Protocol) 트래픽 로그를 이용하여 실험을 수행하였다. 실험에 사용되는 DNP3 프로토콜은 변전소와 SCADA(Supervisory Control and Data Acquisition)사이의 통신에서 사용되고 있는 프로토콜이다[15]. 테스트 패킷 수집을 위한 통신 구성도는 Fig. 3과 같다.

공격자는 클라이언트에게는 서버 IP로 변경 한 패킷을 이용한 공격을 수행하고, 서버에게는 클라이언트 IP로 변경한 공격과 허가되지 않은 클라이언트의 IP(10.0.0.7)를 이용한 공격을 수행한다.

시뮬레이션에서는 Fig. 3과 같이 실제 환경보다 간략화 하여 하나의 FEP(Front End Processor)와 하나의 RTU(Remote Terminal Unit)가 통신하는 모델을 사용하였다. 캡처된 패킷은 2일 동안 23MB, 158,403개의 패킷으로 Table 1과 같이 정상 트래픽을 구성하고, Table 2와 같이 Digital Bond에서 제공하는 공격을 기반으로 공격 트래픽을

Table 1. Elements of normal traffic

Type	Data Type	Period
Solicited	Analog Input	10 s
	Binary Input	5 s
	Binary Output	5 s
	Internal Indication	60 s
Unsolicited	Binary Input Change / Confirm	-
Connection Initiate	-	-

Table 2. Elements of attack traffic

Attack Name	Description
Port Scan	send a request to a range of server port address for finding an active port
Point List Scan	send a request to a range of DNP3 point list for scanning active point
Broadcast Request from Authorized Client	send broadcast message from authorized client to disrupt network traffic
Warm Restart	send restart message to target device for disrupting normal operation
Cold Restart	send restart message to target device for disrupting normal operation
Stop Application	send stop application message to target device for disrupting normal operation
Disable Unsolicited Response	request a retraction of sending unsolicited response message
Unsolicited Response Storm	send a large number of unsolicited response messages to target
Unauthorized Request	send request message from unauthorized server
Non-DNP3 Communication on a DNP3 Port	send packets which have non-DNP3 payload

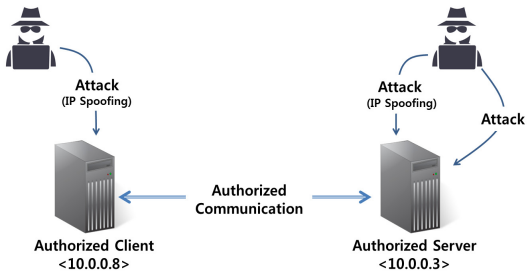

Fig. 3. Topology of experiment

구성하였다.

위 공격들은 모두 DNP3 프로토콜의 정상행위를 규정할 경우 탐지 할 수 있는 공격으로 구성되어 있어, 해당 공격들의 탐지에는 빅데이터 기술을 이용하여 트래픽의 특성을 파악하는 것이 핵심이라 할 수 있다. 실제 환경에서는 용량 문제로 패킷들을 모두 저장하기 보다는 로그 기록만 남기고 있는 경우가 많기 때문에, 생성한 패킷을 이용하여 Fig. 4와 같은 형태의 로그를 생성하였다. 이 때 IP Spoofing등의 공격 대응을 위해 DNP3 프로토콜의 하위 계층인 TCP와 IP 계층의 주요 필드들(IP Address, Port Number 등)도 로그로 생성하였다.

Fig. 4와 같은 형태로 로그를 생성하여 정상행위를 규정하고, 이를 기반으로 공격 탐지를 시도하였다. 이 실험에서 빅데이터 분석을 위해 분산처리 플랫폼인 하둡(Hadoop)과 하이브(Hive)를 이용하여 데이터의 개략적인 특성을 파악하고, 이후 통계 분석 도구인 R을 이용하여 통계적으로 비정상으로 판별되는 트래픽 검출을 수행하였고, 이후 이산 사건 모델 시뮬레이션 도구인 Arena를 이용하여 정상적인 침입 탐지가 가능한지 검증하였다.

5.1 Hadoop 및 Hive를 이용한 데이터 분석

Hadoop/Hive를 사용해 트래픽의 특징들에 대해 개략적인 분석을 수행하였다. 가장 첫 번째로 확인한 내용은 통신 대상들에 대한 식별이다. 이를 위해 수집된 트래픽의 IP와 패킷의 수는 Table 3과 같이 나타났고, 이를 통해 10.0.0.3과 10.0.0.8이 정상적인 서버와 클라이언트이고, 10.0.0.7이 비 허가된 접근을 하고 있다고 판단된다.

이후 대상 트래픽이 사용되고 있는 포트에 대한 내용을 확인하여 30,000번 대의 포트가 비정상적인 한두 개의 패킷만을 전송한 것을 확인하였고,

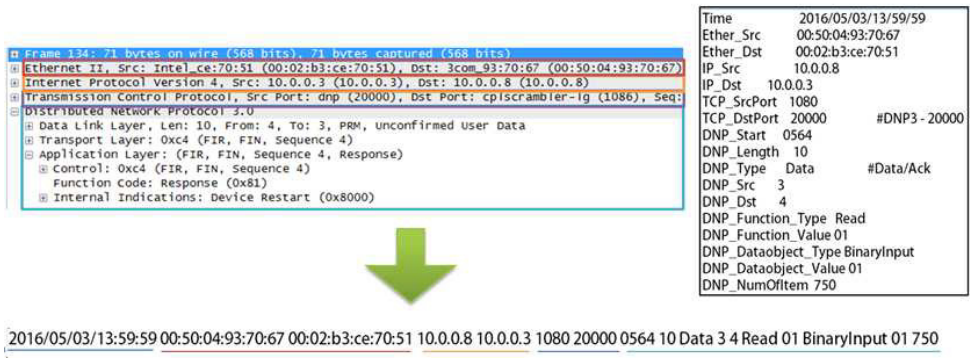

Fig. 4. An extracted protocol field and log form

30,000번대 포트의 동작 내용에 대해 출력한 결과 다음 Fig. 5와 같은 포트 스캐닝 공격을 받은 것으로 추측되는 포트의 리스트를 획득하였다.

DNP3가 사용하는 20,000번 포트로 들어온 패킷 로그를 살펴보면, DNP3의 시작을 나타내는 0564가 정상적으로 표기된 것이 100,781개, 비정상적으로 표시된 것이 총 2개로 확인된다. 비정상적인 값이 사용된 패킷은 DNP3 포트로 다른 프로토콜의 메시지를 전송하여 예상하지 못한 에러 상황을 유발하기 위한 공격으로 예상된다.

DNP3 패킷 로그 중 Item의 개수를 나타내는 DNP_NumOfItem 필드를 확인한 결과 비정상적으로 전체 Item을 요청하고 있는 패킷은 총 3개로 확인되었고, 각 Data Object별로 하나씩 전송되었음이 확인된다.

이러한 내용을 토대로 유추해볼 때, 공격자는 전체 Item 들을 파악하지 못하고 있어 모든 Data Object에 이러한 내용을 요청하여 전송되고 있는 데이터를 파악하고자 하였을 것으로 예상된다.

DNP3 로그 중 DNP3 Function Type에 대한 분석 결과 Read, Response, Confirm을 포함하여 총 10개의 Function이 사용되고 있음이 Table 4와 같이 확인되었다. Function들은 그 종류에 따라 개수가 크게 차이가 있어 비정상적인 내용이 포함되어 있을 것으로 예상된다. 특히 적은 양이 발견된 Function들의 경우 비정상적인 동작으로 판단된다.

Table 3. List of communication IP

Source IP	Destination IP	Count
10.0.0.3	10.0.0.7	1
10.0.0.3	10.0.0.8	21,559
10.0.0.7	10.0.0.3	4
10.0.0.8	10.0.0.3	79,219

Read는 데이터 분석 중 Data Object 별로 주기적인 Request와 Response가 이루어지고 있음이 확인되었다. Unsol Response는 주기성이 간단히 확인되지 않아 양일간의 Unsol Response 메시지의 수를 분석한 결과 특정 시간대에 패킷이 몰려있는 것이 확인되었다. 이 경우 해당 시간대에 Unsol Response 메시지를 이용한 DoS(Denial of Service) 공격이 있었을 것으로 추측 할 수 있으나, 주기적으로 보고하는 메시지가 중첩된 것일 수

```
hive> select * from pklog WHERE tcp_srcport like '%300%';
OK
pklog.time      pklog.ether_src pklog.ether_dst pklog.ip_src    pklog.ip_dst    pklog.tcp_srcport       pklog.t
rc      pklog.dnp_dst   pklog.dnp_function_type pklog.dnp_function_value       pklog.dnp_dataobject_type
03/19/11/11     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30000   20000
03/19/11/13     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30001   20000
03/19/11/15     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30002   20000
03/19/11/17     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30003   20000
03/19/11/19     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30004   20000
03/19/11/20     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30005   20000
03/19/11/23     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30006   20000
03/19/11/25     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30007   20000
03/19/11/27     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30008   20000
03/19/11/29     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30009   20000
03/19/11/31     00:50:04:93:70:67       00:02:b3:ce:70:51       10.0.0.8        10.0.0.3        30010   20000
Time taken: 0.13 seconds, Fetched: 11 row(s)
```

Fig. 5. Suspected log with port scanning attack

Table 4. Number of packets by function codes

Function Type	Count
Read	21,372
Response	21,262
Unsol Response	193
Confirm	185
Enable Unsol Response	14
Disable Unsol Response	2
Warm Restart	1
Stop Application	1
Cold Restart	1

있기 때문에 5.2절에서 통계적 분석 도구를 활용하여 추가 분석을 수행하였다.

이와 같이 트래픽의 특징과 프로토콜의 특성을 바탕으로 살펴본 결과 앞의 공격 분류 중 총 8가지를 손쉽게 파악 할 수 있었고, DNP3에 대한 큰 사전 정보 없이도 트래픽의 특징 도출에는 큰 어려움이 없다는 것이 확인되었다.

5.2 통계 분석 도구인 R을 이용한 정상 행위 규정

Unsolicited Response와 Read Request등 메시지들이 주기성을 가진다는 것을 Hadoop/Hive 분석 결과 개략적으로 확인 할 수 있었다. 따라서 이를 R을 이용하여 Fig. 6과 같이 시각적으로 나타내어 상관관계를 분석하였다.

Read Request와 Response의 경우 전체 시간 동안 계속 진행된 것이 확인 되었고, 앞서 확인한 내용과 동일하게 Warm Restart, Stop Application, Cold Restart, Disable Unsolicited Response 메시지가 전체시간 중 한 번 혹은 두 번 발견되는 것이 확인된다. Unsol_Response의 수와 Confirm의 수가 동일하게 증감하는 것을 토대로 이 두 Function이 연관성을 가지고 있다고 판단되며, 실제 패킷 로그를 확인한 결과 Unsol_Response 메시지를 받은 클라이언트가 서버 측으로 Confirm 메시지를 송신하는 것이 확인되었다.

전체 시간 중 적게 발견되는 Warm Restart, Cold Restart, Stop Application은 기기의 재시작 및 애플리케이션의 정지를 유발하여 정상적인 동작을 방해하는 공격으로 짐작된다. 공격 여부를 판단하기 위해 Stop Application이 발생한 시점에 다른 Function들의 동작 내용에 대한 파악이 필요하다. Fig. 7은 해당 시점에 Read 메시지와 Response메시지의 동작 내용을 상세히 나타낸 것이다. Fig. 7에서 화살 표시로 나타낸 시점에 Read 메시지가 전송되어야 하나, Stop Application으로 인해 예상되는 시점보다 더 늦게 Read 메시지가 발생한 것이 확인된다. Cold Restart, Warm Restart도 동일한 결과를 보이고 있어 이 세 가지 Function은 잠시 동안 대상 시스템을 마비시켜 정상 동작여부를 판별 할 수 없도록 하는 공격으로 판단된다.

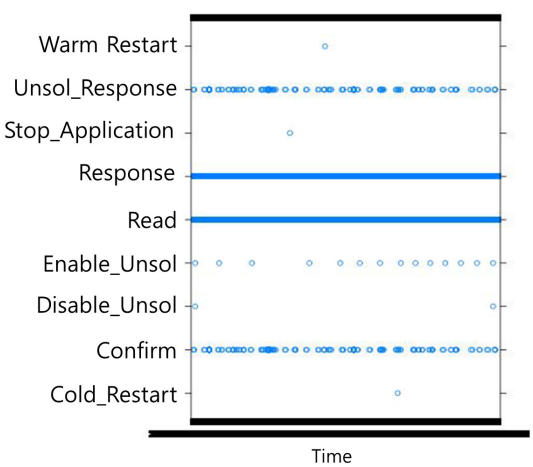

Fig. 6. Kind of DNP3 messages over time

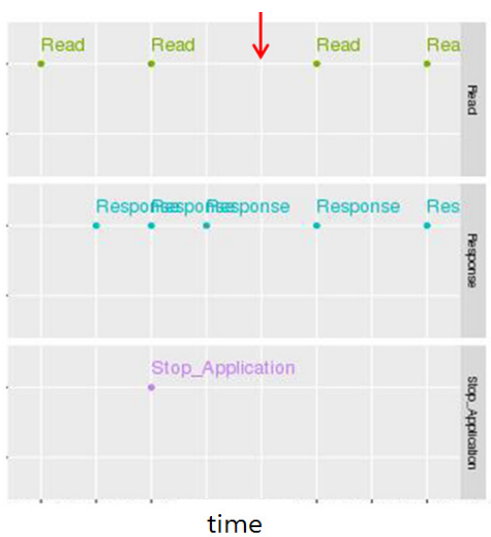

Fig. 7. Stop application occurrence point

Hadoop/Hive를 통해 Unsol Response의 패킷 수가 비정상적으로 많은 시간대가 존재함을 파악하였다. 이에 대한 상세 분석을 수행한 결과 Fig. 8과 같이 특정 시간에 수많은 Unsol Response 메시지가 전달되어 Unsolicited Response Storm 공격이 발생함을 확인 할 수 있다. 공격을 제외한 Unsol Response 메시지를 통계적으로 분석 한 결과 메시지 간의 시간 간격이 최소 1초 이상이라는 특징을 파악하였다.

이와 같이 통계적 분석 도구인 R을 이용하여 앞서 공격으로 의심된 행위들이 공격임을 입증하였고, 이 내용을 토대로 개발한 침입 탐지 방법을 검증 및 결과 분석을 수행한 내용을 5.3절에서 논의한다.

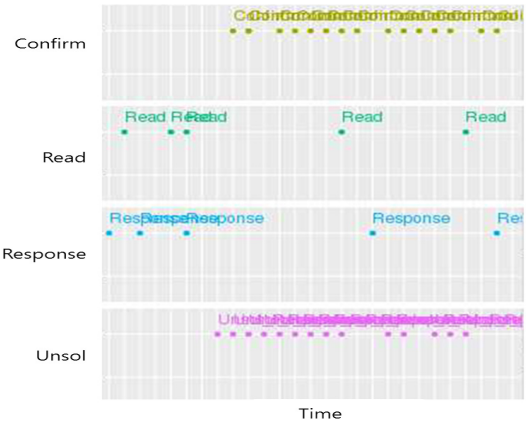

Fig. 8. An unsolicited response storm

5.3 Arena를 이용한 검증 및 검증 결과 분석

앞서 분석한 내용을 통해 총 11가지의 공격에 대한 대응 방안을 수립하였다. 허가된 통신 대상의 IP와 포트, 사용되는 Function, 특정 Function들의 정상 동작 시 나타나는 특징 등을 정상으로 규정하여, 정상행위 기반의 침입 탐지가 가능하다. 하지만 본 실험에 사용된 공격 중 일부는 시그니처를 기반으로 찾을 경우 더욱 간편하게 찾을 수 있는 요소들이 있어 이 둘을 병행하여 침입탐지 방법을 개발하였다.

정상행위를 기반으로 탐지하는 공격은 총 5가지로, 서버와 클라이언트의 IP, Port등 통신과 관련된 주소들, DNP3 프로토콜의 특성, 트래픽의 특징 등을 이용한다. Port Scan 공격의 경우 서버와 클라이언트의 통신에 활용되는 정상포트를 알고 있기 때문에 정상 포트만 통과할 수 있도록 설정하여 공격을 검출한다. Broadcast Request from Authorized Client 공격은 정상 DNP Destination을 알고 있기 때문에 Port Scan과 같은 방법으로 탐지 가능하다. 이 공격의 경우 DNP Destination을 Broadcast로 설정하여 보내기 때문에 시그니처 기반으로도 탐색이 가능하다. Unsolicited Response Storm은 앞서 정상적인 Unsol Response 메시지의 시간 간격이 최소 1초 이상이라는 것을 파악하였기 때문에 이를 정상으로 판단하고, 이보다 더 빠른 속도로 메시지가 전송되는 경우 공격으로 탐지할 수 있다. 이 외 Unauthorized Request, Non-DNP3 Communication on a DNP3 Port와 같은 공격들도 앞서 언급한 DNP3의 특징과 정상 IP/Port를 통해 대응할 수 있다.

시그니처를 기반으로 탐지하는 공격은 총 5가지로 공격에만 존재하는 시그니처를 통해 공격을 파악한다. 일례로 Warm Restart, Cold Restart공격은 Function code가 0x0E, 0x0D인 패킷을 막는 것으로 공격에 대응 가능하다. 이렇게 두 가지 방법을 사용하여 구성한 침입 탐지 시스템의 유효성을 검증하고자 Arena를 이용하여 모델링을 수행하였다.

모델링은 총 3 단계로 Fig. 9와 같이 구성된다. 첫 번째(Submodel - Extracting Data From Log)는 데이터를 읽어오는 단계로 로그파일에서 각 필드들을 추출하여 변수에 할당한다. 두 번째 단계(Detecting Suspected Log)는 읽어온 로그를 정해진 규칙과 비교하여 공격여부를 검사한다. 이때 검사규칙은 정상행위 기반 규칙과 시그니처 기반의 규칙 모두 활용된다. Function Code를 보고 판단 가능한 4가지 공격들은 모두 하나의 서브모델에서 처리하며, 나머지 공격들은 각 공격별로 탐지를 담당하는 서브모델을 생성하였다. 이 단계에서 각 공격들의 탐지와 탐지로그 생성이 수행된다. 세 번째 단계(Submodel - Checking that server is normal)는 두 번째 단계와 동시에 진행되는 단계로 로그를 남긴 패킷의 특성에 따라 이후 진행되는 ACK의 수를 계산하여 공격과는 무관하게 서버의 정상 동작 여부를 검사하도록 구성하였다.

총 1만개의 패킷로그를 이용하여 테스트를 수행하였고, 총 테스트 수행 시간은 1분 15초가 소요되었고, 공격은 모두 탐지되었다. 따라서 본 논문에서 제안한 기법이 유효하며, 빠른 시간에 효과적으로 공격을 검출 할 수 있는 침입탐지 방법을 개발 할 수 있

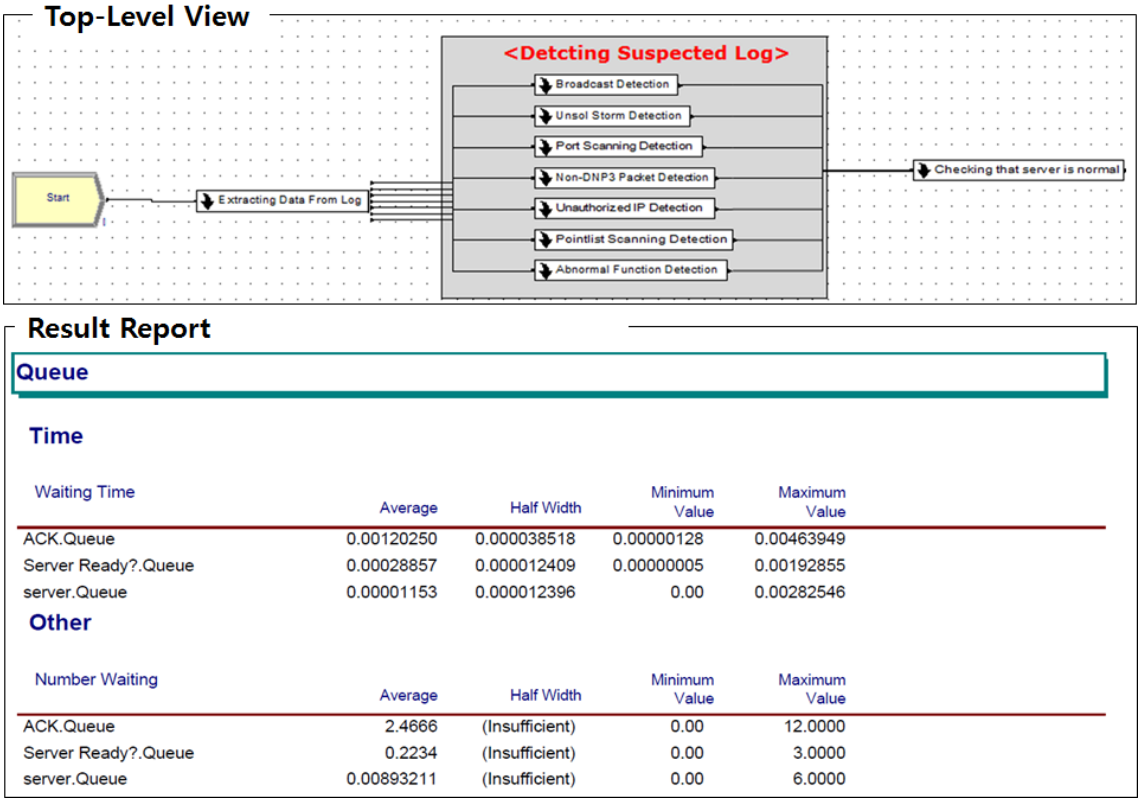

Fig. 9. DNP3 intrusion detection model for simulation

음이 입증되었다.

VI. 결 론

본 논문은 사용되고 있는 많은 양의 트래픽을 빅데이터 분석 도구를 활용하여 정상행위를 규정하는 방안을 제안하고, 실험을 통해 정상행위를 기반으로 생성한 규칙을 통해 공격을 성공적으로 탐지할 수 있음을 보였다. 제안한 기법은 트래픽 분석에 빅데이터 기법을 활용하고, 상세 분석이 필요한 필드들을 추출하여 정교한 정상행위를 규정 할 수 있기 때문에 다양한 프로토콜들이 사용되는 사물인터넷 환경에서 더욱 적합하게 활용 될 수 있을 것으로 판단된다. 본 논문에서 제안하는 정상행위 규정 기법은 프로토콜에 무관하게 대상 구간에 적합한 탐지 모델을 생성 할 수 있다. 따라서 5장 실험에서 수행한 DNP3 프로토콜의 침입 탐지 외에도 다양한 프로토콜에 적용하여 각 통신구간에 적합한 정상행위를 규정하고 이를 이용한 침입 탐지 시스템을 구축 할 수 있을 것으로 사료된다.

본 논문의 실험에서는 산업 제어시스템에서 널리 활용되고 있는 DNP3 프로토콜의 트래픽을 대상으로 제안하는 기법의 유효성을 입증하였다. 사용된 프로토콜이 바이너리 기반의 프로토콜이기 때문에 텍스트 마이닝 등의 기법은 활용되지 않았으나, 응용 계층이 텍스트 기반의 프로토콜일 경우 통계 분석 도구인 R을 이용한 텍스트 마이닝 등의 기법을 적용하여 더욱 상세한 정상 행위의 규정이 가능할 것으로 판단된다. 또한 본 실험에서는 빅데이터 분석 기술을 이용한 정상행위 정의가 유의미함을 보이기 위해 제외하였지만, 필드간의 상관관계 분석을 통한 유의미한 도출도 가능할 것으로 판단된다.

References

[1] Dorothy E. Denning, “Stuxnet: what has changed?,” Future Internet, vol. 4, no. 3, pp. 672-687, July 2012. doi:10.3390/fi4

030672
〔2〕 M. Krotofil and O. Yasynskyi, "Security Analysis of Cyber Attacks in Ukraine," Presented in Miami, 2017, https://www.slideshare.net/MarinaKrotofil/s4-krotofil-afternoonsesh2017
〔3〕 Hong-ryeol Ryu, Sung-mi Jung, and Tae-kyoung Kwon, "New paradigm of evolving threats - Advanced Persistent Threat (APT)," The Magazine of the Institute of Electronics and Information Engineers (IEIE), 41(4), pp. 16-30, Apr. 2014. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE02397067
〔4〕 RTV6 TheIndyChannel.com, "'Raptors ahead' sign gets stares, chuckles," Feb. 2009 http://www.theindychannel.com/news/-raptors-ahead-sign-gets-stares-chuckles
〔5〕 Sang-soo Hong, "〔Technology Trends : SIEM〕 Evolve into an intelligent log management platform," CiOCiSO Magazine, Jan. 2016, http://www.ciociso.com/news/articleView.html?idxno=10993
〔6〕 Do-keun Kim, Seong-bin Pyo, and Chang-hee Kim, "Study on APT attack response techniques based on big data analysis," Journal of Convergence Knowledge, 4(1), pp. 29-34, Jan. 2016. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE06606109
〔7〕 Sang-joon Lee and Dong-hoon Lee, "Real time predictive analytic system design and implementation using Bigdata-log," Journal of The Korea Institute of Information Security & Cryptology, 25(6), pp. 1399-1410, Dec. 2015. doi: 10.13089/JKIISC.2015.25.6.1399
〔8〕 Jong-hyun Kim, Sun-hee Lim, Ik-kyeun Kim, Hyun-suk Cho, et al. "Trend of cyber security techniques using bigdata," ETRI Electrics and Telecommunications Trends, 28(3), pp. 19-29, June 2013. doi:10.22648/ETRI.2013.J.280303
〔9〕 Kyo-il Chung, Hanna Park, Boo-geum Jung, Jong-soo Jang, and Myung-ae Chung, "Bigdata and information security," Korea Institute of Information Technology Magazine, 10(3), pp. 17-22, Sept. 2012. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE02034221
〔10〕 Hadoop, http://hadoop.apache.org/
〔11〕 Hive, https://hive.apache.org/
〔12〕 R project for statistical computing, https://www.r-project.org/
〔13〕 Arena simulation software, https://www.arenasimulation.com/
〔14〕 IoT Security Alliance of KISA, "IoT common security guide for security internalization of ICT convergence products and services," Sept. 2016. https://www.kisa.or.kr/public/laws/laws3_View.jsp?cPage=1&mode=view&p_No=259&b_No=259&d_No=80&ST=&SV=
〔15〕 Sung-moon Kwon and Tae-shik Shon, "Vulnerability and security status of control system DNP3 protocol," REVIEW of KIISC, 24(1), pp. 53-58, Feb. 2014. http://www.dbpia.co.kr/Journal/ArticleDetail/NODE02380941

〈저자소개〉

김 성 진 (SungJin Kim) 학생회원
2014년 2월: 아주대학교 정보 및 컴퓨터공학부 공학사
2014년 3월~현재: 아주대학교 컴퓨터공학과 석박사통합과정
〈관심분야〉 정보보호, 제어시스템 보안, 비정상행위 탐지

김 강 석 (kangseok Kim) 정회원
2007년: 인디애나대학교 컴퓨터공학(박사)
2010년~2016년: 아주대학교 대학원 지식정보공학과 연구교수
2016년~현재: 아주대학교 사이버보안학과 부교수
〈관심분야〉 클라우드 컴퓨팅, 유비쿼터스 컴퓨팅, 모바일 보안, 빅테이터 보안분석