EasyChair Preprint

№ 1276

Development of Big Data Extraction and Learning Platform for Packet Analysis in Industrial Control Syste

Kangbin Yim, Juyoung Seo, Chanmin Kim, Dain Kim and Soyoung Jung

July 10, 2019

산업제어시스템에서의 패킷 분석을 위한 빅데이터 추출 및 학습 플랫폼 개발

Development of Big Data Extraction and Learning Platform for Packet Analysis in Industrial Control System

요 약

산업제어시스템(ICS: Industrial Control System)은 산업 프로세스 운영 및 자동화에 사용되는 장치, 시스템, 네트워크 및 제어 장치 등 산업 구조에 따라 다양한 형태로 구성되는 제어시스템과 그 밖의 관련 기기를 통칭하는 용어이다. 80년대 이후 제조 산업의 급격한 발전과 함께 PLC(Programmable Logic Controller) 형태로 발전했으며, 이는 제어 관점에서의 복잡성을 줄이고 여러 시스템을 동시에 관리할 수 있는 환경을 요구했다. 따라서 다양한 제어 기기들을 하나로 통합하여 운영할 수 있는 Fieldbus와 같은 Ethernet기반의 여러 산업제어시스템 프로토콜이 등장하게 되었다. 그러나 Stuxnet, BlackEnergy, TRISIS-TRITION, IRONGATE 등과 같은 산업제어시스템을 대상으로 하는 공격들은 EWS(Engineering workstation), HMI(Human Machine Interface), SIS(Safety instrumented system) 등 각 장치에 대해 비정상적인 작동을 수행시켜 데이터 수집 및 파괴, 변전소 차단 등 산업시스템에 막대한 영향을 끼쳤으며 최근까지도 공격을 시도하고 있다. 이에 따른 산업제어시스템 취약점에 대한 연구가 진행되고 있으나 폐쇄적인 산업제어시스템의 운용 특성상 각 시스템의 취약점 분석 및 사전 대응을 위한 데이터 셋 확보가 비교적 쉽지 않다. 따라서 본 논문에서는 산업제어시스템에 대한 공격에 대비하기 위해 라즈베리파이 및 각종 센서를 활용하여 데이터의 통합 관리와 자동 제어를 제공하는 인공지능 산업제어시스템 프레임워크를 구축하고, 다양한 시나리오에서 발생할 수 있는 데이터 셋을 확보하여 정규화 학습을 통한 사전 공격 대응이 가능한 인공지능 산업제어 시스템을 구현한다.

ABSTRACT

The Industrial Control System (ICS) uses industrial structure and shape control systems and related systems such as industrial process operations and automation equipment, systems, networks and control devices. The configuration of a PLC (Programmable Logic Controller) must manage the system and manage the environment. Various industrial control system protocols based on fieldbus and Ethernet work. However, attacks targeting industrial control systems such as Stuxnet, Black

Energy, TRISIS-TRITION, and IRONGATE may perform abnormal operations on each device such as EWS (Engineering workstation), HMI (Human Machine Interface) and SIS (Safety instrumented system) Data collection and destruction, substation shutdown, and so on. However, due to the nature of the closed industrial control system, it is relatively difficult to analyze the vulnerability of each system and secure data sets for proactive response. In this paper, we propose an intelligent industrial control system framework that provides integrated management and automatic control of data by using raspberry pie and various sensors to prepare for attack against industrial control system, We propose an artificial intelligence industrial control system that can cope with dictionary attacks through normalization learning.

I. 서 론

산업제어시스템은 80년대 이후 제조 산업의 급격한 발전과 함께 PLC(Programmable Logic Controller) 형태로 발전했으며, 이는 제어 관점에서의 복잡성을 줄이고 여러 시스템을 동시에 관리할 수 있는 환경을 요구했다. 따라서 다양한 제어기기들을 하나로 통합하여 운영할 수 있는 Fieldbus와 같은 Ethernet기반의 여러 산업제어시스템 프로토콜이 등장하게 되었다. 그러나 Stuxnet, Blackenergy 등과 같은 산업제어시스템을 대상으로 하는 공격들은 EWS(Engineering Workstation), HMI(Human Machine Interface), SIS(Safety Instrumented System) 등 각 장치에 대해 비정상적인 작동을 수행시켜 데이터 수집 및 파괴, 변전소 차단 등 산업 시스템에 막대한 영향을 끼쳤으며 최근까지도 공격을 시도하고 있다. 이에 따른 산업제어시스템 취약점에 대한 연구가 진행되고 있으나 폐쇄적인 산업제어시스템의 운용 특성상 각 시스템의 취약점 분석 및 사전 대응을 위한 데이터 셋 확보가 비교적 쉽지 않다.

II. 관련연구

2.1. 산업제어시스템

산업제어시스템이란 산업 프로세스를 운영 및 자동화하는데 사용되는 장치, 시스템, 네트워크 및 제어 장치를 포함하는 여러 유형의 제어시스템 및 관련 기기를 통칭하는 용어이다. 산업 시스템 산업 구조에 따라 산업제어시스템은 다양한 형태로 구성될 수 있으며, 산업제어시스템에서 사용되는 기기와 프로토콜은 제조, 운송, 에너지 전송 등의 산업에 이용된다. 산업제어시스템은 EWS, PLC, HMI 등으로 이루어져 있다.

2.2. EWS

PLC와 직접적으로 연결 및 통신하며, 구동될 래더 로직을 프로그래밍하거나 기기 및 Fieldbus 등의 시스템 구성에 대한 설정값을 변경할 수 있다. 각종 센서나 액추에이터에 명령을 보내거나 동작 과정을 변경할 수 있으며, 그래픽 디스플레이를 통해 정보를 확인할 수 있다.

2.3. PLC

전체 시스템의 제어 구성요소로, 자동 제어 및 감시에 사용하는 제어 장치이며, 프로그램에 의해 입력을 순차적으로 처리하고, 출력 결과를 이용하여 연결된 외부 장치를 제어한다.

2.4. HMI

각 장치의 데이터를 사용자에게 친숙한 형태로 변환하여 보여주는 인터페이스 역할을 하는 장치이며, 이를 통해 관리자가 공정을 감시하고 제어한다.

2.5. Stuxnet

SCADA 시스템을 대상으로 공격하는 멀웨어로 발전소·공항·철도 등 기간 시설을 파괴 목적으로 제작되었다. 2010년 이란의 핵 발전 시설 내 우라늄 농축 시스템을 공격하는 사례로 발견되었다. 이는 원심분리기 마비 및 가동 중단을 야기하였으며, 이로 인해 수 개월간 핵무기 개발이 지연되었다. Stuxnet은 네트워크 또는 USB를 통해 전파되어 원자력 발전소의 소프트웨어와 장비 감염 및 제어가 가능케 함이 확인되었다. 또한 2011년 미국 일리노이 주 상수도 시설 시스템에 침투하여 상수도 펌프 시설 마비 및 수도 공급에 문제를 일으킨 바 있다. 이는 상수도 시설을 직접 공격하는 것이 아니라 SCADA 소프트

웨어를 만드는 업체를 공격하여 계정 정보 탈취 후 상수도 시설 내 접근한 것으로 밝혀졌다.

2.6. Blackenergy

시설물의 물리적 파괴를 목적으로 제작되어 2015년 우크라이나 대규모 정전사태 주원인으로 알려진 멀웨어다. 전력제어시스템 내부 침투 및 여러 변전소의 전력을 연속적으로 차단시킨 바 있다. 내부 직원들에게 매크로가 포함된 MS Office 문서 파일이 첨부된 메일을 송신하여 내부 네트워크에 접근한다. 문서 파일 내 매크로 스크립트가 악성 파일을 실행 및 공격 서버로의 접근을 야기한다. 또한 네트워크 스니핑 툴이 포함된 페이 로드를 통해 로그인 정보 및 시스템 내부로의 접근 등의 움직임을 보였다.

2.7. Trisis-Triton

슈나이더 일렉트릭의 SIS(Triconex Safety Instrumented System)을 대상으로 만들어진 멀웨어다. 실제 중동 지역의 석유 및 가스 추출 공장 등에서 흔하게 사용되는 시스템으로 Triton은 해당 시스템 내 정상 애플리케이션으로 위장하여 존재한다. 시설물의 물리적 파괴를 목적으로 제작되었으며 해당 시스템 기능 및 성능을 조작한다. 이는 2017년 사우디 석유 화학 공장 시설 내 침투하여 안전시스템 가동을 중단시킨 바 있으며 시설물의 안전과 관련된 ICS를 표적으로 삼았음이 밝혀졌다.

2.8. A Dataset to Support Research in the Design of Secure Water Treatment Systems

iTrust는 각종 시스템의 보안 및 안전성을 보장하기 위한 고급 도구 및 방법론 개발을 연구하는 싱가포르 기술대학의 연구센터로 ICS 관련 테스트베드와 데이터 셋을 보유한 연구기관이다. 수처리 시스템(Secure Water Treatment Testbed)인 SWaT와 물 분배 시스템(Water Distribution) WADI, 지능적 전력 제어 시스템(Electric Power Intelligent Control)인 EPIC 세 가지 테스트베드를 보유하고 있으며, 각 테스트베드에서 정상적인 상황의 데이터 셋과 공격이 발생한 경우의 데이터 셋을 확보하고 있다. iTrust는 SWaT를 이용하여 총 11일의 데이터 셋을 확보하였으며, 이 중 처음 7일 간은 정상적인 상태로, 4일 간은 여러 방식의 공격 시나리오를 적용한 상태로 데이터를 수집하였고, 네트워크 트래픽과 SWaT에서 사용된 51개의 모든 센서와 액추에이터의 946,722개의 데이터 수집에 성공하였다.

2.9. Industrial Control System Traffic Data Sets For Intrusion Detection Research

본 연구에서는 표준적인 데이터 셋이 각각의 독립적인 산업제어시스템에서 온전히 적용하기 어려운 문제점을 해결하기 위해 산업제어시스템의 취약점을 판단할 수 있는 플랫폼을 개발했다. 해당 플랫폼에서는 MODBUS 애플리케이션 계층 프로토콜을 사용하는 2개의 산업 제어 시스템에 대한 28가지 공격에서 발생하는 네트워크 트래픽, 프로세스 제어 및 프로세스 측정 기능을 포함하는 4가지 데이터 셋을 확보할 수 있으며, 데이터 셋은 실험실 환경에서의 직렬 포트 데이터 로거로 캡쳐한 네트워크 레코드에서 생성되었다. 해당 데이터 세트를 생성하기 위해 28가지의 공격을 정찰, 응답 주입, 명령 주입, 서비스 거부 공격의 네 가지로 실험했다. 따라서 해당 플랫폼에서 구성된 4개의 데이터 셋은 서로 다른 SCADA 침입 탐지 접근 방식 및 구현 비교에 사용될 수 있음을 확인했다.

III. ICS 구현 및 분석 툴 개발

3.1 ICS 구현

3.1.1 배경 및 환경 설정

데이터 셋을 분석하는 툴을 개발하기 위해 임베디드를 활용하여 산업제어시스템을 구현하였다. 장비로는 라즈베리파이는 B+를 사용하였고, os는 라즈비안 4.14 version을 설치하였다. 다른 부품으로는 서보 모터로 조종되는 로봇 팔이 있고, 그 환경을 받쳐줄 온습도 등 여러 센서들로 구성되어 있다.

3.1.1.1 서보 모터

서보 모터는 명령을 따르는 모터라는 의미에서, 정

확한 위치와 속도를 맞출 수 있다. 회전 각도의 원리는 PWM 방식으로 제어한다. 이 해당 서보 모터는 전체 20ms의 PWM 주기 중 1~2ms사이의 파형을 통해 각도를 제어한다. 예를 들면 1ms만큼 HIGH 신호를 주면 0도를 가리키고, 2ms만큼 HIGH 신호를 주면 180도를 가리키게 된다. 여기서는 로봇 팔을 구성하는 데 사용하였으며, mg996r이라는 모델을 사용하였다.

Fig. 1. mg996r

3.1.1.2 온습도 센서

온습도 센서는 DHT-11로 사용하였고, 여기서는 전체적인 온습도를 파악하여, 시스템 전체 환경에 대한 파악을 목적으로 한다. 아래 그림은 해당 사진이다.

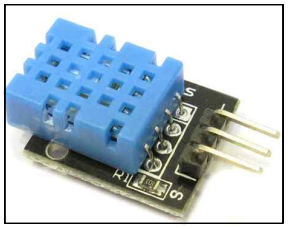


Fig. 2. DHT-11

3.1.1.3 진동 센서

진동 센서는 Sw-420라는 모델을 사용하였고, 로봇에 부착하여 움직일 때마다 진동을 감지하는 목적으로 한다. 아래 그림은 해당 사진이다.

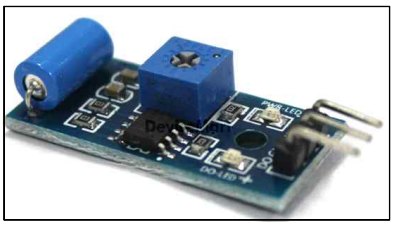


Fig. 3. Sw-420

3.1.1.4 전체적인 사진

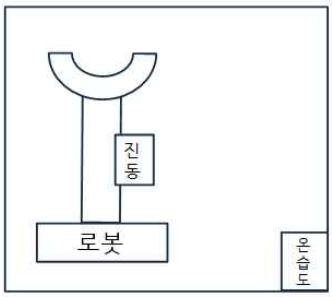


Fig. 4. The overall picture

3.1.2 작동 과정

3.1.2.1 로봇 팔

로봇 팔은 서보모터의 작동으로 움직이며, 여기서는 하나의 물건을 집어서 옮기는 데 사용이 된다. 로봇의 구상도는 아래 그림과 같다.

.

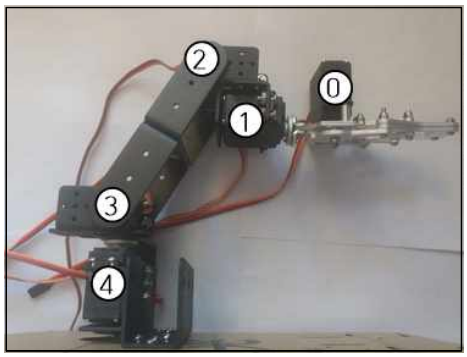
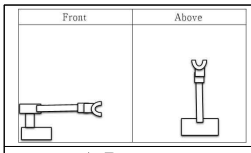
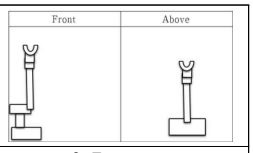
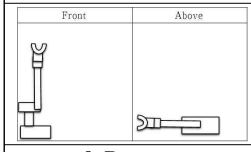
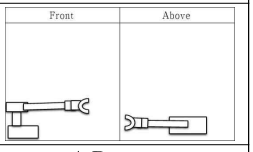
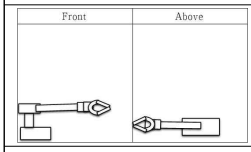
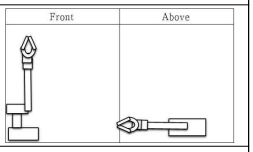
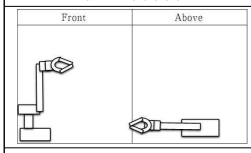
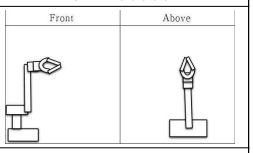
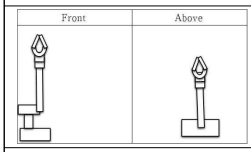
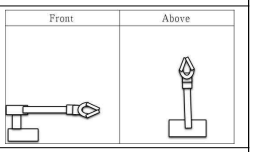


Fig. 5. The overall Robot Arm picture

그림과 같이 5개의 서보모터로 구성되어 있으며, 여기서는 1번 서보모터는 사용하지 않는다. 동작 과정은 다음 표와 같다.

Table 1. Operation process of Robot Arm

 1 Process	 2 Process
 3 Process	 4 Process
 5 Process	 6 Process
 7 Process	 8 Process
 9 Process	 10 Process

각 과정 같은 경우 왼쪽 그림은 옆에서 본 시점이고, 오른쪽 그림은 위에서 본 시점이다. 전체적인 동작은 물건을 집으러 90도 회전한 후에, 물건을 집은 후 다시 돌아오는 동작을 반복적으로 시행한다. 각 과정은 1초씩 시간을 가지고 움직인다. GPIO 핀은 12, 16, 20, 21, 26번을 사용한다.

3.1.2.2 온습도 센서

온습도 센서는 Adafruit_DHT 라이브러리를 사용하였고, GPIO 핀은 2번을 사용하였다. 1초마다 온도와 습도를 받아와 실시간 업데이트를 해준다.

3.1.2.3 진동센서

진동 센서는 Vcc와 Gnd를 연결하면, 자동으로 GPIO input이 들어오는데 결과가 1이면 진동 여부가 감지되었음을 의미하고, 0이면 미감지를 의미한다. 핀은 23번을 사용하였다.

3.1.3 실행 툴 및 사진

3.1.3.1 실행 툴

데이터를 얻기 위해 통합 실행 프로그램을 제작하였다. 환경은 Python을 활용하였으며, 로봇 팔 동작 과정을 표현하고자 Gui로 표현하였다. 라이브러리는 PyQt5를 주로 사용하였고, time, pandas 등 데이터 셋을 확보하고 파일로 저장하는 라이브러리를 사용하였다.

3.1.3.2 실행 사진

코드는 라즈베리파이에서 실행하였으며, 메인화면에서 분석 시작이라는 버튼을 누르면 GPIO를 통해 로봇 및 센서들을 동작시키고, 실시간으로 데이터를 받기 시작한다.

Fig. 6. ICS Tool main window

아래 사진은 분석 시작을 했을 때의 화면이고 로봇 팔 현재 상태를 보여준다. 그리고 그 위에 온도 및 습도를 표시한다. 온도, 습도의 경우 실시간으로 표시해야하기 때문에 쓰레드로 구성하였다.

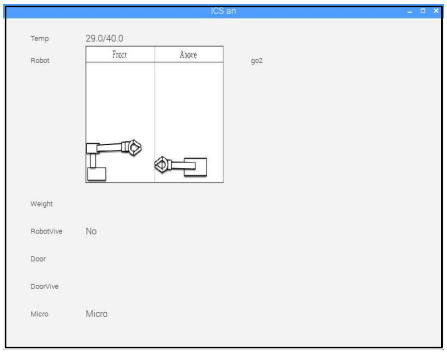


Fig. 7. Collecting Real-time Data set

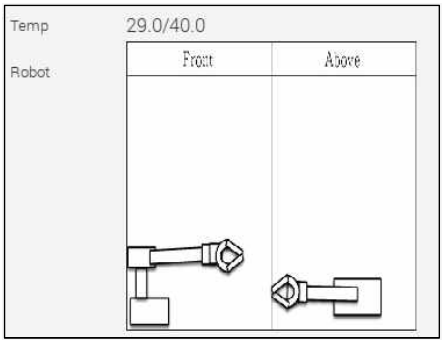
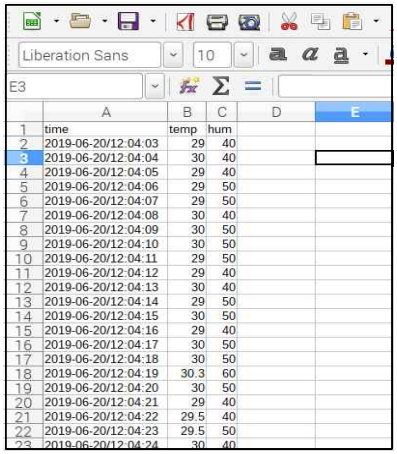


Fig. 8. Collecting Real-time Data set (Detail)

이 데이터들은 csv파일 형태로 현재 시간과 함께 저장된다.



	A	B	C	D	E
1	time	temp	hum		
2	2019-06-20/12:04:03	29	40		
3	2019-06-20/12:04:04	30	40		
4	2019-06-20/12:04:05	29	40		
5	2019-06-20/12:04:06	29	50		
6	2019-06-20/12:04:07	29	50		
7	2019-06-20/12:04:08	30	40		
8	2019-06-20/12:04:09	30	50		
9	2019-06-20/12:04:10	30	50		
10	2019-06-20/12:04:11	29	50		
11	2019-06-20/12:04:12	29	40		
12	2019-06-20/12:04:13	30	40		
13	2019-06-20/12:04:14	29	50		
14	2019-06-20/12:04:15	30	50		
15	2019-06-20/12:04:16	29	40		
16	2019-06-20/12:04:17	30	50		
17	2019-06-20/12:04:18	30	50		
18	2019-06-20/12:04:19	30.3	60		
19	2019-06-20/12:04:20	30	50		
20	2019-06-20/12:04:21	29	40		
21	2019-06-20/12:04:22	29.5	40		
22	2019-06-20/12:04:23	29.5	50		
23	2019-06-20/12:04:24	30	40		

Fig. 9. csv format file of Data set

3.2 분석 툴

3.2.1 배경 및 환경

3.2.1.1 배경

설치된 ICS의 데이터 셋에 대한 분석이 필요하기 때문에 툴을 제작했다. 이를 통해 데이터의 이상 유무를 자동으로 판단하여 패킷을 차단할 수 있게 하는 것이 목표이다. 또한 그래프를 통해 구체적인 분석에 도움을 준다.

3.2.1.2 환경

환경은 python 언어를 활용하여 제작했으며, ICS 실행 툴과 동일하게 PyQt5를 이용하여 Gui형태로 제작하였다. 데이터 셋에 대한 분석을 위한 pandas, numpy 라이브러리, 머신러닝 라이브러리는 sklearn을 사용하였다. 또한 matplotlib 라이브러리를 통해 그래프를 시각화하였다.

3.2.2 실행 사진

3.2.2.1 처음 실행 화면

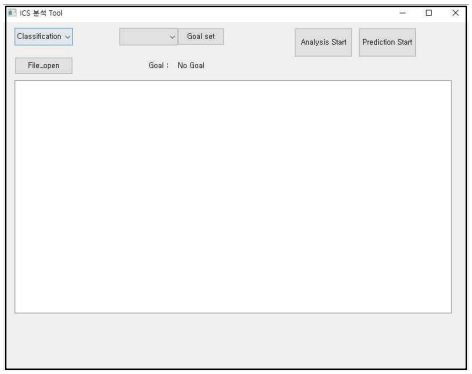


Fig. 10. Analysis Tool - main window

기본 실행 화면은 위의 사진과 동일하고 파일을 불러온 후에 그래프를 통해 분석할 것인지, 머신러닝을 통해 예측 모델을 만들 것인지에 대해 버튼 선택이 가능하다. 그래프 분석을 할 시에는 파일을 불러온 후에 진행해야 한다.

3.2.2.2 데이터 분석 화면

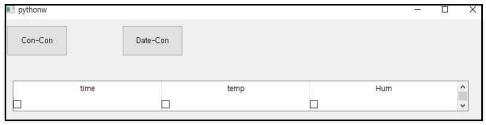


Fig. 11. Analysis Tool - Analysis window

위에 사진은 분석을 진행하는 버튼을 눌렀을 때 나타나는 화면이다. 여기서는 변수들을 선택하여 그래프로 표현할 수 있다. 현재 두 데이터 칼럼을 선택하여 표현할 수 있는데, 연속형, 연속형 변수이거나 연속형, 시계열 변수일 때 그래프로 표현할 수 있다. 여기서 연속형 변수란 몸무게, 나이, 온도와 같이 연속된 숫자로 이루어진 데이터, 시계열 변수는 날짜 데이터를 의미한다.

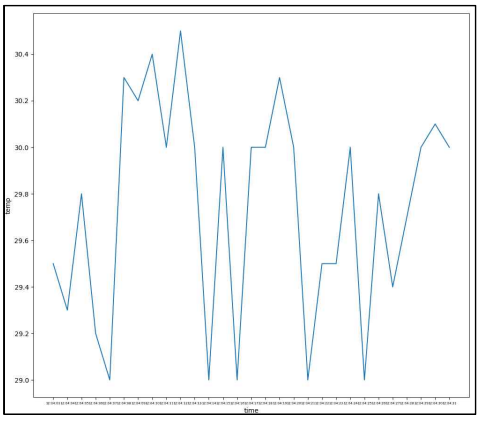


Fig. 12. Analysis Tool - Data Graph ver1

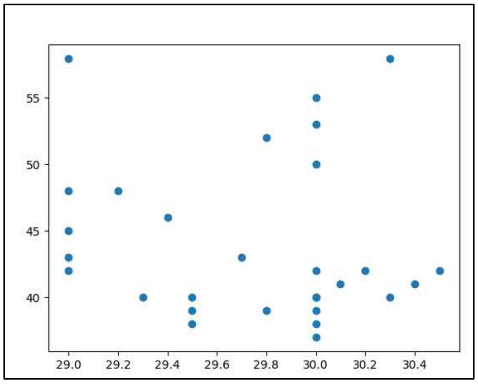


Fig. 13. Analysis Tool - Data Graph ver2

3.2.2.3 데이터 예측 화면

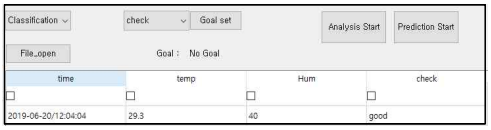


Fig. 14. Setting for data predicting

위 사진을 통해 예측에 사용하고자 하는 칼럼을 체크 박스에 표시해 준 후 구하고자 하는 라벨을 옵션 박스에서 선택하여 Goal set이라는 버튼을 눌러 설정해 준다. 그리고 Prediction Start라는 버튼을 누르면 예측 모델이 생성된다. 이때 예측 모델은 서포트 벡터 머신(svm) 알고리즘을 사용하였다. 이 알고리즘은 지도학습 모델로서 주어진 데이터 집합을 바탕으로 새로운 데이터가 어느 카테고리에 속할지 판단하는 이진 선형 분류 모델을 만드는 데 많이 사용된다.

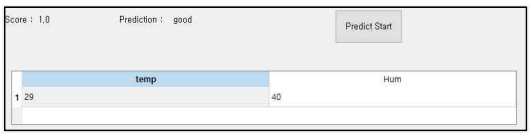


Fig. 15. Data predicting window and example

위의 사진을 통해 모델의 점수와 사용자가 입력한 값에 따른 예측이 가능하게 구성되어있다. 예시로, 정상 데이터인 temp는 29, Hum는 40을 입력하였

을 때 good이라고 나오고, 둘 또는 하나에 비정상적인 데이터를 대입하여 예측하였을 때 bad라고 나오는 것을 볼 수 있다.

IV. 결 론

산업제어 시스템의 폐쇄적인 특징에 따라 데이터 셋 확보의 어려움과 악의적인 사용자로서의 공격에 대응하기 힘들다는 취약점이 존재하였다. 본 연구에서는 라즈베리파이B+와 여러 센서들을 활용해 산업제어시스템을 구현하여 여러 시나리오에 대한 임시적인 데이터 셋을 수집하였고, 이에 대해 이진 분류 알고리즘 중 SVM을 활용하여 지도 학습을 진행하여, 비정상적인 패킷을 식별함으로써 취약점 대응이 가능해졌다. 최종적으로 인공지능 기술을 추가한 패킷 분석 및 관리할 수 있는 산업제어시스템용 프레임워크를 개발하였다. 이 연구를 통해 인공지능 기술을 폐쇄적인 산업제어시스템 취약점 연구에 활용하는 것은 충분한 가능성이 있음을 발견하였고, 실제 ICS에 활용되는 많은 시나리오에 대한 데이터 셋 분석과 학습을 통해 사전 공격에 대한 대처가 가능할 것으로 기대한다.

References

[1] Do-Yeon Kim, "Vulnerability Analysis for Industrial Control System Cyber Security", The Journal of the Korea Institute of Electronic Communication Sciences, p.137-142, 2014

[2] Gyuwon Hwang, "Power Control System Security Evaluation Items Through Analysis of Security Threat Cases", Soongsil University Graduate School, VIII, pp.53-55, 2017

[3] Jae-jun Heo and Sang-choul Lee, "Paths and Countermeasures of Stuxnet's infection", Korea Institute of Information Security and Cryptology, vol. 21(7), pp. 23-29, 2011

[4] Jun-Hyoung Oh, Young-In You and Kyung-ho Lee, "Standard Trends of Infrastructure Accidents and Control Systems", Korea Institute of Information Security and Cryptology, vol.27(2), pp.5-11, 2017

[5] Thomas Morris and Wei Gao , "Industrial Control System Traffic Data Sets for Intrusion Detection Research", ICCIP Critical Infrastructure Protection VIII, vol. 441, pp.65-78, 2014

[6] Sridhar Adepu, Khurum Nazir Junejo and Aditya Mathur, "A Dataset to Support Research in the Design of Secure Water Treatment Systems", iTrust, Center for Research in Cyber Security, Singapore University of Technology and Design, Singapore

[7] Hyerim Bae, Sanghyuck Park, Yulim Choi, "Operational Big Data Analytics platform for Smart Factory", The Journel of the Koera Institute of BigData, 2016

[8] Chae-Soo Kim, Seung-Beom Son ,"A Study on Big Data Cluster in Smart Factory using Raspberry-Pi", IEEE International Conference on Big Data, 2018

[9] Hui ZHAO, Jianrong, "Design Concernsfor Industrial Big Data SystemintheSmart Factory Domain:from Product Lifecycle View", 23rd International Conference on Engineering of Complex Computer Systems, 2018