침입 감지 시스템에서 등락 패턴을 이용한 False Alarm 감소에 관한 연구

박선우, 박현재, 최영준
아주대학교

{romaakk, Estancia, choiyj} @ajou.ac.kr

A Study on the Reduction of False Alarm Using the Fluctuation Pattern in Intrusion Detection Systems

Park Sun Woo, Park Hyun Jae, Choi Young Jun
Ajou Univ.

요 약

본 논문은 IIoT 제어시스템을 위한 AI 기반의 침입 감지 시스템을 학습시키는 데에 있어서, IoT 센서의 측정 오류로 인한 False Alarm 을 최소화하는 것에 대한 연구이다. AI 기반의 침입 감지 시스템(IDS)에서는 정상 데이터가 공격으로 분류되는 False Alarm 의 비율을 낮추는 것이 큰 과제이다. False Alarm 은 디지털화 된 산업현장에서 큰 영향을 끼친다. 각각의 알람에 대하여 시스템의 보안 책임자가 직접 공격인지 아닌지를 확인해야 하기 때문에, False Alarm 의 비율이 높을수록 자연스레 진짜 공격을 놓칠 확률이 커지기 때문이다. 이에, 본 논문은 시계열 순서를 따르는 데이터의 국지적인 등락 패턴을 이용한 AI 모델의 학습 방법을 통해, 임의로 발생하는 센서의 측정 오류에 의한 False Alarm 을 유의미하게 줄이는 데에 도움을 주고자 하였다.

Ⅰ. 서 론

오늘날, 컴퓨터 및 네트워크 기술이 현대인의 실생활 곳곳으로 확산되었다. 기업이나 조직의 시스템들 또한 대부분 컴퓨터 시스템으로 대체되어 자산을 보호하기 시작했다. 공격자로부터 이러한 자산과 네트워크를 보호하기 위해서, 네트워크 내외부의 트래픽을 분석해 공격을 감지해내는 침입 감지 시스템(IDS)의 필요성이 크게 대두되고 있다.

하지만, IDS 도 만능은 아니기에 실제 공격이 아닌 정상 트래픽을 공격으로 판단하는 경우가 종종있다. 이를 False Alarm 이라고 칭한다. False Alarm 은 보안 책임자로 하여금 시스템의 관리를 어렵게 하는 주범으로, 이를 줄이는 것이 IDS 개발의 주된 과제 중 하나이다.

이러한 False Alarm 은 디지털화 된 산업 현장에서 큰 영향을 끼친다. 각각의 알람에 대하여 보안 책임자가 직접 공격인지 아닌지를 확인해야 하기 때문에, False Alarm 의 비율이 높을수록 자연스레 진짜 공격을 놓치게 될 확률이 커지기 마련이고 이는 현장 장비의 손실로 이어질 수 있다.

이에, 본 논문은 IIoT 제어시스템을 위한 AI 기반의 IDS 를 학습시키는 데에 있어서, 임의로 발생하는 센서의 측정 오류로 인한 False Alarm 을 최소화하는 것을 목표로 한다. 이를 위해 시계열 데이터의 국지적인 등락 패턴을 이용하는 방법을 제안한다.

IDS 에서의 False Alarm 을 줄이기 위한 선행 연구로는 침입탐지 시스템에서 Alert 의 패턴 학습을 이용한 False Positive 감소에 대한 연구[1]가 존재한다. 해당 연구는 IDS 를 통해 데이터를 분류, 발생한 Alert 를 Event 생성기에 학습시켜 Knowledge 를 만든다. 이후, 실시간 네트워크에서 들어오는 Alert 를 Event 화 하여 기존의 Knowledge 와 비교하여 False Alarm 을 감소시키는 방법을 제안하였다. 본 연구에서는, Alert 의 패턴을 학습하여 비교하는 대신 IDS 의 학습 과정에서 데이터의 전처리를 통해 False Alarm 을 감소시키는 데에 집중하였다.

Ⅱ. 본론

1. 데이터셋

학습데이터로는 국가보안기술연구소가 지난 02.12 에 공개한 AI 학습용 산업 제어시스템 보안 데이터셋 'HAI 1.0'[2]을 이용하였다. HAI.10 은 IIoT 산업제어시스템 테스트베드에서 다양한 공격을 시뮬레이션한 데이터로, 시스템의 각 포인트에서 측정된 값들과 공격 여부, 종류로 이루어진 시계열 데이터이다. IDS 의 평가에 주로 이용되어온 1999 DARPA 나 NSL-KDD 데이터셋과 비교하였을 때, HAI1.0 이 비교적 최근의 환경과 공격을 더 잘 반영하기에 본 연구의 학습데이터로써 선택하였다.

2. 모델 선택

IDS 개발을 위한 AI 모델로는 XGBoost[3]모델을 선택했다. XGBoost 는 그래디언트 부스팅 알고리즘으로, 앙상블 트리에 기반한다. XGBoost 는 그 탁월한 성능으로, 2016 년 공개이후 많은 기계학습 대회에서 널리 이용되온 바 있다. 그러나, XGBoost 는 시계열 정보를 시간대별로 독립적으로 다루기에 후술할 데이터 전처리를 통해 각 시간대 사이의 관계정보를 추가하는 작업을 해 주었다.

3. 데이터 전처리

먼저, 908,997 개의 데이터를 8:2 의 비율로 시계열 순서를 지키며 학습과 테스트데이터로 나누었다. 각 시간대 별로 센서들의 측정 값을 feature 로 사용하고, 공격 여부를 class 로 사용하였다.

그 후에는 후술할 2 개의 포인트에 대하여 등락 패턴 정보를 추가하였다. 정상상태의 데이터는 일반적으로 주변 시간대와 비슷한 측정 값이 계속 유지된다. 공격의 경우는 탐지를 피하기 위하여 측정 값을 서서히 증가시키거나 감소시키는 패턴이 많으므로, 정상상태처럼 주변 시간대와 비슷한 측정 값이 유지된다. 반면, 임의로 발생하는 센서 오류는 주변 시간대의 측정 값과 비교해 급등하거나 급락하는 모습을 보인다. 주변

시간대의 측정 값의 등락 패턴을 통해서 공격과 센서 오류를 구분지을 수 있다고 판단, 각 포인트별로 등락 패턴 정보를 추가하였다.

포인트의 등락을 표현하기 위해서, 연속된 이전 5 개 시간대의 측정 값을 등락의 기준으로 이용하였다. Window 크기 5 의 Rolling Mean(5mean), Std(5std)을 계산, 등락의 Upper bound 로는 (5mean+ 5std/2)를, Lower bound 는 (5mean-5std/2)로 하여 UP/MID/DOWN 상태를 나누었다. 경계를 기준으로 5 개 시간대에서의 등락을 표현하면 총 243 개의 등락 패턴이 되므로, 더 보편적인 등락 패턴을 얻기 위해 Window 내에서도 3 개씩 평균을 내어 대표하도록 하였다. 결과적으로, 등락 패턴을 27 가지로 나타낼 수 있었다.

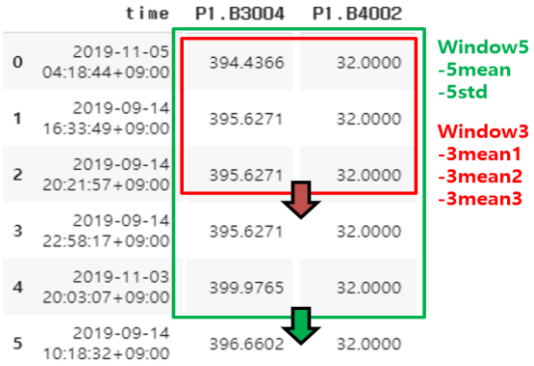

그림 1. 크기 3, 5 의 window

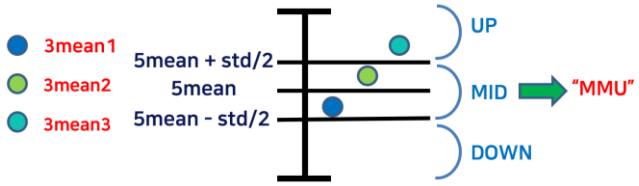

그림 2. 등락 패턴의 계산

등락 패턴 정보를 추가한 데이터에 대하여, PCA 를 통해 데이터의 차원을 축소하여 전처리를 마무리하였다.

4. 테스트 데이터 생성

오류 데이터에 대한 AI 모델의 성능을 평가할 수 있도록 테스트 데이터로써 다양한 오류 데이터셋을 생성하였다. 테스트 데이터에서 공격과 관련된 포인트 A(P1.B3004)와 관련이 없는 포인트 N(P1.B4002)을 오류가 삽입될 포인트로 선정하였다. 기존의 테스트 데이터를 200 개씩 섹터로 나누고 각 섹터에서 임의의 시간대를 선택하였다. 선택된 시간대의 원래의 측정 값에 보정치를 더해 정상 범위를 벗어나도록 했다. 보정치를 더한 오류 값은 아래와 같으며, 각 포인트의 정상범위는 HAI1.0 의 description 을 참고하였다.

$$Value_{new} = Value_{origin} + Mean_{sector} * Ratio * times$$

$$Value_{new} > Max + Mean_{sector} * Ratio * times$$

$$Ratio = 0.1, times = 0.1, times\ 성장배수 = +0.15$$

5. 모델 평가

학습데이터에 대하여 위의 전처리 방식으로 학습된 XGBoost 모델을 실험군으로, 위의 전처리 방식에서 등락 패턴 정보를 제거하여 학습된 모델을 대조군으로 하여 결과를 비교하였다. 테스트 Metric 으로, 분류 정확도와 공격으로 오분류 된 오류 데이터의 비율을 측정하였다.

포인트	N		A		A, N	
오류비율	정확도	오분류 E	정확도	오분류 E	정확도	오분류 E
5%	87.69	6.59	85.20	65.63	84.27	87.84
10%	87.46	6.54	82.54	64.97	80.55	88.59
15%	87.20	6.81	79.84	65.18	76.93	88.14
30%	86.51	6.66	71.72	65.27	65.93	88.22

표 1. 대조군 테스트 결과(%)

포인트	N		A		A, N	
오류비율	정확도	오분류 E	정확도	오분류 E	정확도	오분류 E
5%	85.91	11.14	85.32	25.22	84.61	41.82
10%	85.50	11.02	84.33	25.10	82.93	41.49
15%	85.10	10.74	83.27	25.43	81.22	41.40
30%	83.87	10.59	80.16	25.46	76.03	41.56

표 2. 실험군 테스트 결과(%)

N		A		A, N	
정확도	오분류 E	정확도	오분류 E	정확도	오분류 E
-2.12	4.22	3.44	-39.95	4.28	-46.63

표 3. 실험군-대조군 평균(%)

공격과 무관한 포인트인 N 에만 오류를 넣은 경우에는 정확도가 2%가량 감소하고 오분류된 오류비율이 4%가량 증가하는 결과를 보였다. 반면, 공격과 관련된 포인트 A 혹은 양쪽에 오류를 넣었을 때는 3.44%. 4.28%의 정확도 상승과 39.95%, 46.64%의 유의미한 오분류 비율 감소를 확인할 수 있었다.

Ⅲ. 결론

본 논문에서는 IIoT 제어시스템을 위한 AI 기반의 침입 감지 시스템을 학습시키는 데에 IoT 센서의 측정 오류로 인한 False Alarm 의 최소화를 위해, 시계열 순서를 따르는 데이터의 국지적인 등락 패턴을 이용하는 방법을 제안하였다. 등락 패턴을 이용해 학습한 실험군과 대조군을 통해 공격의 분류정확도 및 공격으로 오분류된 오류데이터의 비율을 비교하였다. 결과로, 분류 성능을 크게 해치지 않으면서 임의로 발생하는 센서의 측정오류에 의한 False Alarm 의 비율을 낮출 수 있었다.

향후 연구 방향으로, AI 학습 모델을 바꾸어 연구할 예정이다. XGBoost 는 각 데이터를 독립적으로 처리하는 데에 비해 LSTM 은 이전의 데이터에 기반한 예측이 가능하여 더 좋은 결과를 얻을 수 있을 것으로 생각한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음 (2015-0-00908)
본 연구는 한국전력공사의 2018 년 선정 기초연구개발 과제 연구비에 의해 지원되었음(과제번호 : 18A-013)

참 고 문 헌

[1] 심철준. "침입탐지 시스템에서 Alet 의 패턴학습을 이용할 False Positive 감소에 대한 연구." 국내석사학위논문 건국대학교 대학원, 2003. 서울

[2] zaroosin, “hai”, 2020, Github, https://github.com/ icsdataset/hai

[3] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.