

Toward Detection and Attribution of Cyber-Attacks in IoT-enabled Cyber-physical Systems

Maruboyina Naga Prasanna

PG scholar, Department of MCA, CDNR collage, Bhimavaram, Andhra Pradesh.

A.Naga Raju

(Assistant Professor), Master of Computer Applications, DNR collage, Bhimavaram, Andhra Pradesh.

Abstract: Securing Internet-of-Things (IoT)-enabled cyber-physical systems (CPS) can be challenging, as security solutions developed for general information/operational technology (IT/OT) systems may not be as effective in a CPS setting. Thus, this article presents a two-level ensemble attack detection and attribution framework designed for CPS, and more specifically in an industrial control system (ICS). At the first level, a decision tree combined with a novel ensemble deep representation-learning model is developed for detecting attacks imbalanced ICS environments. At the second level, an ensemble deep neural network is designed to facilitate attack attribution. The proposed model is evaluated using real-world data sets in gas pipeline and water treatment system. Findings demonstrate that the proposed model outperforms other competing approaches with similar computational complexity.

INTRODUCTION

Internet of Things (IoT) devices are increasingly integrated in cyber-physical systems (CPS), including in critical infrastructure sectors such as dams and utility plants. In these settings, IoT devices (also referred to as Industrial IoT or IIoT) are often part of an Industrial Control System (ICS), tasked with the reliable operation of the infrastructure. ICS can be broadly defined to include supervisory control and data acquisition (SCADA) systems, distributed control systems (DCS), and systems that comprise programmable logic controllers (PLC) and Modbus protocols.

The connection between ICS or IIoT-based systems with public networks, however, increases their attack surfaces and risks of being targeted by cyber criminals. One high-profile example is the Stuxnet campaign, which reportedly targeted Iranian centrifuges for nuclear enrichment in 2010, causing severe damage to the equipment [1], [2]. Another example is that of the incident

targeting a pump that resulted in the failure of an Illinois water plant in 2011 [3].

BlackEnergy3 was another campaign that targeted Ukraine power grids in 2015, resulting in power outage that affected approximately 230,000 people [4]. In April 2018, there were also reports of successful cyber-attacks affecting three U.S. gas pipeline firms, and resulted in the shutdown of electronic customer communication systems for several days [1].

Although security solutions developed for information technology (IT) and operational technology (OT) systems are relatively mature, they may not be directly applicable to ICSs. For example, this could be the case due to the tight integration between the controlled physical environment and the cyber systems. Therefore, system-level security methods are necessary to analyze physical behaviour and maintain system operation availability [1].

ICS security goals are prioritized in the order of availability, integrity, and confidentiality, unlike most IT/OT systems (generally prioritized in the order of confidentiality, integrity, and availability) [5]. Due to close coupling between variables of the feedback control loop and physical processes, (successful) cyber-attacks on ICS can result in severe and potentially fatal consequences for the society and our environment.

This reinforces the importance of designing extremely robust safety and security measurements to detect and prevent intrusions targeting ICS [1]. Popular attack detection and attribution approaches include those based on signatures and anomalies. To mitigate the known limitations in both signature-based and anomaly-based detection and attribution approaches, there have been attempts to introduce hybrid-based approaches [6].

Although hybridbased approaches are effective at detecting unusual activates, they are not reliable due to frequent network upgrades, resulting in different Intrusion Detection System (IDS) typologies [7]. Beyond this, conventional attack detection and attribution techniques mainly rely on network metadata analysis (e.g. IP addresses, transmission ports, traffic duration, and packet intervals).

Therefore, there has been renewed interest in utilizing attack detection and attribution solutions based on Machine Learning (ML) or Deep Neural Networks (DNN) in recent times. In addition, attack detection approaches can be categorized into network-based or host-based approaches.

Supervised clustering, single-class or multi-class Support Vector Machine (SVM), fuzzy logic, Artificial Neural Network (ANN), and DNN are commonly used techniques for attack detection in network traffic. These techniques analyze real-time traffic data to detect malicious attacks in a timely manner. However, attack detection that considers only network and host data may fail to detect sophisticated attacks or insider attacks.

Unsupervised models that incorporate process/physical data can complement a system's monitoring since they do not rely on detailed knowledge of the cyber-threats. In general, a sophisticated attacker with sufficient knowledge and time, such as a nation state advanced persistent threat actor, can potentially circumvent robust security solutions. Furthermore, most of the existing approaches ignore the imbalanced property of ICS data by modeling only a system's normal behavior and reporting deviations from normal behavior as anomalies.

This is, perhaps, due to limited attack samples in existing datasets and real-world scenarios. Although using majority class samples is a good solution to avoid issues due to imbalanced datasets, the trained model will have no view of the attack samples' patterns. In other words, such an approach fails to detect unseen attacks and suffers from a high falsepositive rate [8].

LITERATURE SURVEY

[1] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, and J. Coble, "Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4362–4369, 2019.

The growing number of attacks against cyber-physical systems in recent years elevates the concern for cybersecurity of industrial control systems (ICSs). The current efforts of ICS cybersecurity are mainly based on firewalls, data diodes, and other methods of intrusion prevention, which may not be sufficient for growing cyber threats from motivated attackers.

[2] R. Ma, P. Cheng, Z. Zhang, W. Liu, Q. Wang, and Q. Wei, "Stealthy Attack Against Redundant Controller Architecture of Industrial CyberPhysical System," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9783–9793, 2019.

In an industrial Cyber-Physical System, the controller plays a critical role in guaranteeing reliability and stability. Therefore, redundant controller architecture is a well-adopted approach by Distributed Control Systems (DCS), Supervisory Control And Data Acquisition (SCADA) and other typical industrial Cyber-Physical Systems. They monitor and control the critical industrial process such as power generation, chemical industry, water treatment plant, etc.

[3] E. Nakashima, "Foreign hackers targeted U.S. water plant in apparent malicious cyber-attack, expert says." [Online]. Available: <https://www.washingtonpost.com/blogs/checkpointwashington/post/foreign-hackers-broke-into-illinois-water-plant-controlsystem-industry-expert-says/2011/11/18/gIQAgmTZYN blog.html>

Federal investigators are looking into a report that hackers managed to remotely shut down a utility's water pump in central Illinois last week, in what could be the first known foreign cyber attack on a U.S. industrial system.

The November 8 incident was described in a one-page report from the Illinois Statewide Terrorism and Intelligence Center, according to Joe

Weiss, a prominent expert on protecting infrastructure from cyber attacks.

[4] G. Falco, C. Caldera, and H. Shrobe, "IIoT Cybersecurity Risk Modeling for SCADA Systems," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4486–4495, 2018.

Urban critical infrastructure such as electric grids, water networks, and transportation systems are prime targets for cyberattacks. These systems are composed of connected devices which we call the Industrial Internet of Things (IIoT). An attack on urban critical infrastructure IIoT would cause considerable disruption to society. Supervisory control and data acquisition (SCADA) systems are typically used to control IIoT for urban critical infrastructure.

Despite the clear need to understand the cyber risk to urban critical infrastructure, there is no data-driven model for evaluating SCADA software risk for IIoT devices. In this paper, we compare non-SCADA and SCADA systems and establish, using cosine similarity tests, that SCADA as a software subclass holds unique risk attributes for IIoT. We then disprove the commonly accepted notion that the common vulnerability scoring system risk metrics of exploitability and impact are not correlated with attack for the SCADA subclass of software.

[5] J. Yang, C. Zhou, S. Yang, H. Xu, and B. Hu, "Anomaly Detection Based on Zone Partition for Security Protection of Industrial Cyber-Physical Systems," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4257–4267, 2018.

Industrial control systems (ICSs) are facing increasingly severe security threats. Zone isolation, a commonly adopted idea for stopping attack propagation in general information systems, has been investigated for ICS security protection. It is usually implemented through the perimeter security techniques. However, anomaly states of the physical processes in a compromised field zone may spread into other zones through the inter-zone information interaction. Due to the coupling of the physical processes between different zones, it is difficult to prevent the propagation of attack impact in ICSs.

[6] S. Ponomarev and T. Atkison, "Industrial control system network intrusion detection by telemetry analysis," *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 2, pp. 252–260, 2016.

Until recently, industrial control systems (ICSs) used "air-gap" security measures, where every node of the ICS network was isolated from other networks, including the Internet, by a physical disconnect. Attaching ICS networks to the Internet benefits companies and engineers who use them. However, as these systems were designed for use in the air-gapped security environment, protocols used by ICSs contain little to no security features and are vulnerable to various attacks.

[7] J. F. Clemente, "No cyber security for critical energy infrastructure," Ph.D. dissertation, Naval Postgraduate School, 2018.

The United States power grid is a logical target for a major cyber attack because it connects all of the nations critical infrastructures with electricity. Attackers consistently exploit vulnerabilities of the bulk power system and are close to being able to disrupt electrical distribution. We live in a world that is interconnected, from personal online banking to government infrastructure consequently, network security and defense are needed to safeguard the digital information and controls for these systems. The cyber attack topic has developed into a national interest because high-profile network breaches have introduced fear that computer network hacks and other security-related attacks have the potential to jeopardize the integrity of the nations critical infrastructure.

PROPOSED METHOD

Figure 1 shows the architecture of the proposed framework. In this framework, the attack detection method detects the attacks by analyzing the ICS input features using the combination of ensembled unsupervised DNNs and a decision tree. If an attack is detected, the sample is passed to several DNNs for detailed analysis.

If the attacks were previously unseen/unknown, the unseen attack detection module would detect it and label it as an unseen attack. This will be passed on for detailed security

analysis. Otherwise, the attack attribution method detects the attribute of the attack.

A. Proposed Ensemble Attack Detection Method

The proposed attack detection consists of two phases, namely representation learning and detection phase. Using a conventional unsupervised DNN on an imbalanced dataset yielded a DNN model that mainly learned majority class patterns and missed minority class characteristics.

Most researchers have tried to address this challenge by generating new samples or removing certain samples to make the dataset balanced and then passing the data to a DNN. However, in ICS/IIoT security applications, generating or removing samples are not reasonable solutions. Due to the ICS/IIoT systems' sensitivity, generated samples should be validated in a real network, which is impossible since the generated attack samples may be harmful to the network and cause severe impacts on the environment or human life.

In addition, validation of the generated samples is time-consuming. Moreover, removing the normal data from a dataset is not the right solution since the number of attack samples in ICS/IIoT datasets is usually less than 10% of the dataset, and most of the dataset knowledge is discarded by removing 80% of the dataset.

To avoid the above mentioned problems in handling imbalanced datasets, this study proposed a new deep representation learning method to make the DNN able to handle imbalanced datasets without changing, generating, or removing samples. This model consisted of two unsupervised stacked autoencoders, each responsible for finding patterns from one class. Since each model tries to extract abstract patterns of one class without considering another, the output of that model represented its inputs well.

The stacked autoencoders had three decoders and encoders with input and final representation layers. The encoder layers mapped the input representation to a higher, 800-dimensional space, a 400-dimensional space, and the final 16-dimensional space.

Equations 1 shows the encoder function of an autoencoder. The decoder layers did the opposite

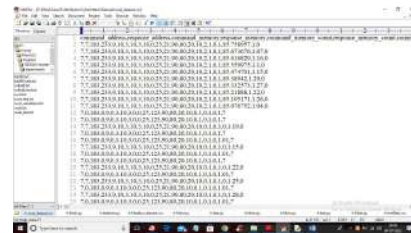
and tried to reconstruct the input representation by starting from the 16-dimensional new representation and mapping it to the 400-dimensional, 800-dimensional, and input representations. Equations 2 shows the decoder function of an autoencoder. These hyperparameters were selected using trial-and-error to have the best performance in f-measure with the lowest architectural complexity.

To implement this project we have designed following modules

- 1) Upload SWAT Water Dataset: using this module we will upload dataset to application and then read dataset and then find different attacks found in dataset
- 2) Preprocess Dataset: using this module we will replace all missing values with 0 and then apply MIN-MAX scaling algorithm to normalized features values and then split dataset into train and test where application used 80% dataset for training and 20% for testing
- 3) Run AutoEncoder Algorithm: using this module we will trained AutoEncoder deep learning algorithm and then extract features from that model.
- 4) Run Decision Tree with PCA: extracted features from AutoEncoder will get transform using PCA to reduce features size and then retrain with Decision tree. Decision tree will predict label for each record based on dataset signatures
- 5) Run DNN Algorithm: predicted decision tree label will further train with DNN (deep neural network) algorithm to detect and attribute attacks
- 6) Detection & Attribute Attack Type: using this module we will upload unknown or un-label TEST DATA and then DNN will predict attack type
- 7) Comparison Graph: using this module we will plot comparison graph between all algorithms
- 8) Comparison Table: using this module we will display comparison table of all algorithms which contains metrics like accuracy, precision, recall and FSCORE.

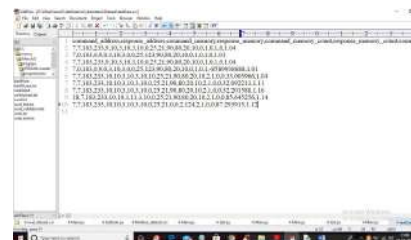
'Normal', 'Naive Malicious Response Injection (NMRI)', 'Complex Malicious', 'Response Injection (CMRI)', 'Malicious State Command Injection (MSCI)', 'Malicious Parameter Command Injection (MPCI)', 'Malicious Function Code Injection (MFCI)', 'Denial of Service (DoS)'

Above are the attacks found in dataset and dataset contains above labels as integer value of its index for example NORMAL label index will be 0 and continues up to 8 class labels. Below screen showing dataset details



In above dataset screen first row contains dataset column names and remaining rows contains dataset values and in last column we have attack type from label 0 to 7. We will use above dataset to train proposed Auto Encoder, decision tree and DNN algorithms.

In below screen we are using NEW test data which contains only signature and there is no class label and proposed algorithm will detect and attribute class labels.



In above test data we have IOT request signature without class labels.

In below screen you can read red colour comments to know about algorithms implementation

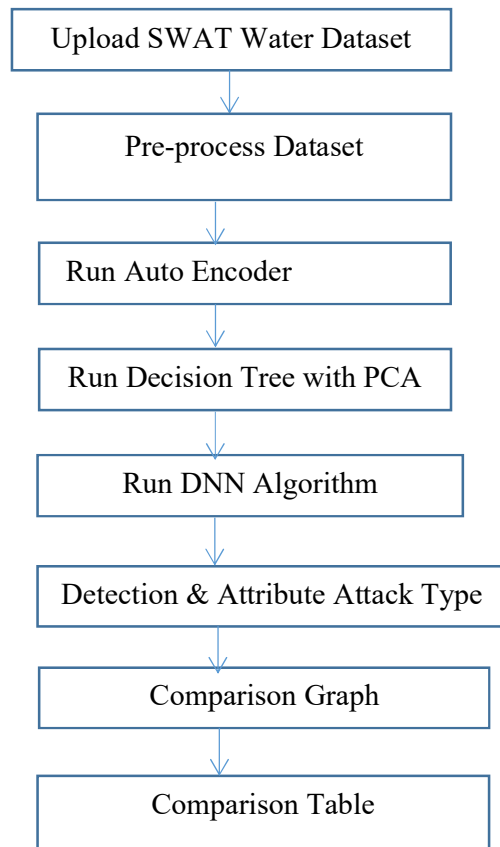
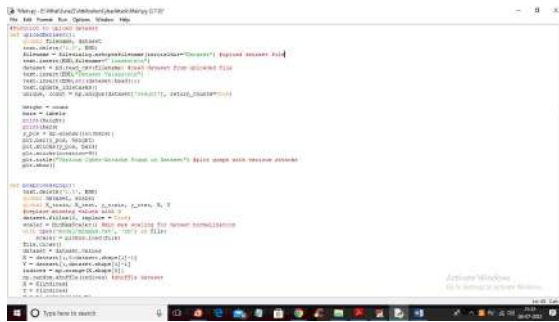


Fig. Flowchart

RESULT

To implement this project author has used SWAT (secure water treatment) and this dataset contains IOT request and response signature and associate each dataset with unique attack label and dataset contains below cyber-attack labels

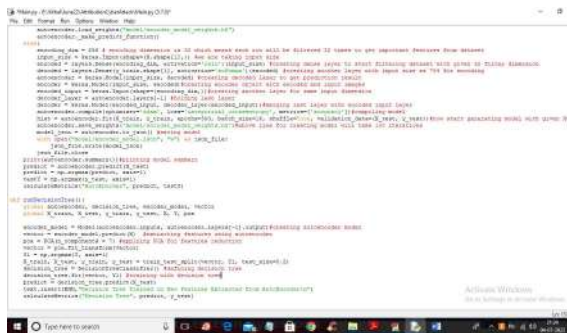


```

% Load the dataset
data = load('SWAT_Water_Dataset.mat');
% Extract features and labels
features = data.features;
labels = data.labels;
% Min-Max Normalization
[features_min, features_max] = minmax(features);
features = (features - features_min) ./ (features_max - features_min);
% Split the dataset into training and testing sets
[features_train, labels_train, features_test, labels_test] = trainTestSplit(features, labels, 0.8);

```

In above screen read red colour comments to know about dataset loading and min-max normalization



```

% Training the dataset with DNN algorithms
% Define the network architecture
net = feedforwardnet([100 10 1]);
% Initialize the network
net = init(net, defaults);
% Train the network
[net, history] = train(net, features_train, labels_train, 'train', 'patience', 10);
% Test the network
[net, history] = test(net, features_test, labels_test, 'test', 'patience', 10);

```

In above screen you can see we are using AutoEncoder, PCA and decision tree to train dataset and in below screen we are using DNN algorithms to train dataset with predicted labels from Decision Tree.



```

% Training the dataset with DNN algorithms
% Define the network architecture
net = feedforwardnet([100 10 1]);
% Initialize the network
net = init(net, defaults);
% Train the network
[net, history] = train(net, features_train, labels_train, 'train', 'patience', 10);
% Test the network
[net, history] = test(net, features_test, labels_test, 'test', 'patience', 10);

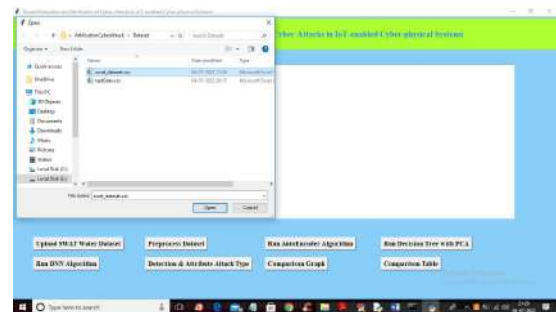
```

In above screen we are training dataset with DNN algorithms

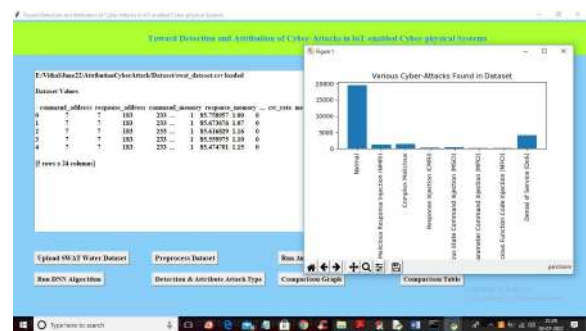
To run project double click on 'run.bat' file to get below screen



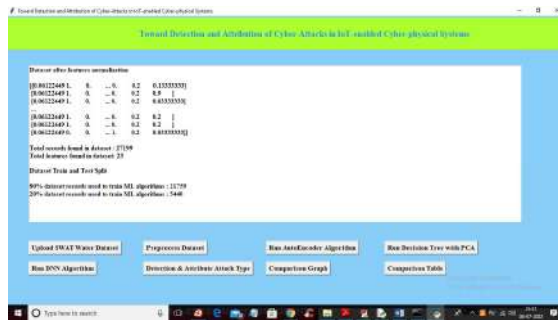
In above screen click on 'Upload SWAT Water Dataset' button to upload dataset to application and get below output



In above screen selecting and uploading SWAT dataset file and then click on 'Open' button to load dataset and get below output



In above screen dataset loaded and in graph x-axis contains ATTACK NAME and y-axis contains count of those attacks found in dataset and we can see 'NORMAL' class contains so many records and other attacks contains very few records so it will raise data imbalance problem which can be solved using AutoEncoder, Decision Tree and DNN. Now close above graph and then click on 'Preprocess Dataset' button to remove missing values and then normalized values with MIN-MAX algorithm



In above screen all values are normalized (converting data between 0 and 1 called as normalization) and then we can see total records in dataset and then dataset train and test split records count also displaying. Now dataset is ready and now click on 'Run AutoEncoder Algorithm' button to train dataset with AutoEncoder and get below accuracy



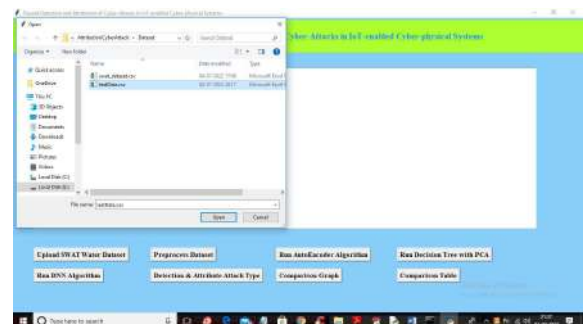
In above screen with AutoEncoder we got 90% accuracy and this accuracy can be enhance by implementing Decision Tree with PCA algorithm and now click on 'Run Decision Tree with PCA' button to get below output



In above screen we can see with decision tree accuracy and precision value is enhanced and now click on 'Run DNN Algorithm' button to further enhance accuracy and get below output



In above screen with DNN we got 99% accuracy and now click on 'Detection & Attribute Attack Type' button to upload test DATA and detect attack attributes



In above screen selecting and uploading 'TEST DATA' file and then click on 'Open' button to get below output



In above screen in square bracket we can see TEST data values and after arrow => symbol we can see detected ATTACK TYPE and scroll down above text area to view all detection

In above table we can see algorithm names and its metrics values such as accuracy and precision and other.

This paper proposed a novel two-stage ensemble deep learning-based attack detection and attack attribution framework for imbalanced ICS data. The attack detection stage uses deep representation

learning to map the samples to the new higher dimensional space and applies a DT to detect the attack samples. This stage is robust to imbalanced dataset and capable of detecting previously unseen attacks. The attack attribution stage is an ensemble of several one-vs-all classifiers, each trained on a specific attack attribute. The entire model forms a complex DNN with a partially connected and fully connected component that can accurately attribute cyberattacks, as demonstrated. Despite the complex architecture of the proposed framework, the computational complexity of the training and testing phases are respectively $O(n^4)$ and $O(n^2)$, (n is the number of training samples), which are similar to those of other DNN-based techniques in the literature. Moreover, the proposed framework can detect and attribute the samples timely with a better recall and f-measure than previous works. Future extension includes the design of a cyber-threat hunting component to facilitate the identification of anomalies invisible to the detection component for example by building a normal profile over the entire system and the assets.

[1] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, and J. Coble, "Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4362–4369, 2019.

[2] R. Ma, P. Cheng, Z. Zhang, W. Liu, Q. Wang, and Q. Wei, "Stealthy Attack Against Redundant Controller Architecture of Industrial CyberPhysical System," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9783–9793, 2019.

[3] E. Nakashima, "Foreign hackers targeted U.S. water plant in apparent malicious cyber attack, expert says." [Online]. Available: https://www.washingtonpost.com/blogs/checkpointwashington/post/foreign-hackers-broke-into-illinois-water-plant-controlsystem-industry-expert-says/2011/11/18/gIQAgnTZYN_blog.html

[4] G. Falco, C. Caldera, and H. Shrobe, “IIoT Cybersecurity Risk Modeling for SCADA Systems,” *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4486–4495, 2018.

[5] J. Yang, C. Zhou, S. Yang, H. Xu, and B. Hu, "Anomaly Detection Based on Zone Partition for Security Protection of Industrial Cyber-Physical

- Systems,” IEEE Transactions on Industrial Electronics, vol. 65, no. 5, pp. 4257–4267, 2018.
- [6] S. Ponomarev and T. Atkison, “Industrial control system network intrusion detection by telemetry analysis,” IEEE Transactions on Dependable and Secure Computing, vol. 13, no. 2, pp. 252–260, 2016.
- [7] J. F. Clemente, “No cyber security for critical energy infrastructure,” Ph.D. dissertation, Naval Postgraduate School, 2018.
- [8] C. Bellinger, S. Sharma, and N. Japkowicz, “One-class versus binary classification: Which and when?” in 2012 11th International Conference on Machine Learning and Applications, vol. 2, 2012, pp. 102–106.
- [9] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>
- [10] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.
- [11] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, “Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things,” IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6822–6834, 2019.
- [12] I. A. Khan, D. Pi, Z. U. Khan, Y. Hussain, and A. Nawaz, “HML-IDS: A hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems,” IEEE Access, vol. 7, pp. 89 507–89 521, 2019.
- [13] T. K. Das, S. Adepur, and J. Zhou, “Anomaly detection in industrial control systems using logical analysis of data,” Computers & Security, vol. 96, p. 101935, 2020.
- [14] J. J. Q. Yu, Y. Hou, and V. O. K. Li, “Online False Data Injection Attack Detection With Wavelet Transform and Deep Neural Networks,” IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3271–3280, 2018.
- [15] M. M. N. Aboelwafa, K. G. Seddik, M. H. Eldefrawy, Y. Gadallah, and M. Gidlund, “A machine-learning-based technique for false data injection attacks detection in industrial iot,” IEEE Internet of Things Journal, vol. 7, no. 9, pp. 8462–8471, 2020.
- [16] W. Yan, L. K. Mestha, and M. Abbaszadeh, “Attack detection for securing cyber physical systems,” IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8471–8481, 2019.
- [17] A. Cook, A. Nicholson, H. Janicke, L. Maglaras, and R. Smith, “Attribution of Cyber Attacks on Industrial Control Systems,” EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, vol. 3, no. 7, p. 151158, 2016.
- [18] L. Maglaras, M. Ferrag, A. Derhab, M. Mukherjee, H. Janicke, and S. Rallis, “Threats, Countermeasures and Attribution of Cyber Attacks on Critical Infrastructures,” ICST Transactions on Security and Safety, vol. 5, no. 16, p. 155856, 2018.
- [19] M. Alaeiyan, A. Dehghantanha, T. Dargahi, M. Conti, and S. Parsa, “A Multilabel Fuzzy Relevance Clustering System for Malware Attack Attribution in the Edge Layer of Cyber-Physical Networks,” ACM Transactions on Cyber-Physical Systems, vol. 4, no. 3, pp. 1–22, 2020.
- [20] U. Noor, Z. Anwar, T. Amjad, and K.-K. R. Choo, “A machine learning-based FinTech cyber threat attribution framework using highlevel indicators of compromise,” Future Generation Computer Systems, vol. 96, pp. 227–242, 2019.