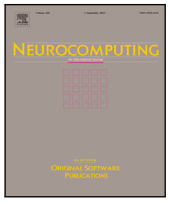
Neurocomputing 568 (2024) 127068

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



A hybrid methodology for anomaly detection in Cyber–Physical Systems

Nicholas Jeffrey [a,b,*], Qing Tan [a], José R. Villar [b]

[a] Faculty of Science and Technology, Athabasca University, Canada
[b] Computer Science Department, University of Oviedo, Spain

ARTICLE INFO

Keywords:
Security threats
Cyber–Physical Systems
Machine learning

ABSTRACT

The rapid adoption of Industry 4.0 has seen Information Technology (IT) networks increasingly merged with Operational Technology (OT) networks, which have traditionally been isolated on air-gapped and fully trusted networks. This increased attack surface has resulted in compromises of Cyber–Physical Systems (CPS) with significant economic and life safety consequences. This paper proposes a hybrid model of anomaly detection of security threats to CPS by blending the signature-based and threshold-based Intrusion Detection Systems (IDS) commonly used in IT networks, with a Machine Learning (ML) model designed to detect behaviour-based anomalies in OT networks. This hybrid model achieves more rapid detection of known threats through signature-based and threshold-based detection strategies, and more accurate detection of unknown threats via behaviour-based anomaly detection using ML algorithms.

1. Introduction

Cyber–Physical Systems (CPS) are integrated systems that combine software and physical components [1]. The adoption of Industry 4.0 has driven the rapid growth of CPS, outpacing advancements in cybersecurity, with new threat models and security challenges that lack a unified framework for secure design, malware resistance, and risk mitigations [2].

Threat detection and prevention is a mature industry in Information Technology (IT) networks, but less so in Operational Technology (OT) networks, largely because traditional Industrial Control Systems (ICS), have not adjusted to the ubiquitous connectivity of Industry 4.0, and still largely consider security to be an afterthought [3].

A modern CPS will be comprised not just of OT components, such as sensors and actuators, but will include IT components with ubiquitous network interconnectivity, which significantly changes the threat profiles historically faced by the operators of OT networks.

As IT networks and OT networks merged to form CPS, it quickly became apparent that security policies were inconsistent for different parts of the CPS. Traditional IT networks have used the so-called CIA (Confidentiality, Integrity, Availability) triad to define the organizational security posture, with each facet listed in order of importance. OT networks reverse that order, with availability being the most important factor, followed by integrity, with confidentiality the least important facet of overall system security

IT networks typically deploy Host-based Intrusion Detection Systems (HIDS), such as signature-based antimalware agents on endpoint devices. However, the endpoint devices on OT networks are typically much more resource-constrained, are unable to run a local HIDS, and will sometimes opt for a passive Network-based Intrusion Detection System (NIDS), or more commonly, no IDS at all, due to historical design assumptions of operating on an isolated and fully trusted network.

Threat detection methodologies can be broadly categorized as signature-based, threshold-based, or behaviour-based. Each of these methodologies have their own strengths and weaknesses, which proves challenging in environments that contain both software and hardware components.

A modern IDS for a CPS must recognize this reality, and provide anomaly detection for both the Cyber and the Physical portions of the CPS. This paper is an extension of previous works [4], and further develops the concepts of a hybrid model of anomaly detection, that combines the use of Machine Learning (ML) with signature-based, threshold-based, and behaviour-based methodologies to best balance the competing goals of low latency, high accuracy, and rapid detection of threats.

This paper proposes a complementary hybridized security stage at the network edge, where Machine Learning (ML) methods face only some part of the potential attacks, relying on IT Security for detecting a significant part of the possible anomalies. In this way, only normal messages are analysed, reducing the computational overload on the edge. The remainder of this study is organized as follows: Section 2 focuses on the state of the art in anomaly detection in CPS. Following that, the approach presented in this study is detailed in Section 3, while

* Corresponding author at: Computer Science Department, University of Oviedo, Spain.
E-mail addresses: uo292630@uniovi.es (N. Jeffrey), qingt@athabascau.ca (Q. Tan), villarjose@uniovi.es (J.R. Villar).

https://doi.org/10.1016/j.neucom.2023.127068
Received 9 February 2023; Received in revised form 30 October 2023; Accepted 19 November 2023
Available online 23 November 2023

Table 1
Summary of related works.

Author	Anomaly detection method
Kaur et al	Feature weighting of imbalanced data sets to avoid overfitting.
Vuttipittayamongkol	How class overlap affects classification accuracy in imbalanced data sets.
Esposito et al	Mitigating overfitting in imbalanced data sets with subset stratification.
Ahmin et al	Use of one-class classifiers for detection of zero-day threats not present in training data.
Abid et al	Distributed intrusion detection on industrial control systems using big data techniques.
Al-Asiri et al	Modelling threshold-based detection as a finite state machine.
Altaha et al	Improving anomaly detection by combining a rules-based ML algorithm with DPI.
Neshenko et al	Using GAN to improve classification accuracy of multivariate data sets.
Siniosoglou et al	Improving classification accuracy by encapsulating Deep Learning into a GAN.
Yilmaz et al	Improving anomaly detection in resource-constrained environments with Transfer Learning.
As-Shabi et al	Improving classification accuracy of imbalanced data sets with LTSM.
Raman et al	Improving anomaly detection in CPS with PNN.
Greggio	Reducing computational complexity in IDS with GMM.

Section 4 describes the experimental methodology. Section 5 describes the experiment and results. Finally, conclusions and future works are detailed in Section 6.

2. Related work

A brief summary of related works is shown in Table 1, with additional detail shown below.

Kaur et al. [5] explore the ML challenges related to imbalanced data sets, comparing different data pre-processing and feature weighting strategies to avoid overfitting and ML classification accuracy. Imbalanced classes are a common challenge in anomaly detection, particularly in CPS when the minority class of attack data is often suppressed by the CPS operator due to financial and life safety concerns, making it difficult to obtain accurate source data for ML models.

Vuttipittayamongkol et al. [6] further develop concepts of class imbalance in ML, focusing on how class overlap can affect classification accuracy, which is particularly concerning to the operators of CPS, where misclassification of minority class instances may lead to interruptions of critical infrastructure availability. Class overlap occurs when outlier data shifts the decision boundary towards the majority class, leading to ML models skewing towards increased false negative misclassifications. Different ML models have varying levels of tolerance for class overlap, with SVM experiencing significantly more degradation than KNN.

Esposito et al. [7] continue to develop mitigations for ML classifiers overpredicting on the majority class in imbalanced data sets. An automated procedure called GHOST is proposed as a generalized method that can be applied to any ML method, by stratifying subsets of the training data to find the optimal decision threshold without needing to retrain the model.

Ahmin et al. [8] propose a new taxonomy of classifying threats based on protocol analysis, traffic analysis, and control process analysis. This has some overlap with the hybrid anomaly detection strategy proposed in this paper, by offloading the detection of known threats to a signature-based IDS, and accomplishes much with low computational requirements, but the AI model suffers from relatively low detection rates, which is less than optional for OT networks due to potential economic and life safety issues.

Ahmed et al. [9] investigate the use of one-class classifiers for anomaly detection in an CPS, noting that the use of unsupervised learning can be useful in detection of zero-day or previously unknown threats, but result in an unacceptably high false positive rate, which potentially affects the availability of the CPS.

Abid et al. [10] propose that traditional signature-based IDS (Snort, Bro, etc.) are only effective at detecting known threats, and a cloud-based distributed IDS can leverage AI/ML by utilizing powerful cloud-based compute resources that are faster than locally available compute capacity inside the corporate network. This strategy does provide rapid AI model training, but still uses AI/ML for all threat detection, rather than offloading simple threats to a threshold-based detection method.

Al-Asiri and El-Alfy [11] take the concept of threshold-based detection and describe how the entire CPS can be modelled as a finite state machine, codifying a set of rules that interrogate the physical measurements from the CPS and compare those readings against an expert system of rules, with readings outside of the defined thresholds in the expert system defined as an anomaly. This can be considered as an extreme case of a one-class classifier, by defining a comprehensive ruleset of all permissible states of the CPS, with anything outside those states defined as a anomalous.

Altaha and Hong [12] further develop the concept of a rules-based ML algorithm, but focus on the predictability of specific traffic patterns between nodes in the CPS, coupled with Deep Packet Inspection (DPI) to obtain a deeper understanding protocol-specific commands passed between the sensors and actuators in the CPS. An unsupervised DL model is proposed to capture and define normal behaviour, but this approach is protocol-specific, making it difficult to generalize across the various legacy communication protocols employed in CPS.

Neshenko et al. [13] propose an unsupervised ML model for CPS running critical infrastructure with life safety concerns, using the CPS sensor readings and actuator states to generate a one-class model of normal behaviour for the CPS, classifying any activity outside of the defined class as anomalous. Neshenko et al. further postulate that the use of a Generative Adversarial Network (GAN) model to rapidly learn the normal behaviour of a CPS provides improved efficiency and accuracy, particularly for complex environments with multivariate data sets.

Siniosoglou et al. [14] further refine the use of GAN models to develop an IDS for anomaly detection, by encapsulating an Autoencoder Deep Neural Network (DNN) into the structure of a GAN. By combining these two DNN, the hybrid Autoencoder-GAN can be used for both anomaly detection and attack classification, providing richer insight into threats to the CPS.

Yilmaz et al. [15] propose the use of Transfer Learning to improve anomaly detection on resource-constrained OT components of a CPS, by generating intrusion detection algorithms for novel attack types based on knowledge previously learned in a related problem domain. The extreme heterogeneity of CPS and rapid evolution of IoT protocols makes the TL methodology attractive for shortening the time between novel malicious traffic and first detection.

As-Shabi and Abuhamdah [16] propose the use of Deep Learning for anomaly detection in IoT environments through the use of long short term memory (LSTM) algorithm to achieve very high accuracy rates for true positive detections of known threats, but suffers due to a lack of training data showing anomalous activity.

Raman et al. [17] propose an anomaly detector based on a Probabilistic Neural Network (PNN) using the popular SWaT data set for validation. PNN is an interesting variant of neural networks that is particularly effective in anomaly detection on highly imbalanced data sets, which is typical for CPS environments. Recognizing that threshold-based anomaly detection techniques require a detailed understanding of the physical topology and process flow of a CPS, the proposed

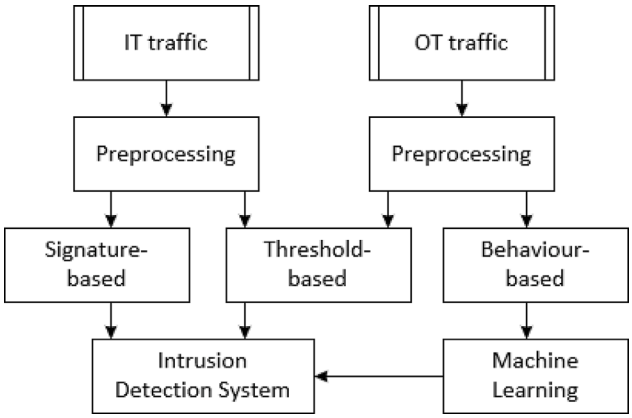


Fig. 1. Logical workflow of data. The IDS for the IT networks follows the standard guidelines, using either signature-based or threshold-based anomalies sent to the IDS. In the OT network, threshold-based anomalies based on immutable physical characteristics of the CPS are sent directly to the IDS, while behavioural anomalies are detected with an ML model, and then forwarded to the IDS.

PNN model can achieve higher classification accuracy than other algorithms, by leveraging Bayesian strategies for mapping input variables to class labels, avoiding the requirement for detailed understanding of the CPS process flow commonly used for threshold-based detection methodologies.

Greggio [18] further develops the concepts of probabilistic anomaly detection with an unsupervised learning algorithm that leverages finite Gaussian Mixture Models (GMM) to provide a compromise between classification accuracy and computational complexity. This method is particularly promising for rapid classification of complex data sets with high dimensionality, but classification accuracy suffers if the data distribution between the classes changes over time.

3. Machine learning based anomaly detection in OT networks

CPS have unique challenges with using ML for anomaly detection, as there is frequently a large amount of available data that shows normal activity, but a lack of real-world data showing anomalous or malicious activity [19]. The lack of training data for anomalous activity has frequently been addressed by the use of artificially generated research data sets, which have varying degrees of fidelity to real-world environments.

This lack of available training data containing anomalous activity can be addressed by the use of one-class classifiers [20], which are used to recognize all benign activity within a single class, and therefore drawing the conclusion that any detected activity outside of that class is anomalous by definition. This is an attractive ML solution, as there is typically a large amount of training data for benign activity, which allows the learning model to be trained in a rapid and inexpensive manner.

This research assumes that the IT network IDS filters all malicious traffic coming from the Internet or even from potentially hostile computers within the organization. Therefore, the CPS security on the OT networks must focus solely on unexpected traffic that does not resemble a threat from the IT network perspective. In this sense, this is a hybridization of the security models, where the IT security services and the OT security services cooperate through a divide-and-conquer approach (see Fig. 1). This scheme helps in reducing the computational effort that the hardware at the edge must bear in detecting cyber-attacks.

There are certain anomaly detection classes that this paper will intentionally avoid using ML for detection. The IT portions of the CPS are assumed to be already protected by a Host-based Intrusion Detection System (HIDS), such as an antimalware agent, which employs signature-based detection techniques for known malicious activity, and threshold-based detection techniques for operational parameters of the CPS running outside of the defined tolerance limits. This paper focuses on the OT portions of the CPS, particularly behaviours that can be modelled with ML.

Certain OT components of the CPS also make use of threshold-based anomaly detection techniques, especially with regard to the physical components of the CPS, which will have defined operational tolerances for physical characteristics such as operating temperature, pressure, vibration, frequency of actuator duty cycles, etc. The design tolerances of these physical characteristics are typically based on life safety regulations that cannot be modified by the operators of the CPS. Since these characteristics can be considered immutable, they can be rapidly and accurately validated by a simple threshold-based anomaly detection strategy, avoiding the time-consuming and costly training of an ML model for this type of threat.

To contrast, the endpoint devices on OT networks are much more varied and heterogeneous than on IT networks, which makes accurate anomaly detection more difficult. To describe that another way, IT networks are much closer to a monoculture, with large numbers of identical systems (i.e. Windows, Linux, etc.), while OT networks are usually very different from organization to organization. This is where we will use ML to train a model to understand what “normal” behaviour looks like for the OT portion of the CPS, and alert on anomalous activity that falls outside the training model.

Moreover, in this study, we assert that the problem of anomaly detection in CPS must be modelled as one-class problems [21], that is, problems where the available data are almost completely gathered from normal operation, implying that there are almost no instances from the anomaly class. Learning two-class models with this highly unbalanced data would lead to generating a balanced data set by replicating a short amount of instances, thereby inducing a bias on the models. Additionally, it may lead to learning models with poor generalization capabilities, rendering them incapable of facing cyber-attacks that were not present in the training data.

However, modelling anomaly detection as a one-class problem leads to learning a model for normal traffic, which would be easily adaptable to changes in the operational procedures of the CPS. In this research, we propose the use of one-class K-Nearest Neighbours (KNN) and one-class Support Vector Machines (SVM) as candidates for modelling the normal traffic because they can easily be deployed on computing devices on the edge. Nevertheless, other more complex modelling techniques, such as one-class Deep Learning models could be used as well for this task. In this latter case, not only data transformations can be utilized, but the network traffic itself can be used directly as the inputs of the learning model.

To merge the two methods, the OT threshold-based and the ML-based IDS, we propose using the any-vote ensemble method, where an alarm due to any of the CPS IDS methods is directly signalled to the system operator in order to follow the corresponding security response procedure. Therefore, for the sake of simplicity, in this research we focus only on the ML part in order to evaluate whether the one-class approach performs better than the two-class solution or not.

4. Methodology

Threat detection methodologies can be broadly categorized as signature-based, threshold-based, or behaviour-based. Traditional antivirus programs are an example of a signature-based threat detection methodology, using a centralized and regularly updated database of signatures of malicious files or traffic to trip an alarm on an IDS and/or IPS. Signature-based detection works well on IT networks thanks to standardized communication protocols and low levels of heterogeneity but suffers from high levels of false negatives on OT networks due to their proprietary communication protocols and heterogeneous physical components.

Threshold-based methodologies rely on known ranges of acceptable operation, which are relatively easy to define on IT networks. Examples of threshold-based threat detections for IT networks include network link utilization, communication latency, processor utilization levels, etc. However, OT networks have proven more difficult to accurately define known ranges of acceptable operation, due to real-world environmental fluctuations. For example, a wireless mesh network of air quality sensors in a smart city environment may have communication latency impacted by fog or rain, making the thresholds of acceptable operation differ based on unpredictable weather conditions.

Behaviour-based methodologies are the most difficult to accurately define on IT networks and are even more challenging for OT networks. Defining an accurate baseline of normal behaviour on an IT network requires a thorough understanding of what normal system activity looks like, and it is rare that IT networks are completely unchanged over their entire lifecycle, making any definition of normal behaviour a moving target at best. These challenges are exacerbated on OT networks, which tend to be even more dynamic due to environmental factors such as weather-related variations in temperature, humidity, ambient light, etc. Additionally, the negative impact of a false positive or false negative detection on an OT network has more significant consequences, including physical equipment damage and life safety concerns.

This study seeks to maximize the strengths and minimize the weaknesses of each of the above threat detection methodologies through a hybrid model that is described in more detail in the following sections.

5. Experiment and results

5.1. Description of the data sets

An assortment of research data sets in the field of anomaly detection for OT networks already exist (SWaT [22], WADI [23], CSE-CIC-IDS2018 [24]), but focus heavily on IT networks rather than OT behavioural patterns. None of the existing data sets are optimized for one-class classifier models, and none consider the use of a hybrid anomaly detection strategy that combines signature-based and threshold-based anomaly detection strategies with an AI learning model.

For these reasons, a new data set was generated programmatically using only benign data, and labelled in a semi-automated manner for loading into the learning model. In this research, we take advantage of an available CPS installation depicted in Fig. 2. A logical view of the data collection workflow in the testbed has already been shown in Fig. 1.

This testbed is based on a scaled-down pilot system for a commercial greenhouse facility, and each sensor and actuator is accessible via either ethernet or serial connectivity, which allows for observability of all components of the CPS. Sensory systems include temperature and soil moisture sensors and transmitters. Telemetry information from each of component in the CPS is forwarded to a centralized log collector for analysis and anomaly detection. Equipment in the testbed can be broadly classified as resource-rich (personal computers, ethernet switches, firewalls, etc.), or resource-constrained (sensors, actuators, microcontrollers, etc.). Single-board computer (SBC) devices such as a Raspberry Pi and ESP32-based microcontrollers are also used for emulation of simple sensors and actuators.

These distinctions between resource-rich and resource-constrained are primarily useful for determining the appropriate method of anomaly detection. Resource-rich devices are typically able to run a resource-intensive agent to collect health metrics and operational characteristics, while the resource-constrained components such as sensors and actuators are often unable to be queried directly, and must rely on passive observation of network activity to determine their health state.

The data set was generated through direct observation of a prototype/testbed environment using a combination of real and emulated processes to generate activity on the sensors and actuators. A small degree of randomness was introduced to the observed data to account for the non-deterministic speed of wifi and ethernet-based communication, as well as environmental fluctuations such as sunlight for photovoltaic power generation, and ambient humidity and temperature affecting the OT components of the CPS. This provided an acceptable range of behaviour considered normal by the learning model, without being overly discriminatory for expected environmental fluctuations. The generated data set includes 2000 instances of normal traffic and 200 instances of potential cyber-attacks. This unbalanced nature of the data set reflects the actual scenario, where cyber-attacks are intrinsically rare and, thus, scarce data is available; conversely, normal traffic is more common by far, although we have included a 10 to 1 ratio only.

The raw data was then transformed programmatically using min–max normalization to scale all the data to values between 0 and 1, in order to avoid features in the data set with larger numerical values from unduly influencing the learning model.

5.2. Experiment setup

Given the unbalanced nature of the data, anomaly detection must be modelled as a one-class problem, where the model learns from data corresponding to the normal class and, when deployed, identifies those instances that do not belong to this class. To our knowledge, this is not the common method used in the literature, where the problem is modelled as a two-classes problem. Therefore, a completed experimentation is designed to evaluate whether one-class modelling performs better or not than two-classes modelling for the anomaly detection of network traffic in OT networks. As long as KNN and SVM are proposed as the one-class models, the same techniques are proposed for the comparison; therefore, two-classes KNN and SVM will be trained and evaluated to obtain a fair comparison of the results.

The procedure is shown in Fig. 3. Both the normal traffic and the anomaly traffic data are downsampled to, on the one hand, keep some anomaly traffic for the final validation stage; on the second hand, the downsampling allows us to obtain a balanced data set for training the two-classes problems. The remaining data not used in the training and testing of the models is preserved for the final validation stage.

Once the data is downsampled, it is normalized and a 10-fold cross-validation is carried out independently for the two-classes problem and for the one-class problem, although the same random seed is used to obtain the same partitions in each case for comparison reasons. In the two-classes problem, all the partitions may include normal and anomaly instances. However, in the one-class problem, the partitions are prepared only with the normal traffic instances; the anomaly instances are used to measure the performance of the models obtained for each fold. Interestingly, the normalization for the one-class problem is determined exclusively with data from normal traffic only.

A final validation stage includes all the data that has not been used in training and testing; this is an unbalanced data set containing instances from normal traffic and from anomalies. The aim of this validation stage is to compare the behaviour of the different modelling techniques included in this comparison, so conclusions could be extracted.

In order to avoid drawing conclusions from biased data, the whole procedure is repeated 10 times from the random downsampling to the final validation stage; the obtained partial results will be aggregated.

To measure the quality of the models the Accuracy, Sensitivity, Specificity, and the Geometric Mean measurements [25] will be used. The Accuracy will give some ideas of the performance on the balanced data set, while Sensitivity and Specificity will help in the final validation stage, where the data will be clearly unbalanced.

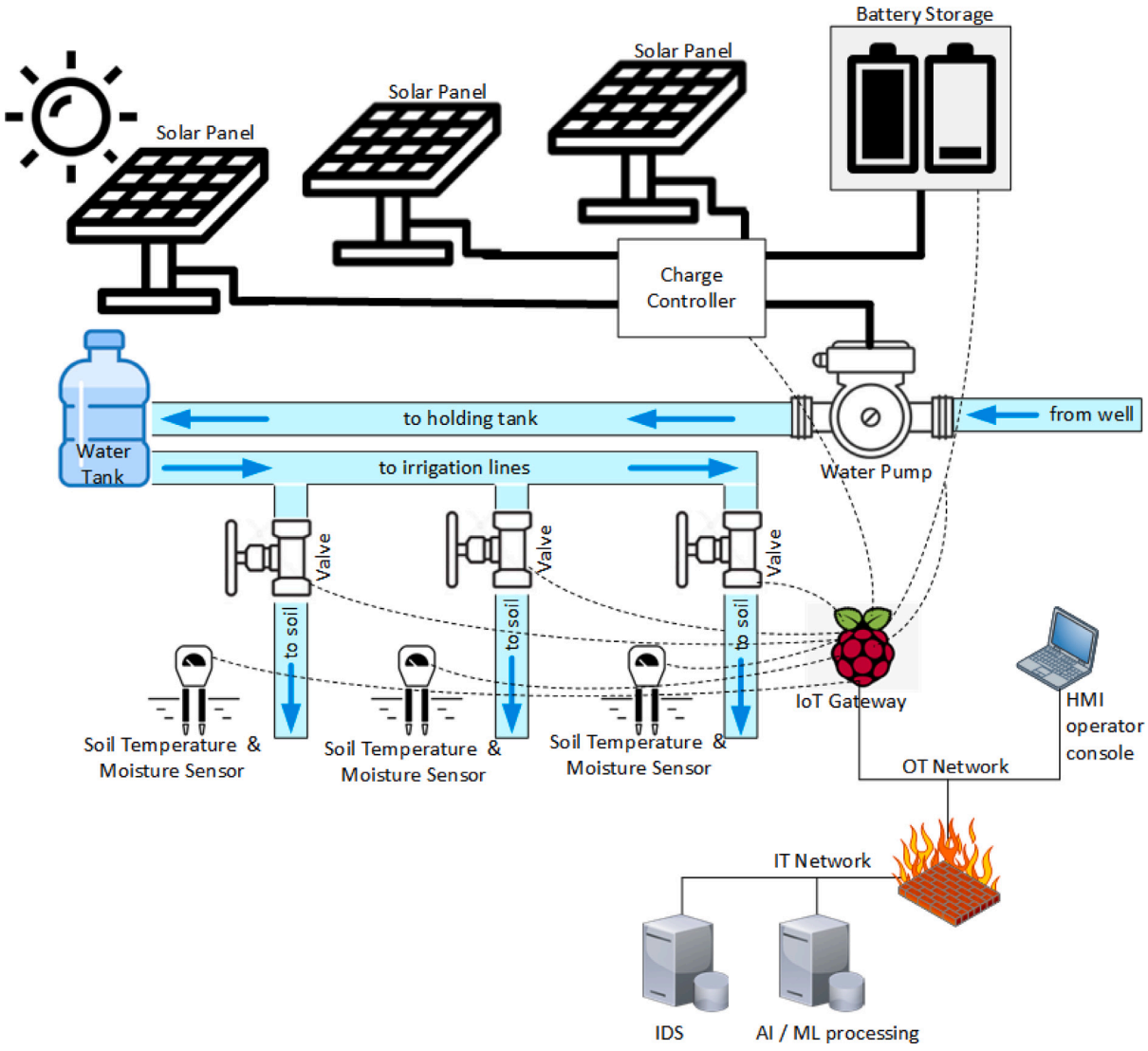


Fig. 2. Schema of the CPS installation used as a prototype for this research.

5.3. Results and discussion

Results from the experimentation are shown in Table 2 for the cross-validation training and Table 3 for the final validation stage.

To evaluate the experimental results, the accuracy, sensitivity, and specificity of each model was calculated and compared.

The 1-class SVM model had the lowest accuracy, with 68%. This model suffered from high false positives (37%) and high false negatives (32%) due to overfitting from assumptions the model makes about linearity when the class ratios are unbalanced.

The 2-class SVM model fared better, with 77% accuracy. The true positive detection was very good (100%), but this model suffers from high false negatives (18%), which are particularly undesirable for operators of CPS, as failing to detect actual attack activity can have significant financial and life safety consequences.

The 2-class KNN model had slightly better accuracy than 2-class SVM, with 82%. This model also suffers from high false negatives (18%) due to outliers affecting the decision threshold, although to a lesser extent than 2-class SVM.

The 1-class KNN model provided the best accuracy (98%), thanks to its higher tolerance of distribution shift in the ratio of the data set classes. While accuracy is high, this model does predict excessive false positives due to overlap in the data points in the unbalanced classes, as well as scarcity of data points in the anomaly class. While false positives are undesirable, the operators of CPS are much more concerned with false negatives (i.e. missing a legitimate cyberattack), and this model does provide the lowest (1%) false negative rate. This supports the original hypothesis that because attack data for CPS is intrinsically rare, anomaly detection for Cyber–Physical Systems should be approached as a 1-class problem.

6. Conclusions and future works

Threat detection in CPS faces unique challenges, with the differing security postures of IT and OT networks making it difficult to provide a unified threat detection strategy. This paper proposes a hybrid methodology that leverages signature-based detection strategies for known attack patterns, threshold-based detection strategies for immutable properties of the CPS, and the use of one-class ML algorithms for behaviour-based detection of anomalies.

This paper details experimentation with 1-class and 2-class ML algorithms against an unbalanced data set, starting from the hypothesis that because malicious activity in CPS was a very small minority of the available data set, greater accuracy can be obtained by using 1-class classifiers to recognize the majority class of benign activity, with any detected activity outside of the majority class as anomalous. The

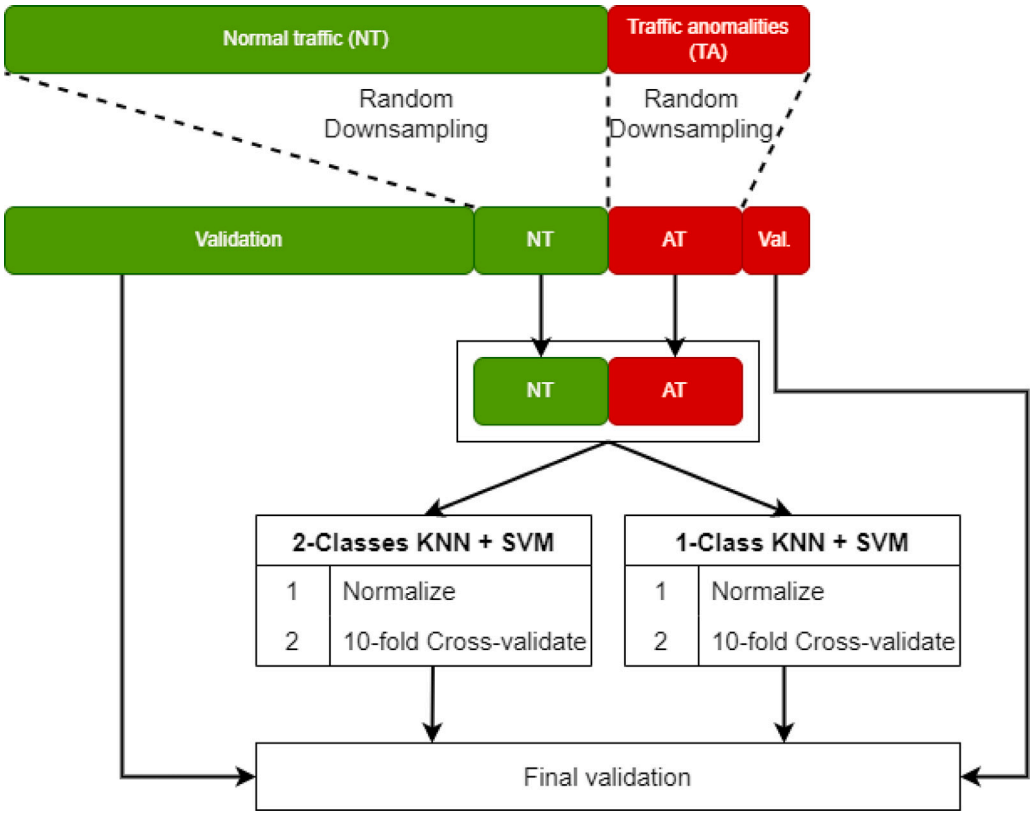


Fig. 3. Experimental setup followed in this research. The normal traffic (green) and the anomaly traffic (red) data are randomly downsampled to balance the train/test data set. This balanced data set is used for training and testing the collection of models. The remaining data is used in the final validation.

Table 2
Cross-validation results for the 1-Class and 2-Class KNN and SVM models.

Fold	1C-KNN				1C-SVM			
	Acc	Sens	Spec	GM	Acc	Sens	Spec	GM
1	0.9432	0.3750	1.0000	0.6124	0.7500	0.6875	0.8125	0.7474
2	0.9602	0.5625	1.0000	0.7500	0.6875	0.6875	0.6875	0.6875
3	0.9489	0.4375	1.0000	0.6614	0.7500	0.7500	0.7500	0.7500
4	0.9545	0.5000	1.0000	0.7071	0.7188	0.8125	0.6250	0.7126
5	0.9545	0.5000	1.0000	0.7071	0.6250	0.5625	0.6875	0.6219
6	0.9659	0.6250	1.0000	0.7906	0.6562	0.6250	0.6875	0.6555
7	0.9716	0.6875	1.0000	0.8292	0.5938	0.5000	0.6875	0.5863
8	0.9545	0.5000	1.0000	0.7071	0.7500	0.8125	0.6875	0.7474
9	0.9716	0.6875	1.0000	0.8292	0.5938	0.6250	0.5625	0.5929
10	0.9375	0.3125	1.0000	0.5590	0.7188	0.8125	0.6250	0.7126
Mean	0.9562	0.5188	1.0000	0.7153	0.6844	0.6875	0.6813	0.6814
Median	0.9545	0.5000	1.0000	0.7071	0.7032	0.6875	0.6875	0.7001
Std	0.0114	0.1252	0.0000	0.0888	0.0633	0.1102	0.0688	0.0637
Fold	2C-KNN				2C-SVM			
	Acc	Sens	Spec	GM	Acc	Sens	Spec	GM
1	0.8750	0.9375	0.8125	0.8728	0.9062	1.0000	0.8125	0.9014
2	0.9375	0.9375	0.8750	0.9057	0.9062	1.0000	0.8125	0.9014
3	0.9062	0.9375	0.8125	0.8728	0.9062	1.0000	0.8125	0.9014
4	0.8750	0.8750	0.8750	0.8750	0.9375	1.0000	0.8750	0.9354
5	0.8438	0.8750	0.8125	0.8432	0.9062	1.0000	0.8125	0.9014
6	0.8750	0.8750	0.8750	0.8750	0.8750	0.9375	0.8125	0.8728
7	0.8750	0.8750	0.7500	0.8101	0.8750	0.9375	0.7500	0.8385
8	0.9375	0.8750	0.8750	0.8750	0.9375	0.9375	0.8750	0.9057
9	0.7812	0.9375	0.6250	0.7655	0.7812	0.9375	0.5625	0.7262
10	0.8438	0.9375	0.7500	0.8385	0.8125	0.9375	0.6250	0.7655
Mean	0.8750	0.9063	0.8063	0.8534	0.8844	0.9688	0.7750	0.8650
Median	0.8750	0.9063	0.8125	0.8728	0.9062	0.9688	0.8125	0.9014
Std	0.0466	0.0329	0.0804	0.0406	0.0511	0.0329	0.1029	0.0682

Table 3
Final validation results for the 4 types of models. TP, TN, FP and FN stand for True Positive, True Negative, False Positive and False Negative, correspondingly.

Method	Acc	Sens	Spec	GM	TP	TN	FP	FN
1C-KNN	0.9782	0.4500	0.9897	0.6673	0.0096	0.9686	0.0101	0.0117
1C-SVM	0.6803	0.6250	0.6815	0.6526	0.0133	0.6670	0.3117	0.0080
2C-KNN	0.8191	0.9250	0.8168	0.8692	0.0197	0.7995	0.1793	0.0016
2C-SVM	0.7734	1.0000	0.7685	0.8766	0.0213	0.7521	0.2266	0.0000

experiment results supported that hypothesis, with the KNN algorithm being the most robust, due to its higher tolerance for unbalanced data sets.

Future works include continued investigation into increasing accuracy through the use of more complex learning models, including GAN for large and heterogeneous environments, decision threshold tuning to minimize misclassification in unbalanced data sets, and further development of complementary detection methodologies that combine ML algorithms for OT networks with signature-based and threshold-based detection strategies for the IT components of CPS.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research has been funded by the Spanish Ministry of Science and Innovation under project MINECO-TIN2017-84804-R, PID2020-112726RB-I00 and the State Research Agency (AEI, Spain) under grant agreement No RED2018-102312-T (IA-Biomed).

References

[1] S. Zanero, Cyber-physical systems, Computer 50 (2017) 14–16, http://dx.doi.org/10.1109/MC.2017.105.

[2] M. Wolf, D. Serpanos, Safety and security in cyber-physical systems and internet-of-things systems, Proc. IEEE 106 (2018) 9–20, http://dx.doi.org/10.1109/JPROC.2017.2781198.

[3] D. Weissman, A. Jayasumana, Integrating IoT monitoring for security operation center, in: 2020 Global Internet of Things Summit, GIoTS, 2020, pp. 1–6, http://dx.doi.org/10.1109/GIOTS49054.2020.9119680.

[4] N. Jeffrey, Q. Tan, J.R. Villar, Anomaly detection of security threats to cyber-physical systems: A study, in: P. García Bringas, H. Pérez García, F.J. Martinez-de Pison, J.R. Villar Flecha, A. Troncoso Lora, E.A. de la Cal, Á. Herrero, F. Martínez Álvarez, G. Psaila, H. Quintián, E.S. Corchado Rodriguez (Eds.), 17th International Conference on Soft Computing Models in Industrial and Environmental Applications, SOCO 2022, Springer Nature Switzerland, Cham, 2023, pp. 3–12, http://dx.doi.org/10.1007/978-3-031-18050-7_1.

[5] H. Kaur, H.S. Pannu, A.K. Malhi, A systematic review on imbalanced data challenges in machine learning: Applications and solutions, ACM Comput. Surv. 52 (4) (2020) 1–36, http://dx.doi.org/10.1007/978-981-4585-18-7_2.

[6] P. Vuttipittayamongkol, E. Elyan, A. Petrovski, On the class overlap problem in imbalanced data classification, Knowl.-Based Syst. 212 (106631) (2021) http://dx.doi.org/10.1016/j.knosys.2020.106631.

[7] C. Esposito, G.A. Landrum, N. Schneider, N. Stiefl, S. Riniker, GHOST: adjusting the decision threshold to handle imbalanced data in machine learning, J. Chem. Inf. Model. 61 (6) (2021) 2623–2640, http://dx.doi.org/10.1021/acs.jcim.1c00160.

[8] A. Ahmim, L. Maglaras, M.A. Ferrag, M. Derdour, H. Janike, A novel hierarchical intrusion detection system based on decision tree and rules-based models, in: 2019 15th International Conference on Distributed Computing in Sensor Systems, DCOSS, Institute of Electrical and Electronics Engineers Inc, 2019, pp. 228–233, URL: https://ieeexplore.ieee.org/document/8804816/.

[9] C.M. Ahmed, G.R. M. R, A.P. Mathur, Challenges in machine learning based approaches for real-time anomaly detection in industrial control systems, in: Proceedings of the 6th ACM on Cyber-Physical System Security Workshop, ACM, 2020, pp. 23–29, http://dx.doi.org/10.1145/3384941.3409588.

[10] A. Abid, F. Jemili, O. Korbaa, Distributed architecture of an intrusion detection system in industrial control systems, in: Advances in Computational Collective Intelligence, Springer International Publishing, 2022, pp. 472–484, http://dx.doi.org/10.1007/978-3-031-16210-7_39.

[11] M. Al-Asiri, E.-S.M. El-Alfy, On using physical based intrusion detection in SCADA systems, Procedia Comput. Sci. 170 (2020) 34–42, http://dx.doi.org/10.1016/j.procs.2020.03.007.

[12] M. Altaha, S. Jong, Anomaly detection for SCADA system security based on unsupervised learning and function codes analysis in the DNP3 protocol, Electronics 11 (2022) 2184, http://dx.doi.org/10.3390/electronics11142184.

[13] N. Neshenko, E. Bou-Harb, B. Furht, A behavioral-based forensic investigation approach for analyzing attacks on water plants using GANs, Forensic Sci. Int. Digit. Invest. 37 (2021) 301198, http://dx.doi.org/10.1016/j.fsidi.2021.301198.

[14] I. Siniosoglou, P. Radoglou-Grammatikis, G. Efstathopoulos, P. Fouliras, P. Sarigiannidis, A unified deep learning anomaly detection and classification approach for smart grid environments, IEEE Trans. Netw. Serv. Manag. 18 (2021) 1137–1151, http://dx.doi.org/10.1109/TNSM.2021.3078381.

[15] S. Yilmaz, E. Aydogan, S. Sen, A transfer learning approach for securing resource-constrained IoT devices, IEEE Trans. Inf. Forensics Secur. 16 (2021) 4405–4418, http://dx.doi.org/10.1109/TIFS.2021.3096029.

[16] M. Al-Shabi, A. Abuhamdah, Using deep learning to detecting abnormal behavior in internet of things, Int. J. Electr. Comput. Eng. (IJECE) 12 (2022) 2108, http://dx.doi.org/10.11591/ijece.v12i2.pp2108-2120.

[17] M.R. Gauthama Raman, N. Somu, A.P. Mathur, Anomaly detection in critical infrastructure using probabilistic neural network, in: V.S. Shankar Sriram, V. Subramaniyaswamy, N. Sasikaladevi, L. Zhang, L. Batten, G. Li (Eds.), Applications and Techniques in Information Security, Springer, Singapore, 2019, pp. 129–141, http://dx.doi.org/10.1007/978-981-15-0871-4_10.

[18] N. Greggio, Anomaly Detection in IDSs by means of unsupervised greedy learning of finite mixture models, Soft Comput. 22 (2018) 3357–3372, http://dx.doi.org/10.1007/s00500-017-2581-z.

[19] S.S. Khan, A. Ahmad, Relationship between variants of one-class nearest neighbors and creating their accurate ensembles, IEEE Trans. Knowl. Data Eng. 30 (9) (2018) 1796–1809, http://dx.doi.org/10.1109/TKDE.2018.2806975.

[20] S. Agarwal, A. Sureka, Using KNN and SVM based one-class classifier for detecting online radicalization on Twitter, in: Distributed Computing and Internet Technology. ICDCIT 2015. Lecture Notes in Computer Science, Vol 8956, Springer International Publishing, 2015, pp. 431–442, http://dx.doi.org/10.1007/978-3-319-14977-6_47.

[21] B. Schölkopf, R.C. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, Support vector method for novelty detection, in: NIPS'99: Proceedings of the 12th International Conference on Neural Information Processing Systems, ACM, 1999, pp. 582–588, URL: https://proceedings.neurips.cc/paper/1999.

[22] K. Kevin Lamshöft, T. Neubert, C. Krätzer, C. Vielhauer, J. Dittmann, Information hiding in cyber physical systems: Challenges for embedding, retrieval and detection using sensor data of the SWAT dataset, in: Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security, ACM, 2021, pp. 113–124, http://dx.doi.org/10.1145/3437880.3460413.

[23] M. Elnour, N. Meskin, K. Khan, R. Jain, A dual-isolation-forests-based attack detection framework for industrial control systems, IEEE Access 8 (2020) 36639–36651, http://dx.doi.org/10.1109/ACCESS.2020.2975066.

[24] J.L. Leevy, T.M. Khoshfogtaar, A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018, J. Big Data 7 (104) (2020) http://dx.doi.org/10.1186/s40537-020-00382-x.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830, URL: https://dl.acm.org/doi/10.5555/1953048.2078195.

Nicholas Jeffrey received the Master of Science degree from Athabasca University in 2021. He is currently pursuing the PhD degree in the Department of Computer Science at the University of Oviedo, Spain.

Dr. Qing Tan received the Ph.D. in Engineering Cybernetics from the Norwegian Institute of Technology in 1993. He is currently an Associate Professor in the School of Computing and Information Systems at Athabasca University, Canada.

Dr. José R. Villar received the PhD in Computer Science from the University of León in 2002. He is currently an Assistant Professor in the Department of Computer Science at the University of Oviedo, Spain.