Computers in Industry 142 (2022) 103741

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.sciencedirect.com/journal/computers-in-industry

Multi-step attack detection in industrial control systems using causal analysis

Zahra Jadidi [a, b, *], Joshua Hagemann [a], Daniel Quevedo [c]

[a] School of Computer Science, Queensland University of Technology, Brisbane, Australia
[b] School of Information and Communication Technology, Griffith University, Australia
[c] School of Electrical Engineering and Robotics, Queensland University of Technology, Brisbane, Australia

ARTICLE INFO

Keywords:
Attack detection
Causality analysis
Industrial control systems
Multi-step attacks

ABSTRACT

In the old generation of industrial control systems (ICSs), their sub-components communicated within private networks and, therefore, it was assumed that ICSs are safe from cyber-attacks. However, new advanced ICS sub-components need Internet connectivity to control and monitor their geographically dispersed structure. Connection to corporate networks and the public Internet create various security issues. The increasing number of attacks has become a serious threat for ICS networks. These sophisticated attacks use multiple steps and affect different devices. A major weakness of existing attack detection methods is that they only detect attacks and they do not help security analysts identify the cause and effect of attacks. Therefore, manual analysis is required to identify and isolate the cause of the attack. Causal analysis can help to track the propagation of an attack. While there is weak security in ICS networks, there is not sufficient research in the causal analysis of attacks in these networks. To address this research gap in ICS networks, we present a solution that detects the causal impact of attacks by investigating causal dependencies in ICS logs. Our ICS causal anomaly detection (ICS-CAD) method consists of two phases. It initially detects attacks and identifies the ICS device generating the malicious traffic. Secondly, it analyses causal relationships between ICS logs to diagnose the attacker's future effect. We use a causal decomposition method to discover causality relationships in ICS logs. The performance of the ICS-CAD is evaluated using two datasets collected in real-world ICS networks. The ICS-CAD provides 98% accuracy in detecting attacks and the causal impact of the detected attacks.

1. Introduction

Industrial control system (ICS) networks are defined as industrial systems (operational technology (OT) networks) integrated with information technology (IT) networks (Tuptuk and Hailes, 2018; Jadidi and Lu, 2021). The connection of vulnerable industrial systems to IT networks has increased the security risks in ICS networks. ICSs are increasingly the target of sophisticated cyber-attacks which act out their behaviour over an extended period. Traditional intrusion detection systems are designed to monitor a single source of data and detect attacks. However, they do not identify the cause and effect of an attack. Despite IT networks, ICS networks contain interconnected vulnerable industrial devices. Therefore, it is critical to deploy a scalable anomaly detection method that can effectively track an attack's cause and effect and predict the devices that can be affected by the attack (Khosravi and Ladani, 2020; Chakraborty et al., 2019).

Attackers in ICS networks usually take multiple actions affecting different components of a network. To defend against these multi-step attacks, the state in which the attack is, as well as what the next action it is likely to take, needs to be known. The sub-components in an ICS network are interlinked and analysing the logs of one device may reveal events triggered in another (Ray et al., 2019). Multi-step attack detection in ICS networks using analysis of correlated logs has not been sufficiently studied. This is the research gap addressed in our paper.

Extracting causal relationships between events in device logs can give insight into the operations of the system. If it is known what device an attack has impacted, then a graph of causal events can be developed and used to create a plan of action to stop multi-step attacks from progressing further.

Causality analysis was employed in some papers (Khosravi and Ladani, 2020; Liu, 2018; Zhang et al., 2020) for anomaly detection in IT networks. However, the novelty of our paper is presenting a combination of anomaly detection and causality analysis to provide a system-wide anomaly detection in ICS networks. In this paper, we

* Corresponding author at: School of Computer Science, Queensland University of Technology, Brisbane, Australia.

https://doi.org/10.1016/j.compind.2022.103741
Received 8 November 2021; Received in revised form 28 June 2022; Accepted 1 July 2022
Available online 26 July 2022

present a causal anomaly detection solution for ICSs, called ICS-CAD, where an anomaly detection will be coupled with causality analysis to detect attacks and subsequently identify the signs of attack influences. Our proposed ICS-CAD solution contains two-layer analysis: .

- Attack detection using deep learning: In this layer, a particular node in the network runs deep neural network-based sequence classification to analyse time series received from multiple ICS devices and detect attacks. The deep learning method can detect the attack type and identify the ICS device generating the malicious behaviour. In the deep learning method, we perform sequence classification based on Long Short Term Memory (LSTM) networks.
- Causality analysis using causal decomposition method: Causality analysis is performed when an attack is detected in an ICS device. This serves to identify the causal impact of the attack and predict other devices which can be infected by this attack in the future. While causality analysis has been employed in some studies, their approaches were complex with low accuracy. However, we have applied a causal decomposition method, which is a well-known method in other areas like medicine, to ICS data. The results showed high accuracy in detecting attacks and the causal impact of the attacks.

The rest of the paper is organised as follows: Section 2 reviews background and related works. Section 3 discusses the architecture of the proposed approach, the anomaly detection method, and the causality analysis method. In Section 4, we explain the datasets used to evaluate the proposed method. Section 5 shows the evaluation results. Section 6 discusses the efficiency of our method. Finally, Section 7 concludes the paper.

2. Background and related works

In the old generation of industrial control systems, their subcomponents communicated within physically isolated networks, and therefore, they were not designed for security. However, new ICS subcomponents need Internet connectivity to control their geographically dispersed structure. Connecting these vulnerable devices to corporate networks and the public Internet creates various security issues. Due to the increasing number of security incidents in ICS networks, anomaly detection in industrial systems has received a lot of attention during the past few years (Sapkota et al., 2020; Gómez et al., 2019; Feng et al., 2017; Abdelaty et al., 2021).

Although advanced cyber-attacks in ICS environments use multiple steps and affect different devices to reach their final target, existing anomaly detection methods are mostly based on local analysis, and they do not investigate the impact of the attacks on other devices (Jadidi et al., 2021).

Research around the detection of multi-step attacks has been done in some papers (Navarro et al., 2018; Lughofer et al., 2020; Ghafir et al., 2018; Chandra et al., 2016; Harikrishnan and Kumar, 2018; Milajerdi et al., 2019; Alshamrani et al., 2019). These works can be categorised into two groups: detection based on attack signatures, and detection based on modelling the attack process. Multi-step attacks have dynamic behaviours, and they follow diverse attacking techniques to reach their goal. Therefore, detection based on attack signatures is not efficient for zero-day attacks. On the other hand, the lack of proper algorithms with high accuracy is the weakness of detection based on attack process modelling (Khosravi and Ladani, 2020).

One of the methods proposed in the literature to detect multi-step attacks in earlier stages is causality analysis of anomalies (Shi et al., 2017). Discovering the cause and effects of attacks can be used to detect multi-step attacks in the earlier steps and stop future steps. Cause detection in ICS networks is a manual task, and isolating the cause is hard and time-consuming. Causality analysis is a challenging problem in ICS security, and it can be used to discover the causal impact of an event on other devices.

Researchers have proposed various causality analysis methods for anomaly detection (Shi et al., 2017; Khosravi and Ladani, 2020). In particular, (Khosravi and Ladani, 2020) proposed a model based on correlating the security and non-security sensor alerts. This model could discover causal relationship sequences. The proposed model analysed alerts received from different sensors, and it could identify the alerts that could form a sequence of attack steps. Finally, the authors used the probability of attacking different hosts in the system to identify the causal attack chains. This approach considers the detection of multi-step attacks and their diverse behaviours in IT networks. Multi-step attack detection is also investigated in Liu (2018); Zhang et al. (2020). However, these proposed solutions are for attacks in IT networks, and they did not investigate the complexity of attacks in ICS networks including the combination of IT and OT networks.

Attack detection in industrial systems using causality countermeasures was investigated by Shi et al. (2017). Transfer entropy-based causality countermeasures were used in this paper to discover the causal relationships between the countermeasures and the system parameters. The authors measured causal relationships using transmission entropy for sensor measurement sequences. Causality based analysis in ICS networks is also introduced by Zhang et al. (2020). In this paper, causal relationships between nodes were extracted using maximum information coefficient and transfer entropy. However, these entropy-based methods provide computational complexity, and they are not suitable for large-scale networks. Nonlinear Granger causality graph method is used in another paper to detect attack targets in cyber-physical systems (Li et al., 2021). However, this method provides low accuracy. Detection of multi-step attacks in ICS networks and accurate prediction of the future steps of the attacks using causality analysis is the research gap which is addressed in our paper.

Here, we present an automated solution, called ICS-CAD, for multi-step attack detection using deep learning-based anomaly detection and causal decomposition method. An ensemble empirical mode decomposition (ensemble EMD) is used in the causal decomposition method (Yang et al., 2018) to decompose a time series into a finite number of intrinsic mode functions (IMFs). This helps to accurately detect the causal interaction between two time series. Our proposed ICS-CAD method will be able to monitor ICS logs, detect attacks and the device generating the malicious behaviour, and identify the causal impact of the attack on other ICS devices.

3. Multi-step attack detection using causality analysis

The architecture of our presented solution called ICS-CAD for detecting multi-step attacks is shown in Fig. 1. Attack detection is the first phase of our ICS-CAD method to detect attacks and the node generating malicious behaviour. Then, causality analysis in the next phase can track the attack steps, and it will identify other devices which will be affected by this attack in the future. After finding the malicious source, performing causality analysis gives direction into where the cyber-attack may affect the ICS network.

Causality based multi-step attack detection is about analysing dependency changes in multivariate time series. When there is a strong causal relationship between two time series, it means the second time series is affected by the first malicious device. This means that it can comprehensively track anomalies throughout an entire system. This is advantageous as many detection systems focus on monitoring individual devices and features but here, we incorporate causality analysis on top of anomaly detection to create a broader system-wide detection system. If an anomaly is detected in one device, the correlated nodes are checked for anomalies to see if the anomaly is isolated, spreading, or a part of a larger chain of anomalies.

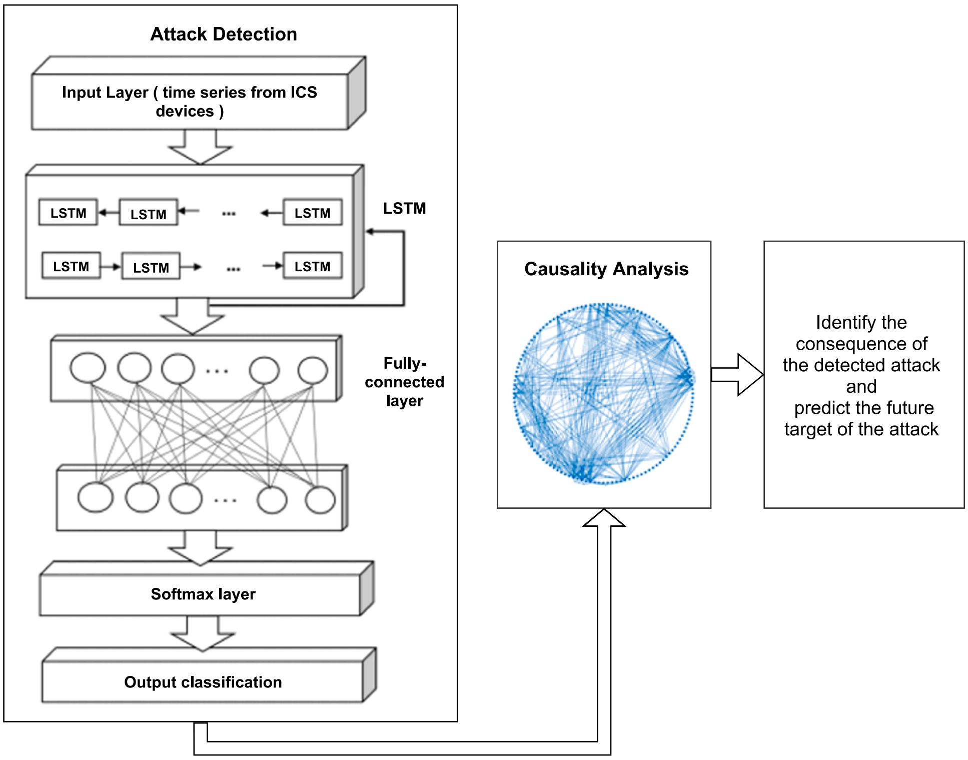

Attack Detection
Input Layer (time series from ICS devices)
LSTM
LSTM
LSTM
...
LSTM
Fully-connected layer
Softmax layer
Output classification
Causality Analysis
Identify the consequence of the detected attack and predict the future target of the attack

Fig. 1. The architecture of the proposed ICS-CAD method.

3.1. Attack detection phase

A deep neural network-based sequence classification (DNN-SC) method is deployed in the attack detection phase to analyse data received from multiple ICS devices and detect the malicious node. This DNN-SC method provides sequence classification to detect malicious time series. The architecture of DNN-SC based attack detector is shown in Fig. 1. This DNN-SC classifier is based on LSTM neural network, and it is a supervised method which needs a labelled dataset. This labelled dataset is a set of time series collected from different sources. In our study, the input time series is received from ICS devices. The output of the deep learning corresponds to the type of attacks in the dataset, and it also identifies the ICS device generating the attack traffic (Xia et al., 2019; Jurgovsky et al., 2018; Luo et al., 2021; Jadidi et al., 2020).

We used an LSTM layer with 100 hidden units in our DNN-SC. The LSTM layer is connected to a fully connected layer followed by a softmax layer and a classification layer. An LSTM neural network was used for anomaly detection because it is well suited for processing and making predictions based on time series data (Chalapathy and Chawla, 2019). Our DNN-SC is evaluated using two publicly available ICS datasets (Anon, 2022; Goh et al., 2016). After an attack is detected, the impact of the attack will be investigated in the causality analysis phase to see if the attack is potentially part of a greater attack.

3.2. Causality analysis phase

A causal decomposition method is used in our paper to discover causal relationships (Yang et al., 2018). This method was chosen in our paper as it has shown high accuracy in identifying causal relationships between time series in other applications like medical applications (Alex et al., 2020). This is achieved by using ensemble empirical mode decomposition (EMD) to decompose a time series into a finite number of intrinsic mode functions (IMFs). The IMFs correspond to different frequencies and residues which make up the components of the time series. The phase coherence between paired IMFs is used in this causality analysis method to quantify the variance-weighted Euclidean distance between paired IMFs decomposed from the original signals, and the paired original and re-decomposed IMFs. A causal relationship is defined by "cause is that which put, the effect follows; and removed, the effect is removed" (Yang et al., 2018). The Pseudocode of this causal decomposition method is provided in Table 1 to summarise the different steps involved in this method. The causal strength between paired IMFs is the output of this method. This causal strength will be discussed in Section 5.

Discovering changes which have been made by an attack requires analysis of all the components of the time series. If the IMF which contains the data of the attack is removed, then the causal response from that attack will also be removed (Tuptuk and Hailes, 2018). A noise level, which is the fraction of the standard deviation of time series, is also used in the causal decomposition method, and it is added to this method to ensure the separability of IMFs. In our paper, the noise level is

Table 1
Causal decomposition Pseudo-code.

Pseudo-Code: Causal decomposition method

```
Input: Time series from industrial control systems
Output: Causality detection in ICS devices
//Causality analysis
foreach pair of input time series do
    - decompose the input time series (t1 and t2) into a set of IMFs
    - use ensemble EMD parameter and added noise level to ensure
    the separability of the IMFs
    foreach pair of IMFs do
        - remove one of the IMFs (e.g. IMF1) from t2 and subtract IMF1
        from the original t2
        - redecompose the time series and calculate the phase coherence
        between the original IMFs of t1 and redecomposed IMF1 of t2
        - repeat this decomposition and redecomposition procedure for
        IMF1 of t1
        - use the variance-weighted Euclidean distance between the phase
        coherence of the original IMFs and redecomposed IMFs to identify
        the causal strength
    end
end
```

manually identified for each pair of time series.

In ensemble EMD, every observed data consist of the original time series and added noise. This method converts signals with different scales to appropriate scales associated with relevant added noise (Mao et al., 2020).

The ensemble EMD in the causal decomposition method is a noise-added data analysis method which uses noise to increase the separability among IMFs during the decomposition (Yang et al., 2018). This method defines IMF components, $S_j(t)$, as shown in Eq. 1.

$$S_j(t) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} S_j(t) + r \times w_k(t) \tag{1}$$

where $w_k(t)$ shows the white noise which is artificially added to the original signal. k means the kth trial of the jth IMF in a noise-added signal. r identifies the magnitude of the added white noise which is a critical parameter in enhancing the orthogonality and separability of the IMFs in ensemble EMD. r is defined as a fraction of a standard deviation of the original signal. Using the added noise in the ensemble EMD, there will be a uniform reference frame in the time-frequency space by presenting the decomposed IMFs by comparable scales which are independent of the nature of the original signals. The strategy of selecting r in this method is to minimise the root-mean-square (RMS) of the pairwise correlation of the IMFs (ideally under 0.05).

In Eq. 1, N shows the number of trials in the ensemble EMD. The value of N should be large, because the added noise in each trial will be cancelled out in the ensemble mean of large trials, and hence, the added noise does not falsify the time series. In our paper, we selected $N = 1000$.

$S_j(t)$ is on both sides of Eq. 1. However, the right-side $S_j(t)$ means the original signal, at time t, which is added to the white noise of a finite amplitude Yang et al. [34].

The noise level used in the causal decomposition method is the fraction of the standard deviation of time series, and in this method, we need to add this noise in the data to ensure the separability of IMFs. In this paper, we utilise the causal decomposition method (Yang et al., 2018) to identify the cause-effect relationships in an ICS network. Different scenarios are selected to evaluate the performance of ICS-CAD. For each scenario, we identify the parameters of the noise level separately. For example, in scenario 41, a noise level of 0.35 is identified. The noise level in this method illustrates how ensemble EMD separates the intrinsic components of a given time series (Yang et al., 2018). Therefore, the noise levels will be changed for different types of scenarios.

4. Datasets

Two datasets were used in our paper to evaluate the efficiency of ICS-CAD, the Power System dataset (Anon, 2022) and the Secure Water Treatment dataset (Goh et al., 2016). These datasets contain the data required for the causality analysis of multi-step ICS attacks. The multi-step attack scenarios used for the evaluation of ICS-CAD are the attacks attacking multiple devices.

4.1. Power system datasets

The Power System framework used for the data collection contains different sections, each including breakers and Intelligent Electronic Devices (Anon, 2022; Pan et al., 2015). There are 128 unique devices used in collecting data for the operation of the systems. The Power System dataset (Anon, 2022) has different scenarios which include normal events and attack events. The attacks provided in this dataset are relay setting change and remote tripping command injection.

4.2. Secure water treatment (SWaT) dataset

The Secure Water Treatment framework consists of six main processes corresponding to the physical and control components of a water treatment facility. There are 51 tags (25 sensors and 26 actuators) in this dataset (Goh et al., 2016). Tag names show their roles. For example, MV is a motorised valve, P is used for pump, LIT denotes level indicator transmitter, UV is an ultraviolet actuator, and AIT is defined as analyser indicator transmitter. The attacks provided in this dataset were spoofing and man-in-the-middle attacks.

Our deep learning-based anomaly detection and the causal decomposition method were evaluated based on multiple scenarios in the Power Systems and SWaT datasets. The scenarios examined include normal operation and attack event scenarios. The scenarios used to evaluate the performance of our ICS-CAD method generate multi-step attacks, shown in Table 2 (Anon, 2022; Goh et al., 2016). The ground truth for causal relationships in this dataset was manually generated.

For both SWaT and Power System Datasets, the details of simulated testbeds, and the interactions between devices in each attack scenario are provided in Goh et al. (2016) and Anon (2022) respectively. Using the information of device interactions and the timestamp of attack samples provided in Goh et al. (2016); Anon (2022), we could obtain the ground truth of the causal interactions.

5. Evaluation results

The performance of ICS-CAD in Power System and SWaT datasets will be discussed in this section.

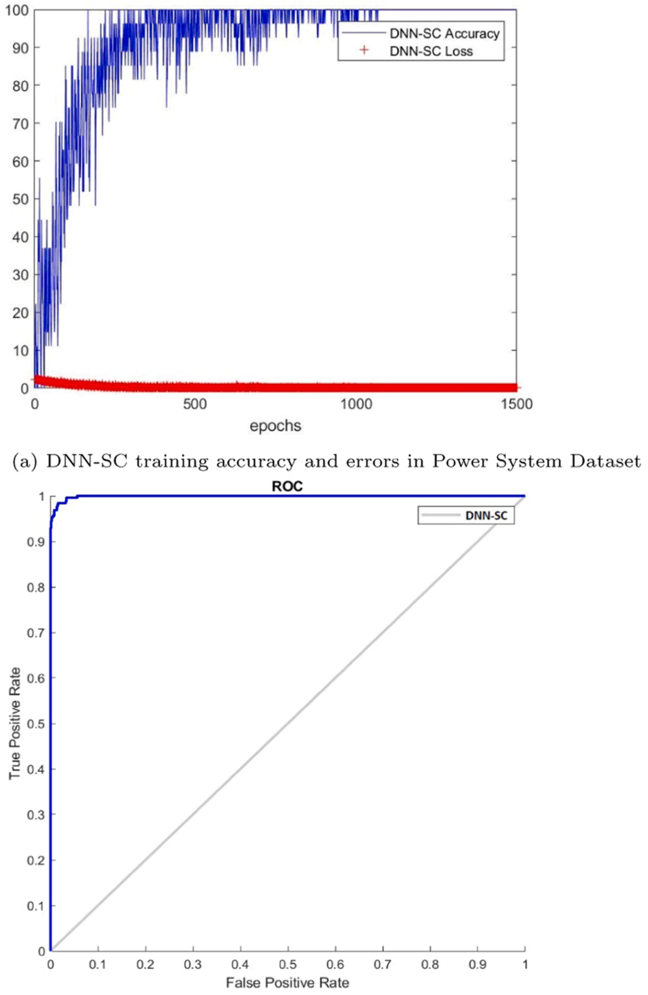
5.1. Attack detection

In this section, the DNN-SC method is initially evaluated using the Power System dataset. Then, the results of this method in the SWaT dataset will also be discussed. The Power Systems dataset has different types of events (Table 2) which shows real-world scenarios. DNN-SC is used to classify the input sequence data received from ICS devices. An LSTM uses time-series data as an input and sends it to the classification layer to detect the malicious time series. The training dataset (Power System dataset) contains time series of multiple IEDs in ICS networks. Fig. 2a illustrates the accuracy of DNN-SC with different iterations in the training phase in the Power System dataset. The accuracy in this figure is 100 % after 1000 iterations. Fig. 2b is the ROC curve of the DNN-SC, and it shows the high True Positive rate of the DNN-SC method in the Power System dataset.

The datasets used in this paper are divided into a training set (70 %), a validation set (10 %) and a testing set (20 %) for the evaluation of our DNN-SC method. To show the real-world situations, testing datasets contain both known (10 %) and new samples (10 %). We included

Table 2
Scenarios for the power system and SWaT datasets.

Scenario	Description of Scenarios Power System Dataset
41	Normal operation load changes
19	Remote tripping command injection: Command injection to Intelligent Electronic Device 1 (IED1) and IED2
35	Attack sub-type (Disabling relay function): Fault on Line maintenance 1 (L1) with IED1 and IED2 disabled
39	Attack sub-type (Disabling relay function): Two IEDs disabled with line maintenance (L1)
	SWaT Dataset
22	Attack points: UV401 and P-501 actuators, and AIT-502 sensor. This attack stops UV-401; changes the value of AIT-502; and keeps P-501 on.
23	Attack points: P-602 and MV-302 actuators, and DIT-301 sensor. This attack changes the value of DPIT-301; keeps MV-302 open; and keeps P-602 closed.
30	Attack points: LIT-101 sensor, and P-10 and MV-201 actuators. The attack turns P-101 and MV-101 on continuously, and changes the value of LIT-101. Then, LIT301 level becomes low and P-102 starts itself.



(a) DNN-SC training accuracy and errors in Power System Dataset

(b) ROC curve for DNN-SC in Power System Dataset

Fig. 2. Performance of DNN-SC in Power System Dataset.

unknown samples in the testing dataset to evaluate the performance of DNN-SC in unknown attacks. The performance metrics provided in Table 3 show the efficiency of our DNN-SC in new attacks in the scenarios defined in this paper.

In this study, it is assumed that the labelled datasets are available for the classifier. In the labelled SWaT and Power System datasets, there are multiple executions of each attack. We aim to investigate how causal analysis can be combined with deep learning-based anomaly detection to track attack activities. Therefore, a supervised classifier was selected as it has shown a high detection rate (Haylett et al., 2021). However, in our future work, we intend to implement an unsupervised learning-based causality analysis to be able to detect anomalies in unlabelled datasets and improve the anomaly detection results in zero-day attacks with new behavioural patterns.

Table 3
Performance of DNN-SC compared with other studies.

Dataset	Method	Precision	Recall	F1	Accuracy
Power System	DNN-SC	0.98	0.94	0.96	0.97
Power System	Random Forest (Farrukh et al., 2021)	0.95	0.95	0.95	0.95
Power System	K-nearest Neighbours (Elmrabit et al., 2020)	0.88	0.91	0.90	0.85
Power System	Adaptive Boosting (Elmrabit et al., 2020)	0.73	0.96	0.83	0.72
SWaT	DNN-SC	0.99	0.89	0.93	0.98
SWaT	SVM (Kravchik and Shabtai, 2018)	0.95	0.70	0.80	.
SWaT	1D CNN combined records (Kravchik and Shabtai, 2018)	0.97	0.79	0.87	.
SWaT	1D CNN ensembled records (Kravchik and Shabtai, 2018)	0.87	0.85	0.86	.

The performance of DNN-SC in detecting anomalies of the Power System dataset is shown in Table 3. F1 score in this table is defined as Eq. 2 (Jadidi et al., 2020).

$$F_1 = 2 \cdot \frac{Precision.Recall}{Precision + Recall} \tag{2}$$

where Recall shows our method's ability to detect anomalies, and Precision shows the percentage of correct alarms generated by our system.

Table 3 shows the evaluation results of the DNN-SC in the SWaT dataset and Power System dataset. The performance of DNN-SC is also compared with other papers in this table. As it is shown, DNN-SC provides high precision and F1 score comparable with other machine learning-based anomaly detectors.

The results of papers (Farrukh et al., 2021; Elmrabit et al., 2020; Kravchik and Shabtai, 2018) are limited to one of our ICS datasets provided in Table 3. Papers (Farrukh et al., 2021; Elmrabit et al., 2020) are based on Power System dataset, and paper (Kravchik and Shabtai, 2018) is based on SWaT dataset. Each paper investigated different sets of machine learning methods to analyse the relevant dataset and find a classifier with high performance metrics. Therefore, the sets of machine learning methods evaluated in these papers are different, and our DNN-SC has been compared with different methods.

The authors of the paper (Kravchik and Shabtai, 2018) used Precision, Recall, and F1 score to evaluate the efficiency of their anomaly detectors, and the Accuracy values of their methods in SWaT dataset are not available to be presented in Table 3.

Sequential dependency between input sequence data is used in DNN-SC for classification of the data into five classes including four scenarios in Power System dataset (Table 2), and another class for detecting the device generating the malicious behaviour. For SWaT dataset, there are four outputs, three scenarios and a node for detecting the malicious device.

In Table 3, SVM is a one-class SVM classifier. 1D CNN combined records show a single convolutional neural network classifier in which a combined architecture is used including a stack of convolutional layers which send the output to LSTM layers to make the prediction. 1D CNN ensembled records show a single convolutional neural network in which an ensemble approach is used to combine the detection results of individual CNNs. This ensembled approach was used to determine which attacks were not detected by the CNN methods.

5.2. *Causality analysis*

As discussed in Section 4, four scenarios are used in the Power System dataset to evaluate the performance of our ICS-CAD method. Initially, it is assumed that scenario 41 is detected by the DNN-SC detector. At this stage, causality analysis will be performed in ICS-CAD to identify the consequences of event 41 on other devices.

The number of ensembles was 1000 times in scenario 41. The number of ensembles is defined as the parameter that affects the error of IMFs. This error shows the degree of distortion influenced by the added noise. It is suggested to set the degree of distortion less than 0.05 in the ensemble EMD (Jurgovsky et al., 2018).

For each pair of time series, the causal decomposition method provides *n* number of IMFs decomposed from the data. We have visualised IMFs decomposed from the time series of two Intelligent Electronic

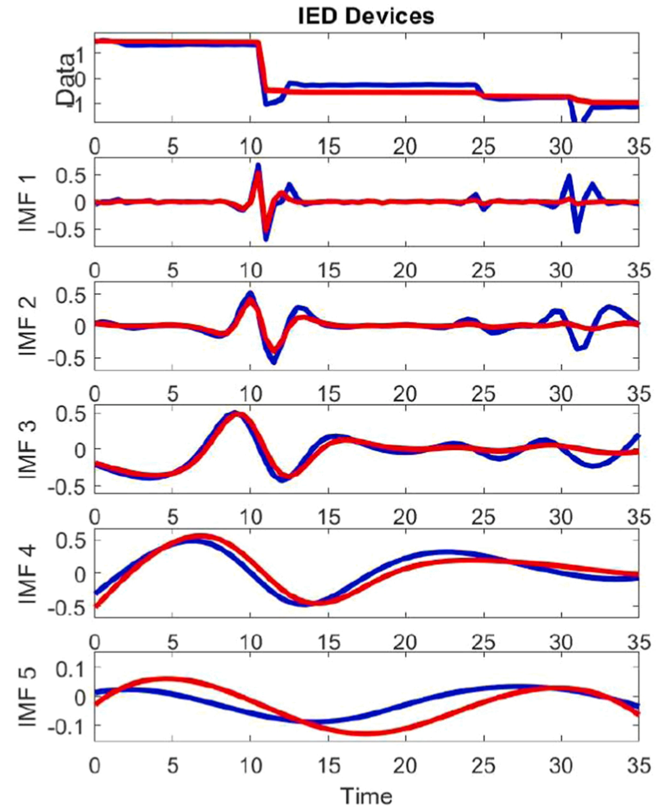


Fig. 3. IMFs decomposed from time series of two IED devices in Power System Dataset.

Devices (IEDs) in scenario 41 in Power System dataset in Fig. 3. This visualisation shows the relationship between the intrinsic components of the time series.

Then, the final output of the causal decomposition method is causal strength between IMFs decomposed from two time series. The causal strength calculated for IED1 and IED2 in Power System Dataset is shown in Fig. 4 where a ratio of 0.5 shows no causality is detected and a ratio approaching 0 or 1 shows a strong causal relationship from either IED1 or IED 2 respectively.

This figure shows that among IMFs decomposed from two time series, the strongest causal influence is in IMF3 from IED 1 to IED2. The purpose of causal decomposition is to discover the effectiveness of the covariation principle of cause and effect for identifying causality relationships (Yang et al., 2018; Tuptuk and Hailes, 2018). Using Fig. 4, we could generate causal graphs of attacks in Section 5.3 (Fig. 5).

5.3. Accuracy of causality analysis

To analyse the accuracy of the results, a 20 % test was performed. Each test dataset was sampled at 20 % of the original number of entries and the causal relationships were calculated. The accuracy of causal detection is defined as Eq.3.

$$Accuracy = \frac{TP + TN}{TN + FP + TP + FN} \tag{3}$$

where True Positive (TP) shows the correct detection of the causal relationships. True Negative (TN) is when our causal identifier correctly identifies the lack of causal relationship. False Positive (FP) is the wrong detection of a causal relationship, and False Negative (FN) shows that the causal relationship is not detected.

To create a model for causal relationships between devices, every

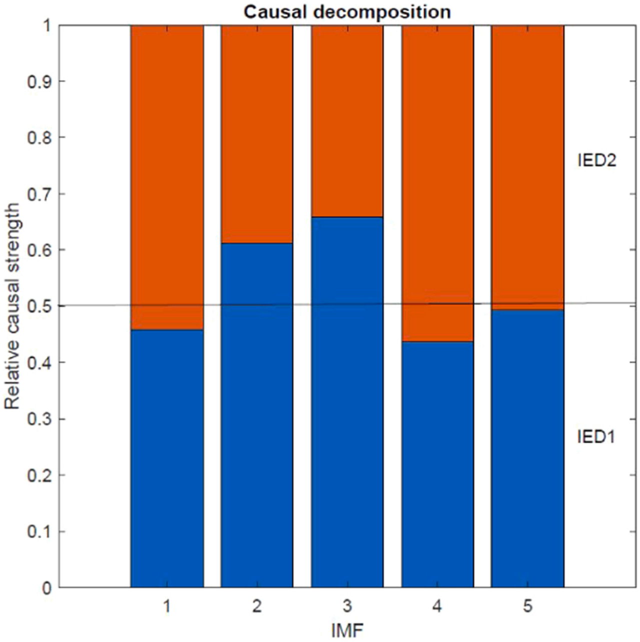


Fig. 4. Causal strength between IED1 and IED2.

device's time-series log was analysed against every other device log within the system via the causal decomposition function. The causal decomposition function discovers the causal strength score of each IMF for the time series. The causal strength score is recorded for every relationship between the devices and ranges from 0 to 1. A noise level is the fraction of the standard deviation of time series. The level of noise used in the EMD and the number of ensembles can be changed as input parameters for this function. Values of 0.35 for noise and averaging of 1000 ensembles were used in scenario 41. For all attack scenarios in the Power System dataset and SWaT datasets, noise levels were separately identified.

A threshold value should also be defined in the causal decomposition method. In this paper, the accuracy was calculated over a range of thresholds to identify which causal threshold gave the most accurate data. The highest accuracy scores and the respective threshold values are recorded in Table 4.

This table shows the impact of the threshold values on the accuracy of causal detection. To verify the accuracy of causal interactions, the ground truth for each scenario was tested against the respective causal score.

Based on our experiments, we found out that if the threshold is too low, there is an increased chance that the causal relationships are false positives. On the other hand, having a very high threshold might eliminate the target causal relationship of the attack. As shown in Table 4, high threshold values decreased the accuracy values. Causal relationships can be measured against a threshold score which defines the strength of the relationships. These relationships can be plotted into a directed graph for visualisation (Fig. 5). The causality score calculated for each device log ranges from 0 to 1. The number of causal relationships identified by the causal decomposition method is changed based on the threshold values. As can be seen in Fig. 5, for a low threshold score of 0.5, two-hundred and twenty-one causal relationships were discovered for Scenario 41. For a score greater than 0.75, only fifteen causal relationships were identified. The impact of threshold values on the accuracy of the causal decomposition method in each scenario is shown in Table 4. Based on the type of attacks detected in ICS-CAD, the threshold values identified in Table 4 will be used to find the consequence of the attack.

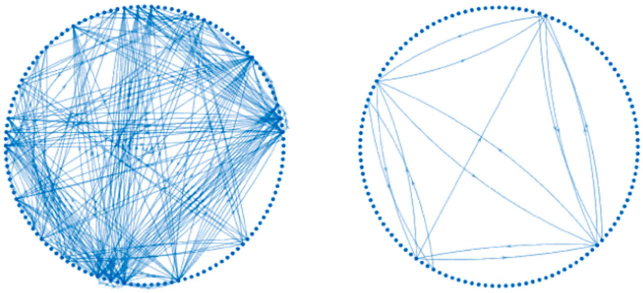


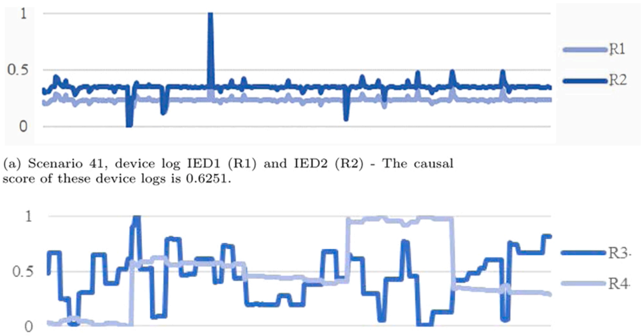
Fig. 5. Causal relationships discovered for Scenario 41, 221 causal relationships with a threshold score greater than 0.5 (left) and 15 causal relationships with a threshold greater than 0.75 (right).

The threshold score of 0.75 was selected as the optimum threshold that can provide the greatest accuracy in scenario 41 in the Power System dataset. The higher the causal relationship score, the closer the correlation of a time series is to another, and the more likely it is that the causal relationship can be recognised by simple human pattern recognition. This is evidenced in Figs. 6a and 6b.

In the Power System dataset, the values of IEDs are constantly changing (Figs. 6a and 6b). In this behavioural pattern, the causal decomposition method could accurately detect causal relationships by analysing the data during the attack. However, for the SWaT dataset, the results of the causal decomposition method were affected by the constant values of device logs, Fig. 7. In constant values like SWaT dataset, data before the attack should also be provided to the causal decomposition method. To make the causal decomposition method applicable to both behavioural patterns provided in these datasets, for each detected attack, both logs before and during the attack were analysed. The causal decomposition method could detect all the causal relationships in both datasets after we included the data prior and data during the attack. Table 5 shows the impact of including this data on our causal decomposition method.

Table 4
20% Accuracy test for causal relationships.

Scenario	Accuracy	Threshold value
	Power system dataset	
41	0.391	0.50
	0.52	0.624
	1	0.75
	0.8	0.80
19	0.352	0.568
	0.524	0.590
	1	0.729
	0.85	0.780
35	0.346	0.560
	0.572	0.541
	1	0.746
	0.9	0.80
39	0.346	0.581
	0.552	0.553
	1	0.712
	0.91	0.760
	SWaT Dataset	
22	0.401	0.412
	0.512	0.571
	1	0.691
	0.89	0.75
23	0.328	0.312
	0.601	0.454
	1	0.741
	0.88	0.80
30	0.438	0.415
	0.636	0.529
	1	0.752
	0.91	0.80



(a) Scenario 41, device log IED1 (R1) and IED2 (R2) - The causal score of these device logs is 0.6251.

(b) Scenario 41, device log IED3 (R3) and IED4 (R4) - The causal score of these device logs is 0.3005.

Fig. 6. Causal scores between different IEDs in Power System Dataset.

6. Discussion

The proposed ICS-CAD method in this paper contains two consecutive phases, attack detection and causality analysis, to provide accurate multi-step attack detection in ICS networks. A deep learning-based sequence classifier employed in the attack detection phase showed high accuracy in detecting attacks in two datasets, the Power System dataset and the SWaT dataset.

In this paper, we used labelled datasets in which both normal and attack samples are labelled. We assumed that labelled datasets are available for training our anomaly detector. Our supervised classifier detects attack samples, and then, the timestamp of the detected attack is used to extract the data prior and during the attack from the original dataset. To generalise our solution to unlabelled datasets, the efficiency of unsupervised learning in causality analysis will be studied in our future work.

The causal decomposition method was used in the second phase to identify the causal impact of the detected attacks. In the causal decomposition method, determining the required threshold value was essential to extract meaningful causal relationships from events in device logs. If the threshold is too low, then, there is an increased chance that the causal relationships are false positives. Conversely, having a high threshold might eliminate the target causal relationship of the attack. The threshold value needed to optimise causal results is dependent on the dataset and attack types.

The performance of the causal decomposition method was also dependant on the behavioural patterns of the ICS devices. For changing patterns, like the Power System dataset, the attack data must only exist at some point within the time series for the causal decomposition method to correctly identify the causal relationships. To apply the causal decomposition method to datasets with constant values like the SWaT datasets, we need to include data prior and during attacks to increase the accuracy of causality analysis. In this paper, we used the combination of data prior and during attacks as the input of the causal decomposition

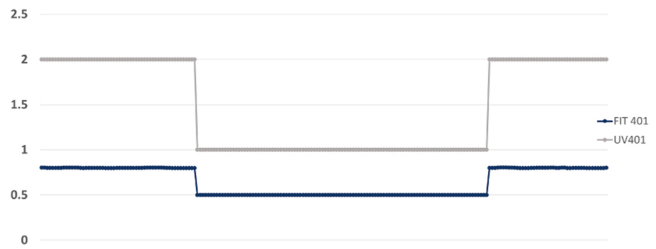


Fig. 7. The causal score of a flow rate sensor (FIT) sensor and a UV actuator in SWaT dataset is 0.71.

Table 5
Impact of including data prior and data during the attack on the causality decomposition method.

Scenario	Attack data included	Identify Causal relationship between attack sensors
	SWaT Dataset	
22	Data prior	No
	Data points prior and during	Yes
	Data during only	No
23	Data prior	No
	Data points prior and during	Yes
	Data during only	No
30	Data prior	No
	Data points prior and during	Yes
	Data during only	No
	Power System Dataset	
19	Data prior	No
	Data prior and during	Yes
	Data during only	Yes
35	Data prior	No
	Data prior and during	Yes
	Data during only	Yes
39	Data prior	No
	Data prior and during	Yes
	Data during only	Yes

method in both datasets, and hence, our ICS-CAD method is applicable to both types of changing and constant values.

The performance results, presented in Tables 3 and 4, showed that ICS-CAD could detect 98% of the attacks and all causal relationships for each detected attack. It means that the causal decomposition method could detect 100% the causal impact of identified attacks. This causal impact illustrates different actions of a multi-step attack which are interacting with each other.

7. Conclusion

Current advanced cyber-attacks utilise multiple steps and affect several devices. Detecting these multi-step attacks in ICS networks is a growing concern that has not been sufficiently investigated in research papers. In this paper, a two-layer analysis called ICS-CAD is proposed for multi-step attack detection in ICS networks. The first layer was an attack detector which provided high accuracy in the detection of attack types in two real-world ICS datasets. The results were compared with other studies. Then, causality detection was introduced in the second layer, and it could detect the causal impact of the detected attacks in both ICS datasets. Using this causality analysis, we could identify the events which are affected by the detected attack.

The future work of this research is working on the attacks which act in different cyber-kill-chain phases and investigating the performance of causality analysis in the detection of these types of attacks. In addition, unsupervised learning will be investigated in our future work to make our solution applicable to unknown scenarios and unlabelled datasets.

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.

Acknowledgement

The authors acknowledge the support of the Commonwealth of Australia and Cyber-security Research Centre Limited.

References

Abdelaty, M.F., Doriguzzi-Corin, R., Siracusa, D., 2021. DAICS: a deep learning solution for anomaly detection in industrial control systems. IEEE Trans. Emerg. Top. Comput. 10 (2), 1117–1129.

Alex, M., Tariq, U., Al-Shargie, F., Mir, H.S., AlNashash, H., 2020. Discrimination of genuine and acted emotional expressions using EEG signal and machine learning. IEEE Access 8, 191080–191089.

Alshamrani, A., Myneni, S., Chowdhary, A., Huang, D., 2019. A survey on advanced persistent threats: techniques, solutions, challenges, and research opportunities. IEEE Commun. Surv. Tutor. 21 (2), 1851–1877.

AnonMississippi State University Critical Infrastructure Protection Center, 2022 Industrial Control System Cyber Attack Data Set,. [Online]. Available: ⟨https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets⟩.

Chakraborty, S., Shah, S., Soltani, K., Swigart, A., 2019. Root cause detection among anomalous time series using temporal state alignment. Proceedings of the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). IEEE,, pp. 523–528.

Chalapathy, R., Chawla, S., 2019. Deep learning for anomaly detection: A survey. arXiv: http://arXiv.org/abs/arXiv:1901.03407.

Chandra, J.V., Challa, N., Pasupuleti, S.K., 2016. A practical approach to E-mail spam filters to protect data from advanced persistent threat. Proceedings of the 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT. IEEE, pp. 1–5.

Elmrabit, N., Zhou, F., Li, F., Zhou, H., 2020. Evaluation of machine learning algorithms for anomaly detection. Proceedings of the 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security). IEEE, pp. 1–8.

Farrukh, Y.A., Ahmad, Z., Khan, I., Elavarasan, R.M., 2021. A sequential supervised machine learning approach for cyber attack detection in a smart grid system. Proceedings of the 2021 North American Power Symposium (NAPS). IEEE, pp. 1–6 (November).

Feng, C., Li, T., Chana, D., 2017. Multi-level anomaly detection in industrial control systems via package signatures and LSTM networks. Proceedings of the 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). IEEE, pp. 261–272.

Ghafir, I., Hammoudeh, M., Prenosil, V., Han, L., Hegarty, R., Rabie, K., Aparicio-Navarro, F.J., 2018. Detection of advanced persistent threat using machine-learning correlation analysis. Future Gener. Comput. Syst. 89, 349–359.

Goh, J., Adepu, S., Junejo, K.N., Mathur, A., 2016. A dataset to support research in the design of secure water treatment systems. Proceedings of the International conference on critical information infrastructures security. Springer, Cham, pp. 88–99.

Gómez, Á.L.P., Maimó, L.F., Celdran, A.H., Clemente, F.J.G., Sarmiento, C.C., Masa, C.J. D.C., Nistal, R.M., 2019. On the generation of anomaly detection datasets in industrial control systems. IEEE Access 7, 177460–177473.

Harikrishnan, V.N., Kumar, G.T., 2018. Advanced persistent threat analysis using Splunk. Int. J. Pure Appl. Math. 118 (20), 3761–3768.

Haylett, G., Jadidi, Z., Thanh, K.N., 2021. System-wide anomaly detection of industrial control systems via deep learning and correlation analysis. Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, Cham, pp. 362–373.

Jadidi, Z., Lu, Y., 2021. A threat hunting framework for industrial control systems. IEEE Access 9, 164118–164130.

Jadidi, Z., Foo, E., Hussain, M., Fidge, C., 2021. Automated detection-in-depth in industrial control systems. Int. J. Adv. Manuf. Technol. 1–13.

Jadidi, Z., Dorri, A., Jurdak, R., Fidge, C., 2020. Securing manufacturing using blockchain. arXiv: http://arXiv.org/abs/arXiv:2010.07493.

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.E., HeGuelton, L., Caelen, O., 2018. Sequence classification for credit-card fraud detection. Expert Syst. Appl. 100, 234–245.

Khosravi, M., Ladani, B.T., 2020. Alerts correlation and causal analysis for APT based cyber attack detection. IEEE Access 8, 162642–162656.

Kravchik, M., Shabtai, A., 2018. Detecting cyber-attacks in industrial control systems using convolutional neural networks. In Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and Privacy. pp. 72–83.

Li, Q., Xu, B., Li, S., Liu, Y., Xie, X., 2021. Nonlinear Granger causality graph method for data-driven target attack in power cyber-physical systems. Trans. Inst. Meas. Control 43 (3), 549–566.

Liu, Y., et al., 2018. Towards a timely causality analysis for enterprise security. In NDSS.

Lughofer, E., Zavoianu, A.C., Pollak, R., Pratama, M., Meyer-Heye, P., Zörrer, H., Radauer, T., 2020. On-line anomaly detection with advanced independent component analysis of multi-variate residual signals from causal relation networks. Inf. Sci. 537, 425–451.

Luo, Y., Xiao, Y., Cheng, L., Peng, G., Yao, D., 2021. Deep learning-based anomaly detection in cyber-physical systems: progress and opportunities. ACM Comput. Surv. 54 (5), 1–36.

Mao, X., Yang, A.C., Peng, C.K., Shang, P., 2020. Analysis of economic growth fluctuations based on EEMD and causal decomposition. Phys. A: Stat. Mech. Appl. 553, 124661.

Milajerdi, S.M., Gjomemo, R., Eshete, B., Sekar, R., Venkatakrishnan, V.N., 2019. Holmes: real-time apt detection through correlation of suspicious information flows. Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP). IEEE, pp. 1137–1152.

Navarro, J., Deruyver, A., Parrend, P., 2018. A systematic survey on multi-step attack detection. Comput. Secur. 76, 214–249.

Pan, S., Morris, T., Adhikari, U., 2015. Developing a hybrid intrusion detection system using data mining for power systems. IEEE Trans. Smart Grid 6 (6), 3104–3113.

Ray, I., Zhu, Q., Haney, M., 2019. In: Rieger, C. (Ed.), Industrial Control Systems Security and Resiliency. Springer, pp. 1–276.

Sapkota, S., Mehdy, A.K.M., Reese, S., Mehrpouyan, H., 2020. FALCON: framework for anomaly detection in industrial control systems. Electronics 9 (8), 1192.

Shi, D., Guo, Z., Johansson, K.H., Shi, L., 2017. Causality countermeasures for anomaly detection in cyber-physical systems. IEEE Trans. Autom. Control 63 (2), 386–401.

Tuptuk, N., Hailes, S., 2018. Security of smart manufacturing systems. J. Manuf. Syst. 47, 93–106.

Xia, Z., Yi, P., Liu, Y., Jiang, B., Wang, W., Zhu, T., 2019. GENPass: a multi-source deep learning model for password guessing. IEEE Trans. Multimed. 22 (5), 1323–1332.

Yang, A.C., Peng, C.K., Huang, N.E., 2018. Causal decomposition in the mutual causation system. Nat. Commun. 9 (1), 1–10.

Zhang, R., Cao, Z., Wu, K., 2020. Tracing and detection of ICS anomalies based on causality mutations. Proceedings of the 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE, pp. 511–517.