BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Computer Science Research
https://ojs.bilpublishing.com/index.php/jcsr

ARTICLE

Intrusion Detection through DCSYS Propagation Compared to Auto-encoders

Fatima Isiaka[1*] Zainab Adamu[2]

1. Department of Computer Science, Nasarawa State University, Keffi, Nigeria
2. Department of Computer Science, Ahmadu Bello University, Zaria, Nigeria

ARTICLE INFO

Article history
Received: 16 August 2021
Accepted: 23 August 2021
Published Online: 26 August 2021

Keywords:
Dynamic control system
Deep learning
Artificial neural network
Auto-encoders
Identify space model
Benign
Anomalies

ABSTRACT

In network settings, one of the major disadvantages that threaten the network protocols is the insecurity. In most cases, unscrupulous people or bad actors can access information through unsecured connections by planting software or what we call malicious software otherwise anomalies. The presence of anomalies is also one of the disadvantages, internet users are constantly plagued by virus on their system and get activated when a harmless link is clicked on, this a case of true benign detected as false. Deep learning is very adept at dealing with such cases, but sometimes it has its own faults when dealing benign cases. Here we tend to adopt a dynamic control system (DCSYS) that addresses data packets based on benign scenario to truly report on false benign and exclude anomalies. Its performance is compared with artificial neural network auto-encoders to define its predictive power. Results show that though physical systems can adapt securely, it can be used for network data packets to identify true benign cases.

1. Introduction

This paper contains an introductory viewpoint to network security, different threats to a network setting, deep learning and the design of different autonomous systems for network security. The goal is to approach this in a simplified form, by first taking a look at the different forms of threats, understanding deep learning and using deep learning in development of safe and autonomous systems for a safe network. This section contains an overview of network security, threats and deep learning with Auto encoders and an overview of Dynamic Control System (DCSYS).

One of the major disadvantages of a network setting is the insecurity. If the internet can be used for online banking, social networking and other services, one may risk a theft to personal information such as name, credit card information, address e.t.c. Unscrupulous people or bad actors can access these information through unsecured connections by planting software or what we call malicious software and use personal details for their benefit. The presence of anomalies is one of the disadvantages. Internet users are constantly plagued by virus on their system and get deactivated when a harmless link is clicked on (true benign case). Most computers connected to internet are always very prone to targeted virus attacks and they often end up getting crashed Unfortunately, the ability to send and receive mails created a means for cyber criminals

**Corresponding Author:*
Fatima Isiaka,
Department of Computer Science, Nasarawa State University, Keffi, Nigeria;
Email: fatima.isiaka@outlook.com

DOI: https://doi.org/10.30564/jcsr.v3i3.3578

to off-load anomalies. Some malware attached to emails could wreak havoc to computers and create backdoor for attackers to infiltrate the network system. The attackers can lure victims into disclosing sensitive information through techniques like phishing scams. The on-going evolvement and introduction of new and advanced threats to network systems have called for the need to build autonomous, safe and interactive systems that can counteract unauthorised access, attacks and access true benign case without false alert. To address this, the paper seeks to discourse the following objectives:

- address cases of false positive in an actual benign scenario.
- reduce the occurrence of advanced threats.
- create a dynamic control intrusion detection system (DCSYS).
- compare its performance with ANN auto-encoders.

These systems will be as a form of interactive tool in a controlled environment that can filter out anomalies, virus intrusion and reproduce interceptions protocols that identifies true benign cases.

1.1 Auto Encoders

In an intrusion detection system (IDS), one of its most positive rule is being able to identify when the configuration has changed or when some network traffic indicates a problem such as when capital one data has been breached (Figure 1). Artificial neural network (ANN) is one of the most faultless IDS as a Deeping learning method and a typical auto-encoder (AE) is a type of ANN used to learn efficient data coding in an unsupervised tender manner [1-4]. Its aim is to learn a representation or encoding for a set of prime data, which is typically for dimensional reduction. Training the network to ignore outlying cases is one its characteristics. Also the reconstruction side is learned, here the auto-encoder generates from the reduced encoding by representing its original input [5-8]. The AE are applied to a lot of problems from facial recognition to acquiring the class semantic meaning and representation of words. It can serve as form of feedback loop for network security data analysis [10,11] for pre-served browsing activities online. One of its equivalent is the variational encoder, a model based on ANN that provides probabilistic manner for describing an observation in a constant latent space. Ultimately, rather than building an encoder which outputs a single value to describe each latent state features, an encoding characteristics can be formulated that describe a probability distribution for each latent class of attributes. Figure 2 shows a typical network architecture for an auto-encoder with μ and σ as the latent class of attributes that produces new set of classes . The variational auto-encoder is a neural network where the middle layer of the network is made of mean and standard deviation that are sampled from the normal distribution derived from input parameters. In AE and VAE all layers use state-of-the-art convolutional neural networks which can be made explicit with inclusion of hyper-parameters as shown in the diagrams below for both AE and VAE.

Despite the deep learning applications to IDS, there are some of its drawbacks such as identifying true benign cases which are sometimes detected as false positive. Here we tend to apply a dynamic control system that can deal with such case and compare its performance to auto-encoders.

Figure 1. Capital one data breach (Curtsey: Google image)

 DOI: https://doi.org/10.30564/jcsr.v3i3.3578

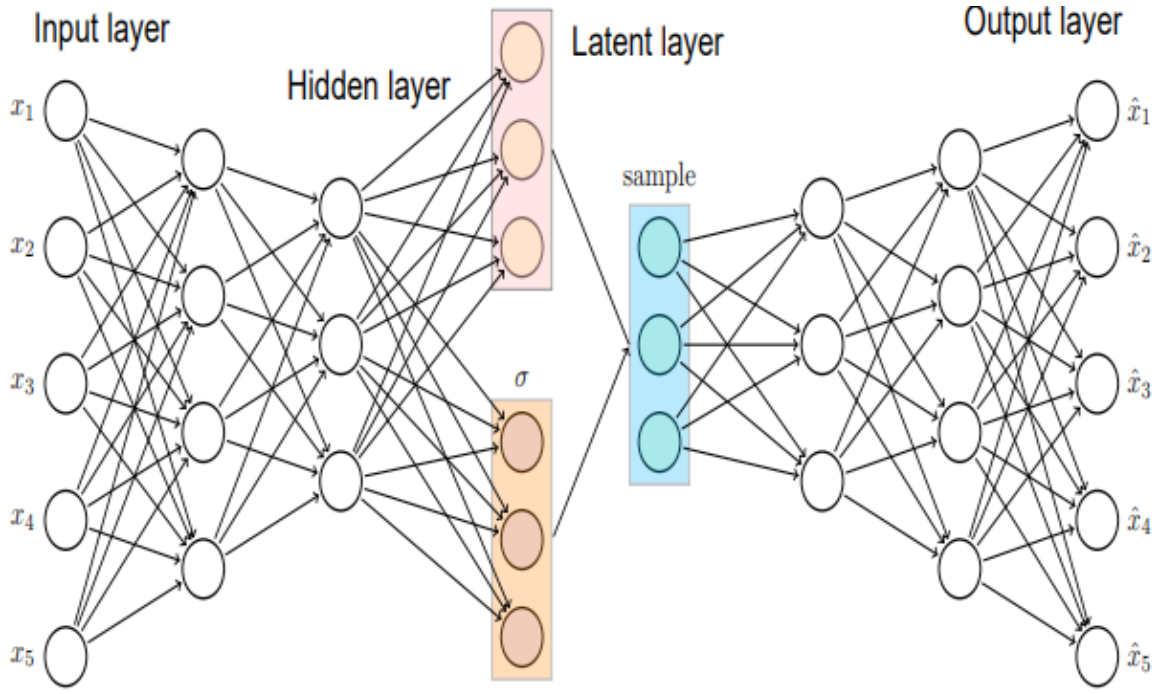


Figure 2. A High-level illustration of AE with hidden and latent layers.

1.2 Dynamic Control Systems

Dynamic control systems are based on discrete time-variant computation which are well known for dealing with physical systems that result in perfect detected signals [11-13]. A dynamic control system manages the command, direction, and regulates the behaviour of other systems using its control loops. It can range from a controller used for controlling processes or machines and powered systems.

The models are comprised of a linear feedback systems, a control loop which includes control algorithms and actuators that attempt to regulate variable set-point (SP). An example is a PID controller algorithm [12,9,10] that controls and restores an actual instance of speed process to the desired speed in an optimum level, with minimal delay or false alert, by controlling the predicted power output of an engine controlling process.

1.3 The Proposed DCSYS Model

The setup function allows for the input of data not limited to the CICA dataset but for a generalised scenario. Predictions are made based on the benign and non-benign cases from an identity state space model set, used as input.

Here we intend to use the control model's process formation to help fish-out false benign cases and comparing its predictive ability with auto-encoders. A window-based control panel was designed as a standalone, solely to address true benign case in the Canadian Institute for Network Security (CICA) data.

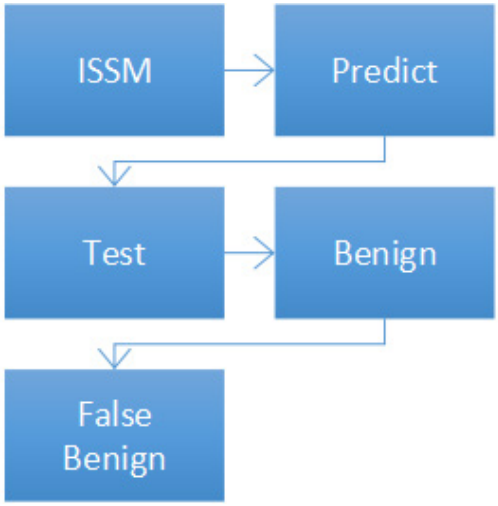


Figure 3. Model Kernel for DCSYS.

2. Method and Dataset

The CICA dataset (Figure 3) contains benign cases that leaves room to address false positive, the "Benign" cases were treated as a single entity to differentiate between "false benign" and "true benign" case. The DCSYS model is encapsulated has an identified state space model given as:

$$\frac{dx}{dy}(t) = Ax(t) + Bu(t) + K\,e(t)$$
$$y(t) = Cx(t) + Du(t) + e(t) \quad (1)$$

where are the state space converted matrices, *u(t)* is the

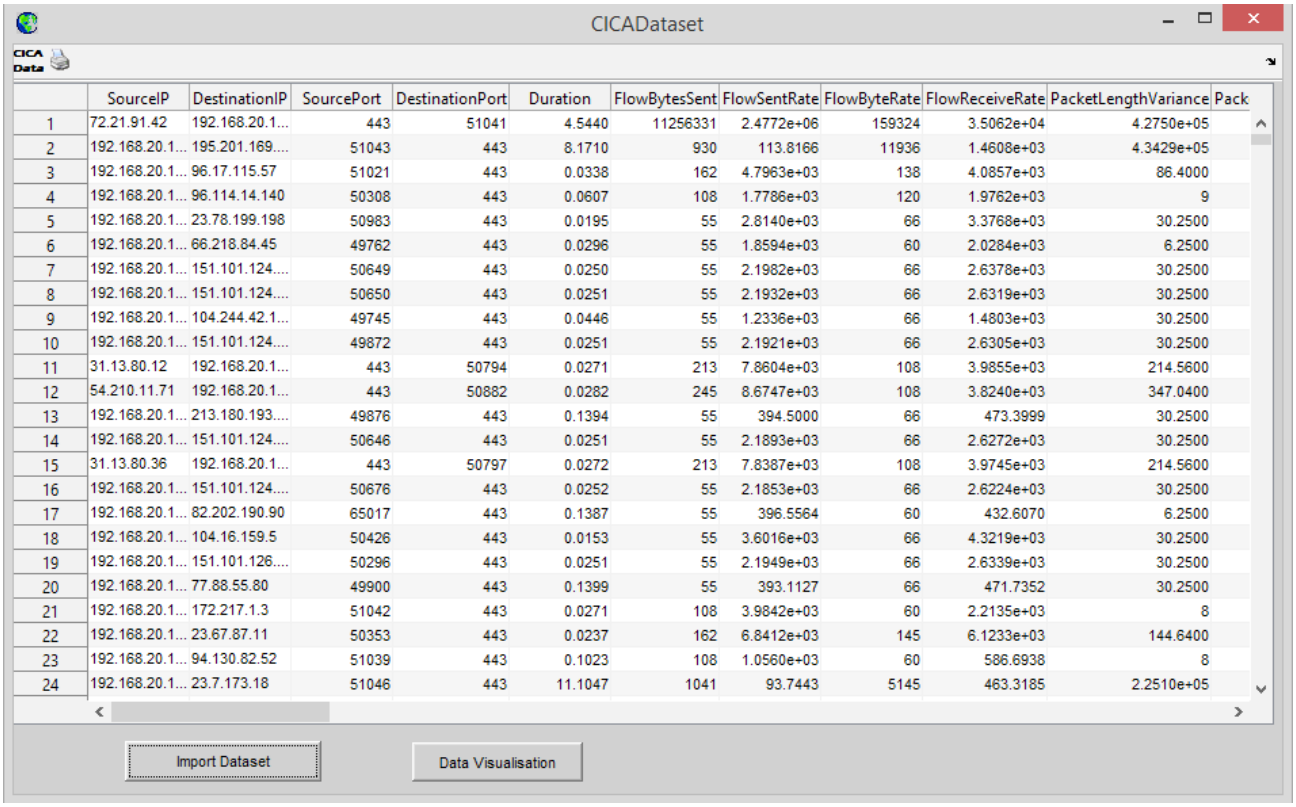
DOI: https://doi.org/10.30564/jcsr.v3i3.3578

input attributes or data packet features, *y(t)* is the output, *e(t)* denotes the disturbance and *x(t)* the vector matrix of signal propagation. All state space matrices are considered free parameters. The matrix is set to zero value, which means no feed-through by default except for the static systems. The default behaviour of the system is modified by feed-through and disturbance pairs. The DCSYS is modified using an interactive form that inputs datasets for visualisation and analysis.

The CICA dataset is a time series data transformed into an identified data series, containing a time-domain output signal *y* and an empty input signal respectively. The sampling interval is specified using an arbitrary constant parameter for the dataset. The output data by default have the same domain as the input data.

The DCSYS takes in data packet and generates an alert (Figure 5) whenever a "false benign" is detected, this report is generated from the denial of serves features created from the data packet. The corresponding source port and destination port is then identified and relays options for either blocking or deleting the data packet associated to these addresses. All cases are treated as benign on till a false case is encountered. The identified matrix created from the data packet is then run through the DCSYS (Figure 6) which generates report on the discrete time identified state space model (MDL) and the ROC or performance of the detection session, this is to authenticate the genuinety of the report and the area under the curve (AUC) transmits the models performance. Bode reports show the frequency response, magnitude, phase of frequency response and the predictive power of each feature attributed to the CICA data, this leaves room for which attribute contributes to the predictive ability of the system or which attribute is impassive. From the Figure 5, the data attribute "*Packet Length Median*" to the "*Source Port*" seem to contribute most significantly to the models predictive power that addresses false benign cases. The system can also generate report on whether exclusion of anomalies and attributes is pre-eminent for the model.

To generalise the predictive performance of the model, a plain comparative analysis is demonstrated with ANN auto-encoders on different epochs (*2, 3, 4*) and *order 1-3* for the DCSYS model on four different scenarios (for anomalies removed, anomalies included, four authentic attributes with anomalies included, and dataset with all attributes). The proceeding the section discusses the results obtained from the analysis as compared to an auto-encoder.



	SourceIP	DestinationIP	SourcePort	DestinationPort	Duration	FlowBytesSent	FlowSentRate	FlowByteRate	FlowReceiveRate	PacketLengthVariance	Pack
1	72.21.91.42	192.168.20.1...	443	51041	4.5440	11256331	2.4772e+06	159324	3.5062e+04	4.2750e+05	
2	192.168.20.1...	195.201.169....	51043	443	8.1710	930	113.8166	11936	1.4608e+03	4.3429e+05	
3	192.168.20.1...	96.17.115.57	51021	443	0.0338	162	4.7963e+03	138	4.0857e+03	86.4000	
4	192.168.20.1...	96.114.14.140	50308	443	0.0607	108	1.7786e+03	120	1.9762e+03	9	
5	192.168.20.1...	23.78.199.198	50983	443	0.0195	55	2.8140e+03	66	3.3768e+03	30.2500	
6	192.168.20.1...	66.218.84.45	49762	443	0.0296	55	1.8594e+03	60	2.0284e+03	6.2500	
7	192.168.20.1...	151.101.124....	50649	443	0.0250	55	2.1982e+03	66	2.6378e+03	30.2500	
8	192.168.20.1...	151.101.124....	50650	443	0.0251	55	2.1932e+03	66	2.6319e+03	30.2500	
9	192.168.20.1...	104.244.42.1...	49745	443	0.0446	55	1.2336e+03	66	1.4803e+03	30.2500	
10	192.168.20.1...	151.101.124....	49872	443	0.0251	55	2.1921e+03	66	2.6305e+03	30.2500	
11	31.13.80.12	192.168.20.1...	443	50794	0.0271	213	7.8604e+03	108	3.9855e+03	214.5600	
12	54.210.11.71	192.168.20.1...	443	50882	0.0282	245	8.6747e+03	108	3.8240e+03	347.0400	
13	192.168.20.1...	213.180.193....	49876	443	0.1394	55	394.5000	66	473.3999	30.2500	
14	192.168.20.1...	151.101.124....	50646	443	0.0251	55	2.1893e+03	66	2.6272e+03	30.2500	
15	31.13.80.36	192.168.20.1...	443	50797	0.0272	213	7.8387e+03	108	3.9745e+03	214.5600	
16	192.168.20.1...	151.101.124....	50676	443	0.0252	55	2.1853e+03	66	2.6224e+03	30.2500	
17	192.168.20.1...	82.202.190.90	65017	443	0.1387	55	396.5564	60	432.6070	6.2500	
18	192.168.20.1...	104.16.159.5	50426	443	0.0153	55	3.6016e+03	66	4.3219e+03	30.2500	
19	192.168.20.1...	151.101.126....	50296	443	0.0251	55	2.1949e+03	66	2.6339e+03	30.2500	
20	192.168.20.1...	77.88.55.80	49900	443	0.1399	55	393.1127	66	471.7352	30.2500	
21	192.168.20.1...	172.217.1.3	51042	443	0.0271	108	3.9842e+03	60	2.2135e+03	8	
22	192.168.20.1...	23.67.87.11	50353	443	0.0237	162	6.8412e+03	145	6.1233e+03	144.6400	
23	192.168.20.1...	94.130.82.52	51039	443	0.1023	108	1.0560e+03	60	586.6938	8	
24	192.168.20.1...	23.7.173.18	51046	443	11.1047	1041	93.7443	5145	463.3185	2.2510e+05	

Figure 4. The Canadian Network Security Database

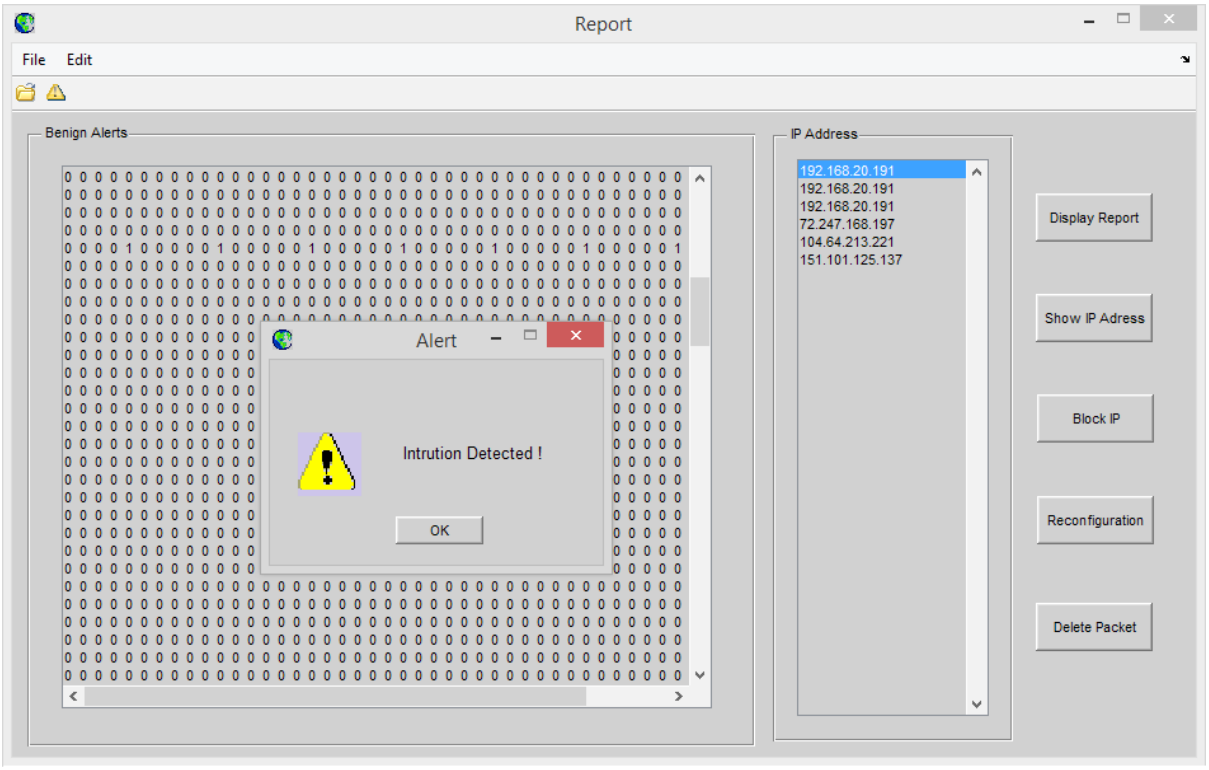


Figure 5. Data message report screen for data packets.

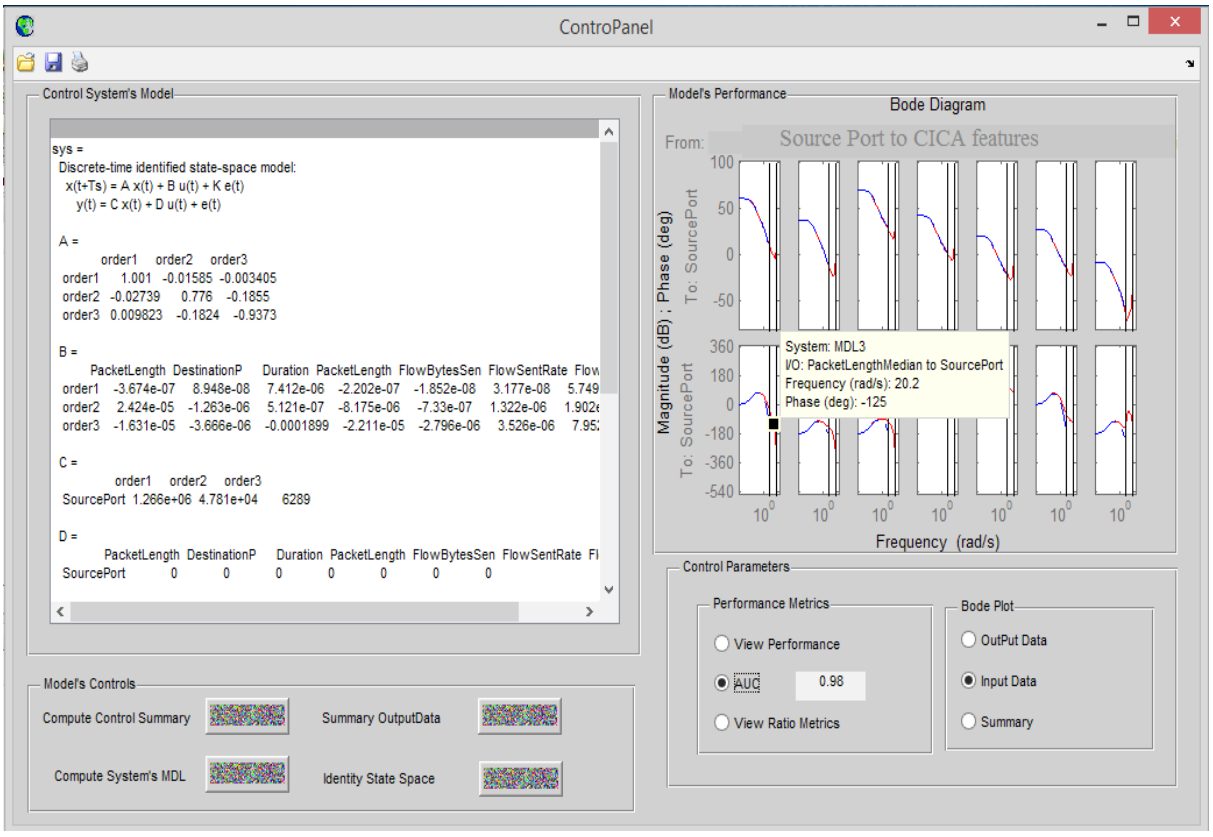


Figure 6. Control panel for report generator on the discrete time identified state space model (MDL).

3. Results

The Receiver Operating Characteristics (ROC) Curve comparison for DCSYS and Auto-encoders at different order and epochs for anomalies removed shows that DCSYS outperforms the auto-encoder with *0.90* accuracy at order one and *0.80* at order three (Figure 7). The auto encoder has an accuracy of *0.70.* This shows a close performance. As compared to anomalies included, the ROC curve comparison for DCSYS and Auto-encoders at different order and epochs shows a performance of *0.90* and *0.80* for the DCSYS that outperforms the auto-encoders (Figure 8). The ROC Curve Comparison for DCSYS and Auto-encoders at different order and epochs for four authentic attributes shows a performance of *0.80* for the auto-encoder at three epoch and very low performance for the

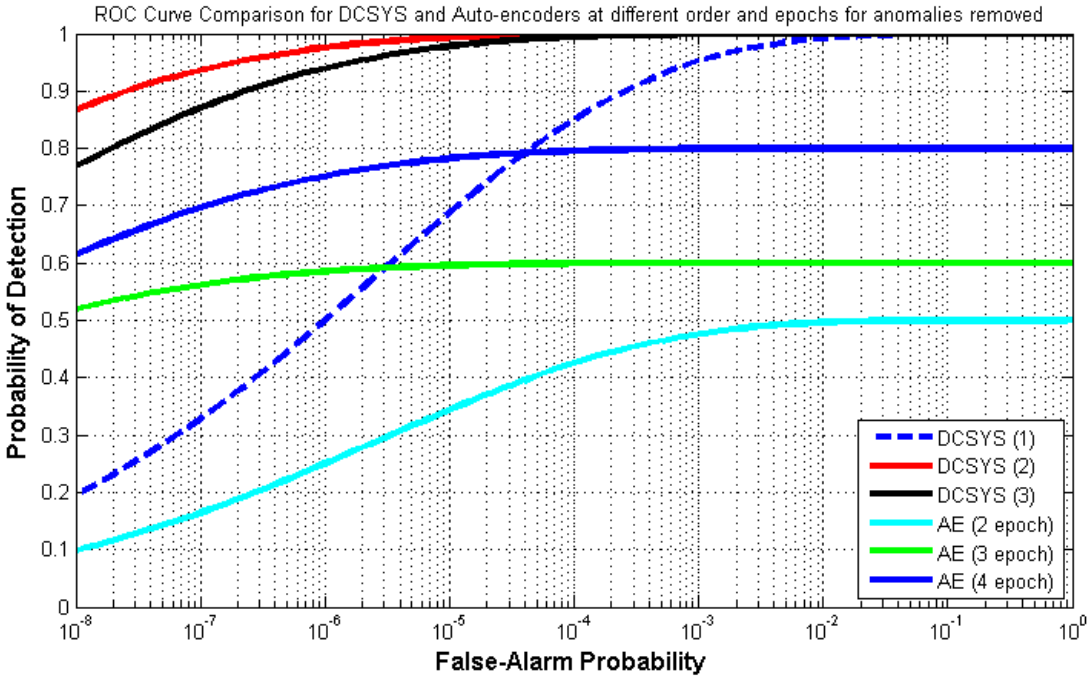


Figure 7. ROC Curve Comparison for DCSYS and Auto-encoders at different order and epochs for anomalies removed in Benign scenario 1.

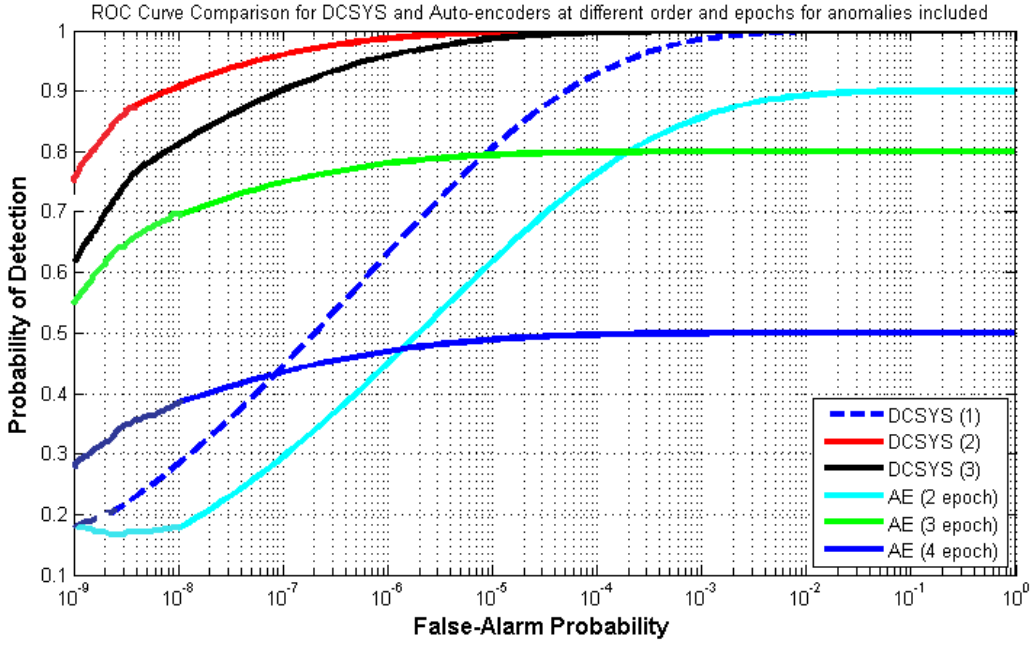


Figure 8. ROC Curve Comparison for DCSYS and Auto-encoders at different order and epochs for anomalies included in Benign scenario 2.

DCSYS (Figure 9), the auto-encoder's detection ability of false alarm is intensified for the case of authenticity of the CICA dataset's features as when compared to the control system's models predictive power. The ROC curve comparison for DCSYS and Auto-encoders at different order and epochs for dataset with all attributes included (Figure 10) shows that the DCSYS outperformances the auto-encoders. The overall predictive performance is in favour of

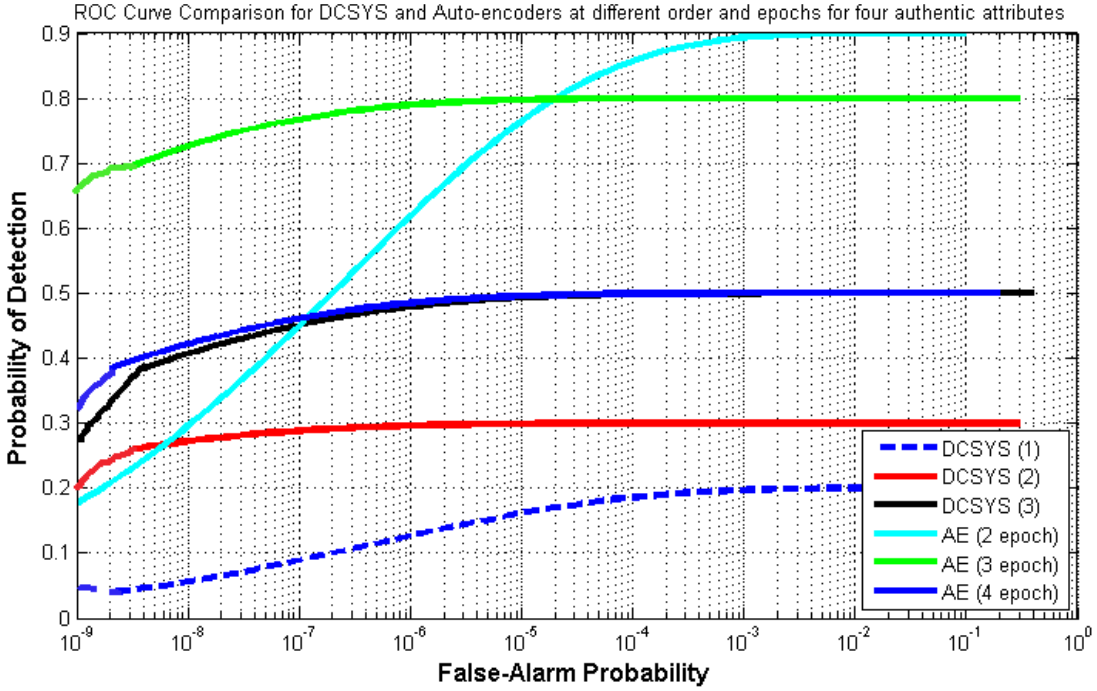


Figure 9. ROC Curve Comparison for DCSYS and Auto-encoders at different order and epochs for four authentic attributes in Benign scenario 3.

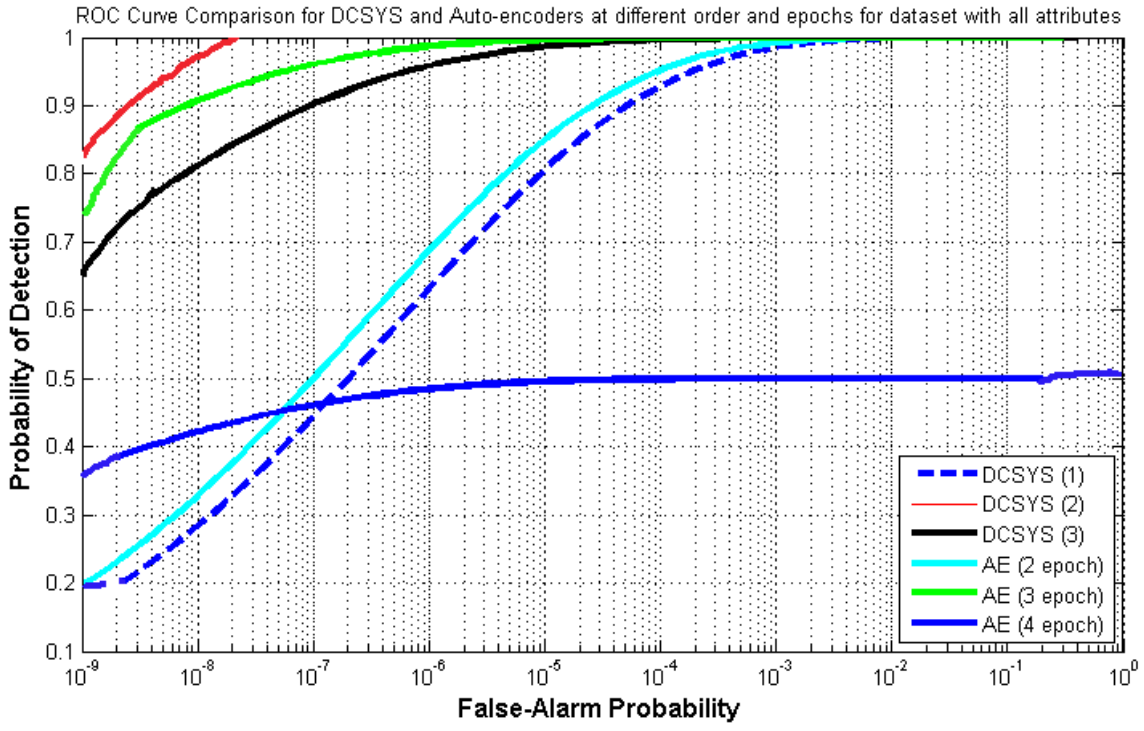


Figure 10. ROC Curve Comparison for DCSYS and Auto-encoders at different order and epochs for datasets with all the attributes in Benign scenario 4.

DOI: https://doi.org/10.30564/jcsr.v3i3.3578

the DCSYS in the case of redundant features.

4. Conclusions

This paper seeks to investigate IDS with DCSYS propagation in a network setting that involves a CICA data packet. As discussed in the paper, one of the major disadvantages that threaten the network protocols is the insecurity. In most cases, bad actors can access information through unsecured connections by planting software or what we call malicious software otherwise identified as anomalies which can be presented as false benign in a benign oriented scenario. The presence of these anomalies is one of the disadvantages. Internet users are constantly plagued by this false benign cases on their system and get activated when a harmless link is clicked on. Deep learning is very adept at dealing with such cases, but sometimes it has its own fault when dealing with benign cases. The paper tends to adopt a dynamic control system (DCSYS) that addresses data packets based on benign scenario to truly report on false benign and exclude anomalies. To define its predictive ability, its performance is compared with an auto-encoder on different epochs. Results show that performance of 0.9 above is quite adaptive enough to validate its predictive ability. Though physical systems can adjust securely, it can be used for network data packets to identify true benign cases. Future work is to further compare its performance with both recurrent and convolutional neural network and improve its performance metrics as a standalone tool that could be embed in an IDS.

References

[1] Phan, N., Wang, Y., Wu, X., & Dou, D. (2016, February). Differential privacy preservation for deep auto-encoders: an application of human behavior prediction. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[2] Kipf, T. N., & Welling, M. (2016). Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*.

[3] Soui, M., Smiti, S., Mkaouer, M. W., & Ejbali, R. (2020). Bankruptcy prediction using stacked auto-encoders. *Applied Artificial Intelligence*, *34*(1), 80-100.

[4] Ding, Y., Tian, L. P., Lei, X., Liao, B., & Wu, F. X. (2021). Variational graph auto-encoders for miRNA-disease association prediction. *Methods*, *192*, 25-34.

[5] Wang, W., & Gómez-Bombarelli, R. (2019). Coarse-graining auto-encoders for molecular dynamics. *npj Computational Materials*, *5*(1), 1-9.

[6] Wu, X., & Cheng, Q. (2021). Deepened Graph Auto-Encoders Help Stabilize and Enhance Link Prediction. *arXiv preprint arXiv:2103.11414*.

[7] Wu, X., & Cheng, Q. (2021). Deepened Graph Auto-Encoders Help Stabilize and Enhance Link Prediction. *arXiv preprint arXiv:2103.11414*.

[8] Gomede, E., de Barros, R. M., & de Souza Mendes, L. (2021). Deep auto encoders to adaptive e-learning recommender system. *Computers and Education: Artificial Intelligence*, *2*, 100009.

[9] Silva, A. B. O. V., & Spinosa, E. J. (2021). Graph Convolutional Auto-Encoders for predicting novel lncRNA-Disease associations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

[10] Arifoglu, D., Wang, Y., & Bouchachia, A. (2021). Detection of Dementia-Related Abnormal Behaviour Using Recursive Auto-Encoders. *Sensors*, *21*(1), 260.

[11] Zino, L., & Cao, M. (2021). Analysis, prediction, and control of epidemics: A survey from scalar to dynamic network models. *arXiv preprint arXiv:2103.00181*.

[12] Zhou, P., Chen, W., Yi, C., Jiang, Z., Yang, T., & Chai, T. (2021). Fast just-in-time-learning recursive multi-output LSSVR for quality prediction and control of multivariable dynamic systems. *Engineering Applications of Artificial Intelligence*, *100*, 104168.

[13] Sun, C., Chen, J., Cao, S., Gao, X., Xia, G., Qi, C., & Wu, X. (2021). A Dynamic Control Strategy of District Heating Substations Based on Online Prediction and Indoor Temperature Feedback. *Energy*, 121228.

 DOI: https://doi.org/10.30564/jcsr.v3i3.3578