Improving Data Protection in Industrial Control System Networks Using Machine Learning Technique

JUDE CHINEDU AKAMADU[1], PROF. JAMES EKE[2], EMETU CHUKWUMA KALU[3]
[1, 2, 3] *Department of Electrical and Electronic Engineering Faculty of Engineering Enugu State University of Science and Technology (ESUT) Enugu, Nigeria*

Abstract- ***This work was embarked on to combat the security vulnerabilities trending on the present-day Industrial Control Systems (ICS) Supervisory Control and Data Acquisition (SCADA) system infrastructure. This was observed after series of literatures were discussed and research gap was established. The study proposed to solve the problem using machine learning. Data of the ICS Denial of Service Attack was collected and then develop a neural network-based algorithm with it to detect threat on ICS. The training was done using back propagation algorithm. The system was implemented using Mathlab and neural network toolbox. The model was simulated and the result showed good threat detection performance with regression value of 0.973 and detection response time of 12ms which is very good. The percentage improvement when compared with the characterized test bed is 13.3% which is very good.***

Indexed Terms- ***Data protection, control system networks, and machine learning***

I. INTRODUCTION

In today's competitive market domain, companies have demand to improve their process efficiency using a low economic cost industrial automation to optimize productivity. Here the collaborative integrated system performs a big role with the functions of self-organization, rapid deployment, flexibility, and inherent intelligent-processing (Bailey et al., 2013).

In the industrial control system (ICS), the operators can perform remote control of various plants such as the continuous stir tank reactor, water treatment plant, power plant, nuclear reactor, biological reactor among others. This remote monitoring and control process is achieved using communication protocols between the human interface machine, sensors, actuators, controllers and Supervisory Control And Data Acquisition (SCADA) system. Historically this SCADA components have dedicated and private networks. Recently the increase in the size and complexity of the modern day industrial facilities has resulted to wide range of remote management, monitoring and supervision of these equipment through an open network called the internet. This exposes the SCADA systems to various forms of network threats and internet attack (Rinaldi et al., 2011).

Industrial control system securities have been handled traditionally using the conventional information technology (IT) security practices such as the use of network authentication codes and low level encryption formats. However, the goal of IT security techniques cannot cope with the high monitoring, supervision and physical complex industrial components that make up the SCADA network of an industry. This is to say that in IT while an electronic mailing system can afford short delay without any effect. for an industrial control system, a slight delay can result to devastating effect and environmental hazards, financial losses and even loss of life (Erickson et al., 2019).

Machine learning and artificial intelligence techniques have been employed widely to combat these challenges of industrial control security, thus developing intelligent intrusion detection systems using various techniques which will be discussed in the literature review. However, the system accuracy, rate of threat detection, speed of threat detection, system reliability and future threat prediction response abilities have all been called into question recently, due to the dynamic nature of the modern form of industrial threat.

Therefore, this research will employ a hybrid machine learning techniques, which will be trained using a

more advanced recent threat dataset to develop an improved real time threat detection and prediction system on an industrial control system.

- Problem Statement

In (Peerenboom, 2015), the SCADA system that controlled a Queensland sewage treatment plant in England was accessed and controlled by a former employee of the software development department, after he was sacked. The impact resulted to about 800,000 liters of raw waste being pumped into a nearby river, which resulted to the loss of aquatic habitats and huge financial loss.

- Aim and Objectives of the Study

The aim of this research work is the improvement of data protection in industrial control system networks using machine learning techniques, with the following setout objectives;

i. To perform systematic review and establish research gap
ii. To characterize an industrial SCADA system under various network conditions
iii. To develop a model of the industrial control system (ICS) network transfer function
iv. To train a machine learning model with DDos() attack to secure the ICS network
v. To implement the model with Matlab and evaluate the performance

- Significance of the Study

a) To improve the security of industrial control network infrastructures
b) To protect Industrial control systems against distributive denial of service attack
c) To back off network threats and hackers intelligently
d) To ensure that high security alert is achieved against external ICS network threat

- Scope of the Study

This work improves data protection in industrial control system using machine learning technique. The data will be protected against distributive denial of service attack using DDos training dataset, implemented at D-hap Integrated Industrial Service Limited.

II. MATERIALS AND METHODS

Here we will discuss the materials and methods employed for the development of the proposed system. It will begin with the characterization of an industrial SCADA network under normal and abnormal condition, then a new system will be developed using the necessary engineering materials and models. This will be implemented using simulink and other necessary implementation tools.

The materials used for the system development includes;

- Controllers
- Process design equipments
- The communication modules
- Routers and switches
- The process plants to be protected

- Methods
- Characterization

This work characterizes D-hap integrated industrial service limited, located at Ituku Enugu. The characterization was done to evaluate the network performance of the case study ICS under operating condition and also when attacked with network threat. This data was collected through the company's demilitral zone (a section for third party collection of real time SCADA data).

- Description of the network structure

The SCADA network consists of a programmable logic controller (PLC), human machine interface (HMI), Check point 1200R security gateway software and a chemical plant model made of a sensor, actuator, orifice and induction motor.

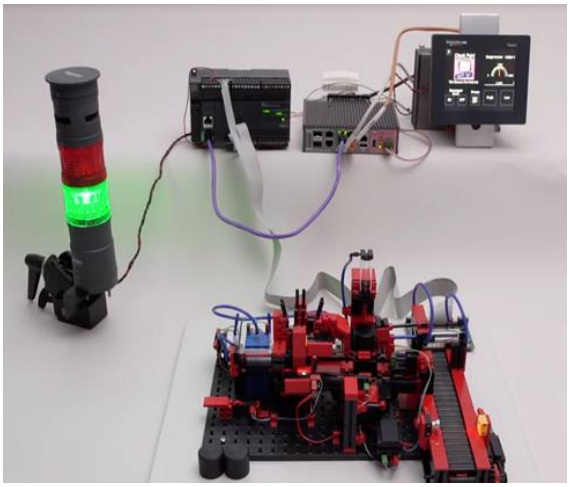

Figure 1. Model of the SCADA system
(Source: D-hap integrated service ltd.)

Table 1: data log of the ICS at normal condition

Data date and time	Flow (E3M 3)	Flow Time (hr)	Static Pressure (kPa)	Diff. Pressure (kPa)	Temperature (C)	Total Liquids (m3)	Orifice (in)
Mar 13 2020/ 08:00:01	30.27	24	780	78	15.5	7.68	0.88
Mar 12 2020/ 08:00:01	29.75	24	766.5	76.65	15.1	7.79	0.88
Mar 11 2020/ 08:00:00	29.54	24	761.25	76.12	15.1	7.63	0.88
Mar 10 2020/ 08:00:00	30.56	24	787.5	78.75	15.5	7.85	0.88
Mar 09 2020/ 08:00:01	30.18	24	777.75	77.78	15	7.67	0.88
Mar 08 2020/ 08:00:01	29.98	24	772.5	77.25	15.2	7.84	0.88
Mar 07 2020/ 08:00:01	29.1	24	750	75	15.5	7.57	0.88
Mar 06 2020/ 08:00:01	30.09	24	775.5	77.55	15.8	7.76	0.88
Mar 05 2020/ 08:00:01	29.54	24	761.25	76.12	15.4	7.86	0.88
Mar 04 2020/ 08:00:00	30.53	24	786.75	78.68	15.2	7.6	0.88
Mar 03 2020/ 08:00:01	30.36	24	782.25	78.22	15	7.73	0.88
Mar 02 2020/ 08:00:00	30.53	24	786.75	78.68	15.7	7.82	0.88
Mar 01 2020/ 08:00:02	30.3	24	780.75	78.08	15.4	7.82	0.88
Feb 29 2020/ 08:00:01	29.8	24	768	76.8	15.7	7.81	0.88
Feb 28 2020/ 08:00:01	29.31	24	755.25	75.53	15.1	7.5	0.88
Feb 27 2020/ 08:00:02	29.86	24	769.5	76.95	15.4	7.66	0.88
Feb 26 2020/ 08:00:04	29.28	24	754.5	75.45	15.5	7.54	0.88

(Source: D-hap Integrated Service Ltd.)

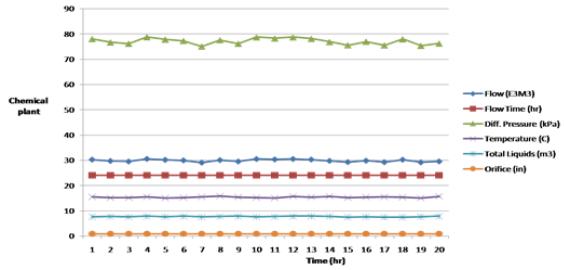

Figure 2. Performance of the plant at normal condition

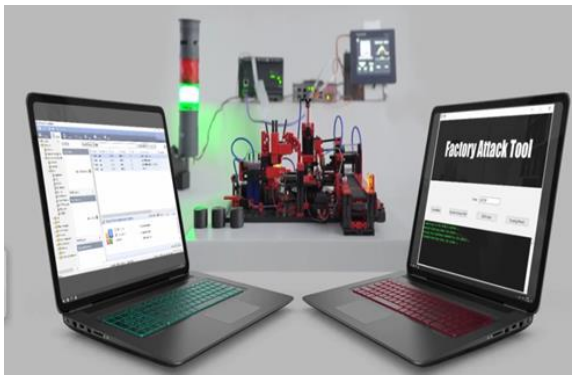

Figure 3. Model of the control system with attack tool (source: D-hap integrated service)

Table 2: data log of the ICS network under attack

Report time (hr)	Flow (E3M3)	Static Pressure (kPa)	Diff. Pressure (kPa)	Temperature (C)	Total Liquids (m3)	Orifice (in)
01.00	30.27	780	78	25.5	7.68	1.88
02.00	39.75	866.5	76.65	25.1	7.79	1.88
03.00	39.54	961.25	76.12	25.1	7.63	2.88
04.00	30.56	987.5	78.75	25.5	7.85	2.88
05.00	20.18	777.75	77.78	25.0	7.67	2.88
06.00	29.98	872.5	77.25	25.2	7.84	2.88
07.00	49.10	2750	75	25.5	7.57	4.88
08.00	30.09	1775.5	77.55	25.8	7.76	4.88
09.00	69.54	4761.25	76.12	55.4	7.86	7.88
01.00	60.53	4786.75	78.68	55.2	7.6	7.88
11.00	60.36	4782.25	78.22	55.0	7.73	7.88
12.00	60.33	4786.75	78.68	55.7	7.82	7.88
13.00	60.30	4780.75	78.08	55.4	7.82	7.88
14.00	69.80	3476.80	76.8	55.7	7.81	7.88
15.00	69.41	4755.25	75.53	55.1	7.5	7.88
16.00	69.86	4769.5	76.95	55.4	7.66	7.88
17.00	69.28	4754.5	75.45	55.5	7.54	7.88

(Source: D-hap integrated service ltd.)

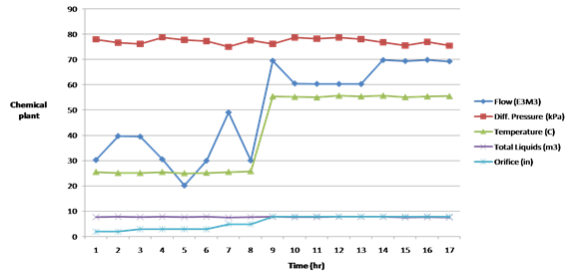

Figure 4. Plant performance under attack

- The proposed system

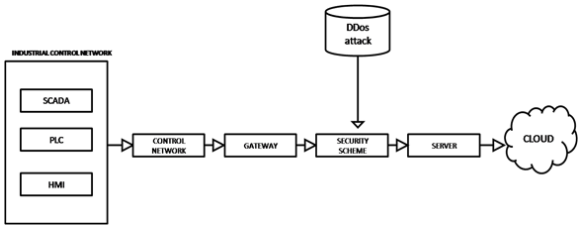

Figure 5. Block diagram of the proposed system

III. SYSTEM DESIGN

The proposed system will be designed using the following mathematic model

- Transfer function of the industrial control system
- ICS threat model
- Design of the ANN Security Scheme
- Nonlinear Auto Regressive Model
- The training model (back propagation algorithm)
- The classification model
- Prediction model.

Existing back propagation algorithm Flowchart

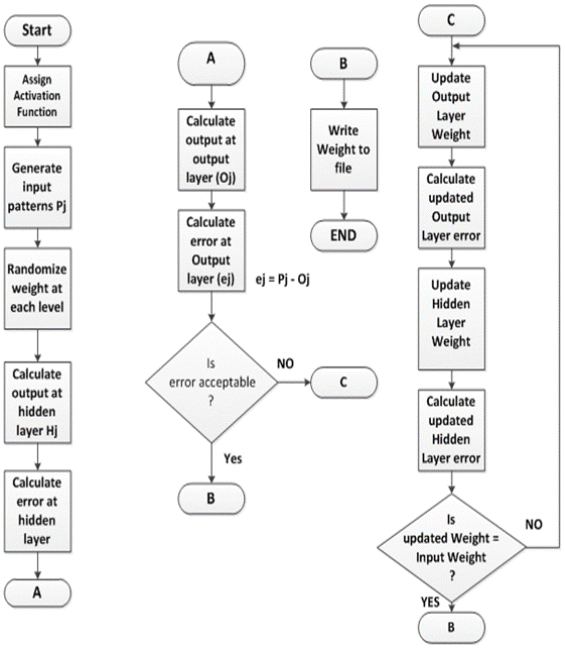

Figure 6: Flow chart for the existing back propagation algorithm

RESULTS AND DISCUSSIONS

Here we will discuss the results of the simulated Developed system. This result was based on the performance of the ANN technique used for the training and threat prediction process. The result was guided by the performance metrics presented in the neural network training tool and was used to evaluate the performance of the new system. Ten fold validation processes was used for the system validation and the average result was detailed as the overall system industrial threat detection regression.

RESULTS

The result was generated after the simulation of the neural network training tool based on the simulation parameters provided in table 1. The results were measured starting with the confusion matrix presented in figure 4

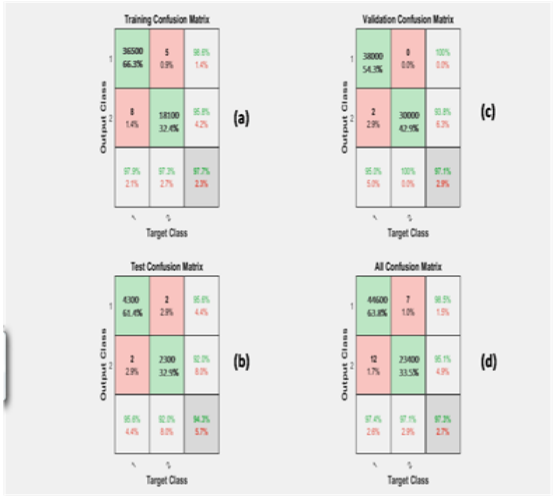

Figure 7: ANN training confusion matrix

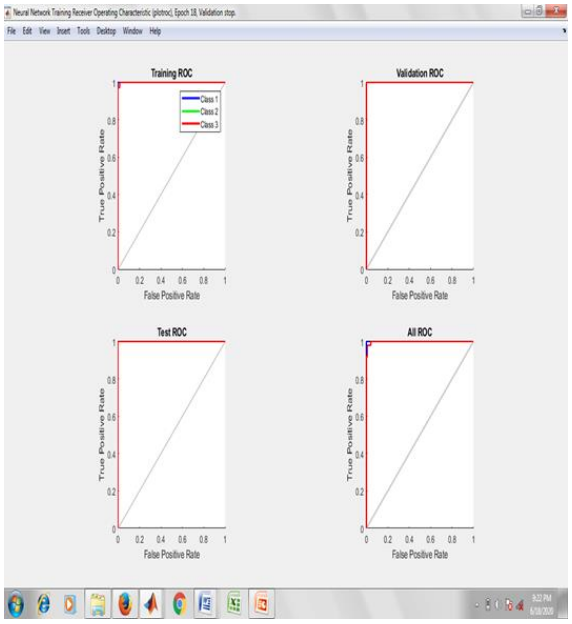

Figure 8: receiver operator characteristics result

Table 3: Results of the ROC curve

Dataset	False positive rate%	True positive rate%
Training	0.23	0.977
Testing	0.57	0.943
Validation	0.29	0.971
All	0.27	0.973
Overall training performance	0.27	0.973

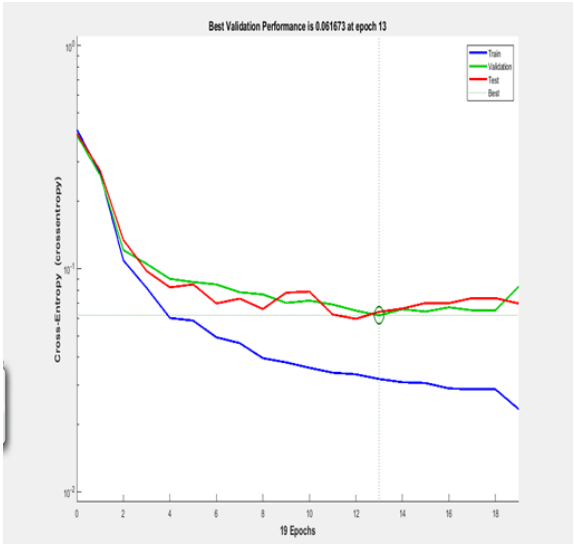

Figure 9: mean square error performance

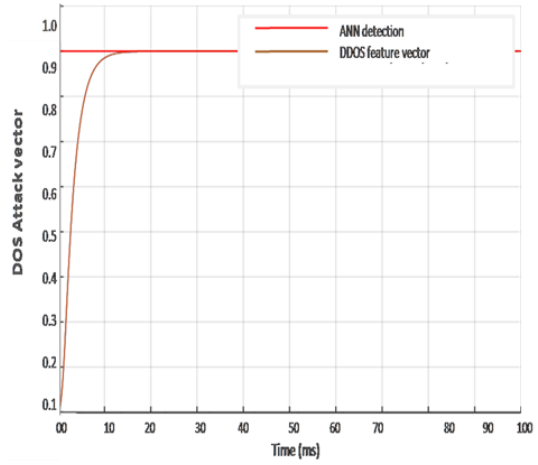

Figure 10: Response of the ANN to threat

- Discussions

This work has presented the performance of the ANN training techniques respectively. The performance of the training process has been evaluated using the receiver operator characteristics, confusion matrix and mean square error performance. This was done to determine the rate and accuracy of the ANN during threat detection and isolation from the industrial control network. From the result it was observed that the 80000 feature vectors of DDos threat which were used for the training process achieved were correctly classified with regression value of 0.973 this shows that the prediction rate of the threat attack is very accurate. The implication of this result is that the prediction accuracy based on any input DDos threat from hackers will be achieved at a precision value of 0.97. The mean square error performance was used also evaluates the training root mean square error. From the result obtained, the desired correlation between the training, test and validation datasets were observed with a best validation performance of 0.061673, this value justified the accuracy achieved using the confusion matrix.

- Fold Cross Validation

To validate the system performance, tenfold validation technique was adopted. This process iteratively trains the multi dataset at various times while the training performance is recorded and averagely computed to achieve validated response. This training process for the neural network will consider the root mean square error, precision value, true positive rate, false positive rate and accuracy based on the performance measured designed in equation 3.13 to 3.17; This result will be recorded below in the table 4;

Table 4: Tenfold validation performance

Iteration	FPR	TPR	MSE
1	0.009	0.991	0.0095612
2	0.042	0.958	0.0773790
3	0.021	0.979	0.0428200
4	0.076	0.924	0.0264560
5	0.023	0.977	0.0156900
6	0.057	0.963	0.0739220
7	0.011	0.989	0.0447392
8	0.056	0.944	0.0629877
9	0.076	0.974	0.0457021
10	0.030	0.970	0.0705542
Average	0.021	0.973	0.0356810

From the table 4; the false positive rate, true positive rate, accuracy and mean square error performance of the iteration processes are recorded as the validation tool of the neural network. The average performance of was achieved after series of simulation were done based on the validation model in equation 3.13. The iteration was averagely measured and the result shows a true positive result of 0.973, false positive of 0.021 based on a standard mean square error value of 0.0356810.

CONCLUSION AND RECOMMENDATIONS

Conclusion

Today industrial processing is using the digital data which are coming from various network providers and IT infrastructures, as a result there is risk of network threat as one cannot trust the source of the internet. Multi-dimensional methodologies have been applied overtime to combat this challenge of network insecurity in the control system industry.

Recommendations

- This work should be adopted and apply to improve the conventional ICS network
- The use of more threat datasets should be employed to train the ANN
- Improve training algorithm that is adaptive can be used for training the ANN for even more performance and response to threat.

Contributions to knowledge

- A neural network intelligent security scheme was developed of the industrial control network security
- High threat detection regression value of 0.973 was achieved
- The case study D-Hap integrated services was secured against DDOS threat

ACKNOWLEDGMENT

My sincere gratitude goes to God Almighty as I appreciate the efforts of my loving father and mother Mr and Mrs. Akamadu, my loving wife and kids. I wish to express my profound gratitude to all who in one way made contribution in this research work, mostly to my ever-ready supervisor in the person of Prof. James Eke, Prof. G.N Onoh, Prof. I.I Eneh, Dr. Abonyi, Dr. Alor and so many that time and space will not allow me to mention. All I have to say is thank you and God bless.

REFERENCES

[1] Almalawi, A, Yu, X, and Tari Z, (2014), "An unsupervised anomaly-based detection approach for integrity attacks on SCADA" systems. ComputSecur 46:94–110

[2] Asogwa, T.C and Ituma, C. (2018); "real time facial recognition system for digital crime management**"** International Journal of Computer Applications (0975 – 8887) Volume 94; No 12. Computer science; Ebonyi State, University, Abakaliki, Nigeria

[3] Bailey D, and Wright E. (2013), Practical SCADA for Industry, IDC Technologies, 2013.

[4] Berge J., (2012), Fieldbuses for Process Control: Engineering, Operation, and Maintenance, ISA, 2012.IEEE Control Systems Magazine, http://www.ce.cmu.edu/~hsm/im2004/readings/CIIRinaldi.pdf

[5] Boyer, Stuart, S. (2017) "SCADA Supervisory Control and Data Acquisition", 2nd Edition, ISA, 2017.

[6] Chukwuemeka O. A. (2016) "A model of hybrid agent software system for combating indigenous spam on GSM platform"; Ebonyi; International

Journal of Computer Trend and Technology (IJCTT)

[7] Erez N, Wool A (2015) Control variable classification, modeling and anomaly detection in Modbus / TCP SCADA systems. Int J CritInfrastructProt 10:59–70.

[8] Erickson, Kelvin, and Hedrick, John, Plant Wide Process Control, Wiley & Sons, 2019.

[9] Falco, Joe, et al., IT Security for Industrial Control Systems, NIST IR 6859, 2013, http://www.isd.mel.nist.gov/documents/falco/ITSecurityProcess.pdf.

[10] Frazer, 2011 Process Measurement and Control – Introduction to Sensors, Communication Adjustment, and Control, Prentice-Hall, Inc., 2011.

[11] Hadziosmanovic D, Sommer, R, Zambon, E, andHartel P (2014) "Through the Eye of the PLC: Towards Semantic Security Monitoring for Industrial Control Systems," ACSAC '14 Proceedings of the 30th Annual Computer Security Applications Conference

[12] Jiang, J, and Yasakethu, L (2013) Anomaly detection via one class SVM for protection of SCADA systems. Proc - 2013 IntConf Cyber-Enabled DistribComputKnowlDiscovCyberC 2013 82–88.

[13] Li, T., Fredrich, J., Fistruc, N. Pavefisic. and Y. Davies (2010) "The Complete Gabor-Fisher Classifier for Robust Face Recognition". EURASIP Journal on Advances in Signal Processing, vol.2, article ID 847680.

[14] Lakhina, A., Crovella, M., and Diot, C., (2004). Characterization of network-wide anomalies in traffic flows. In Proceedings of the 4th ACM SIGCOMM conference on Internet measurement, IMC '04, pages 201–206, New York, NY, USA, 2004. ACM.

[15] Linda, O (2019) Neural Network based Intrusion Detection System for critical infrastructures. International Joint Conference on Neural Networks.

[16] Maglara LA, and Jiang J (2014) "Intrusion detection in SCADA systems using machine learning techniques". 2014 Sci Inf Conf 626–631.

[17] Moya JM, Araujo Á, Banković Z, et al (2009) Improving security for SCADA sensor networks with reputation systems and self-organizing maps. Sensors 9:9380–9397.

[18] Nader, P (2013), "Intrusion detection in scada systems using one-class classification," roc 21st European Signal Processing Conference (EUSIPCO 2013) 1–5

[19] Nader P, Honeine P, Beauseroy P (2014) lp-norms in One-Class Classification for Intrusion Detection in SCADA Systems. Proc IEEE Transactions on Industrial Informatics, 10:4.

[20] Peerenboom, James (2015), Infrastructure Interdependencies: Overview of Concepts and Terminology, Argonne National Laboratory, http://www.pnwer.org/pris/peerenboom_pdf.pdf .

[21] Rinaldi, et al., Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies,

[22] Rrushi, J, and Kang KD (2009), "Detecting anomalies in process control networks". IFIP

[23] Russell, R, P., Sinha, B.and Y. Ostrovsky (2006), "Face Recognition by Humans": Nineteen Results All Computer Vision Researchers Should Know About. Proceedings of the IEEE, 94(11)

[24] Yang, D, Usynin A, A, and Hines, J (2005)," Anomaly-based intrusion detection for SCADA systems", 5th Intl Top Meet Nucl Plant Instrumentation, Control Hum Mach Interface Technol (NPIC&HMIT 05) 12–16

[25] Yasakethu SLP, Jiang J (2013) Intrusion Detection via Machine Learning for SCADA System Protection. 1st IntSymp ICS SCADA Cyber Secur Res 101–105

[26] Yong,zhoug (2013). Survey of techniques for face detection, recognition and feature extraction approaches"; International Journal of Multidisciplinary Educational Research; vol. (3), Nehal

[27] Yulmez, G. and kmen,G (2017), "Design and implementation of face recognition system using Face thermal detection hardware" ; 2nd International Conference on Advanced Signal,

Image and Video Processing; Spain. 2(3) ; pp 555-677; 1277-2343.

[28] Zahraa, Q J (2016) “Design and Implementation of Real Time Face Recognition System (RTFRS)”; International Journal of Computer Applications (0975 – 8887) Volume 94 – No 12. Computer Engineering Department, College of Engineering, University of Baghdad Al-Jadriyah, Baghdad, Iraq

[29] Zhang, L., Yu, S., Wu, D., and Watters, P., (2011). “A Survey on Latest Botnet Attack and Defense,” Proc. of 10th Intl’ Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), IEEE, pp. 53-60, November 2011.

[30] Zeechan, M. (2009); “Face recognition using PCA and SVM"(2011); 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication, (ASID 2009), pp. 97 – 101. Hong Kong, 20-22.

[31] Zhou, C, Huang, S, Xiong, N, et al (2015), “Design and Analysis of Multimodel- Based Anomaly Intrusion Detection Systems in Industrial Process Automation, IEEE Trans Syst Man, Cyber System 45:1345–1360.

[32] Zimmer C, Bhat B, and Mueller, F (2009), “Time-Based Intrusion Detection in Cyber-Physical Systems”. 30th IEEE Real-Time SystSymp 89–92