A novel method to detect adversaries using MSOM algorithm’s longitudinal conjecture model in SCADA network

K.SANGEETHA[1], S.VENKATESAN[2], S. SHITHARTH[3]

[1]Research Scholar, Department of CSE, Sri Satya Sai University of Technology & Medical Sciences, Sehore.

[2]Associate Professor, Department of CSE, Aurora’s Technological and Research Institute, Hyderabad.

[3]Assistant Professor, Department of CSE, Vardhaman College of Engineering, Hyderabad.

Corresponding author: Shitharth.it@gmail.com

***Abstract*— In a neural network, for classifying any high dimensional data, SOM (Self Organizing Maps) is highly preferred. Self-Organizing Maps (SOM) are unsupervised neural networks that cluster high dimensional data and transform complex inputs into easily understandable inputs. It works on data sets and used to find previously unknown patterns. Patterns help in categorizing elements for finding an association between them. They can detect the adversaries and defects in the data. Neural networks are considered as one of the prolific unsupervised learning methods which are a fast, and powerful technique that can be used to solve many real-world problems. These networks are widely used for data representation. An unsupervised algorithm has to understand the patterns in the data and then further process the desired output. This research proposes a modified neural network algorithm called MUSOM (mutated Self Organizing Maps). This algorithm has two major conjectures based on attacker’s response on incoming packets from SCADA sensor. It has two components 1.nodes and 2.neurons. These nodes and neurons are interconnected with each other. One is used to accomplish fault tolerance, error correction and dimensionality reduction concerning the unknown anomalies. The next one is to find the outlying anomalies in the network.**

***Keywords*—Self Organizing Maps (SOM), Mutated Self Organizing Maps (MSOM) Intrusion Detection System (IDS), quantization error (QE) Supervisory Control and Data Acquisition (SCADA),**

I. INTRODUCTION

Supervisory control and data acquisition system are such an integral part of the latest automation industries. This receives data from various sources like sensors, RTU (Remote Terminal Units) and smart meters. The major tasks performed by SCADA is to monitor the connected data fetching sources. SCADA systems are mainly used to control and monitoring purposes in various industrial applications. It can be used for a small office building to monitor environmental conditions also used to monitor complex conditions in a nuclear power plant. SCADA. To protect control systems, systems are evaluated before being deployed in production. So the operators have a good understanding of what types of vulnerability those systems may be introducing into their environment. One of the challenges of control systems is that many of them have been developed in an environment that works very well in operations, but they don’t have all of the cyber security safeguards built into them. Sensor nodes which sense physical phenomenon that occur around them. These sensor nodes are majorly used for medical purposes, agriculture, industrial purposes, and so on. SCADA system uses wired or wireless sensor networks to transport the data from the master station. SCADA systems mainly use wireless sensor networks due to their frequent changing topology nature and the possibility of reconfiguration of networks. Using the wireless sensor networks this information or data is transmitted through a router, firewall, and switches. The first layer of protection is a router. The router should be configured with a VPN tunnel on the router side. Firewalls in control systems are used to protect unauthorized access. Data should be encrypted, to increase the level of information security, while accessing information through the internet. Various security threats are evolving every day like unauthorized access to the control software, virus infection and one more major threat is intruders sending malicious packets to host devices. By sending these packets anyone can control the SCADA devices.

Most of the risks come from nodes with limited resources whereas SCADA systems have globally interconnected nodes. These nodes are easily accessible to intruders. Traditional security measures are not always suitable because of their limited resources. To prevent security attacks, SCADA systems should be secured with enhanced architecture. The architecture consists of the various communication channel that increases the resistance of the sensor networks and data frames. To maintain data integrity and privacy of data some encryption schemes are implemented in wireless sensor networks. Despite various disadvantages in sensor networks, there are some safety measures to overcome security challenges in SCADA. Those are of using specialized protocols and proprietary interface, developing various control and monitoring methods, Authority control for external connections, Analyzing possible attacks, various malicious detection methods and finding reasons for vulnerabilities. Despite various disadvantages in sensor networks, it has one advantage which is called redundancy or fighting against attacks. Machine learning is classified into three types. Supervised learning, Unsupervised Learning, and Reinforcement Learning. SCADA networks mainly use unsupervised learning algorithms to find security solutions with clustering or association methods.

II LITERATURE SURVEY

One of the major problems in sensor networks is Resource Limitation. As a result, we cannot use complex encryption or authentication algorithms. Whenever a failure occurs these algorithms should get updated which is also a difficult task. There are some chances of side-channel attacks to obtain security keys. [1,2] Buttyan and Hubaux [3,4] addressed efficient routing behavior using various routing protocols. In this approach, packets forwarding is done by the nodes across the network, and each node will receive a per-hop payment for forwarding packets. This payment information is stored in a counter and shared within the networks. Jose, alvaro[5] proposed an approach to enhance routing. All the information of misbehaving nodes can be stored in the local reputation system. In this approach packet information or reputation system information shared globally through the networks. So that all the nodes are warned about malicious nodes and based on reputation ratings these nodes are isolated from the routing process. Most of the security protocols in routing can be accessed at the network level. But in this paper, they have proposed a security infrastructure using SOM algorithms which use the information from the application layer.[6]Avoiding non-forwarding nodes are done by using rating methods. These rating methods are used to find misbehavior nodes too. But in this approach, it detects the misbehavior nodes. It cannot isolate the nodes from the routing process[7]. Buchegger and Le Boudec proposed a CONFIDENT routing approach for solving various security problems. This protocol makes misbehavior nodes as less attractive nodes for the routing process. Nodes examine misbehaving nodes and store this information in a local reputation system. Eventually, it conveys this information to neighbor nodes also.[8] Machine learning techniques are proposed to obtain solutions for security problems. SCADA network system has a high level of flexibility and adaptability, yet it consumes notable resources. Feature sets of these techniques are easily accessible by the attacker.[9,10]SCADA systems are vulnerable to LOWSPAN or IP based networks. Lowspan architecture for SCADA systems is used for various applications. When SCADA systems used in industrial environments, it requires high receptivity, high acceptance, stability, and measurements. The efficient extension of LOWSPAN is 6LOWSPAN.To identify various threats and vulnerabilities various encryption and decryption algorithms are used. In this approach, session keys are encrypted using public-key encryption algorithms and for decrypting the data it uses symmetric algorithms. encrypt session keys and symmetric algorithms[11]. Alexey and Anton addressed the issues and purpose of key management and attack detection. A combination of three algorithms is used to calculate routing information frames which is determined by various routing methods. These routing methods use ZIGBEE specifications. It uses the encryption technique and protecting key data by establishing trust between clients at the initial stage and data transferring process. Zigbee feature set supports data encryption that determines the changes in key distribution encryption [12]SCADA networks have been incorporated with the internet which in turn has significantly increased threats to critical infrastructure. In this case, Scada systems should implement an architecture with various data frames that protect the architecture from external attacks.

III. ANOMALY DETECTION

In this anomaly detection, our MSOM proposes two ways of detection. The first way is to calculate the median distance (MD) between every node with its neighbor nodes.[13] Then those median values are compared with one another. In any case, if any of the MD values significantly varies from the rest then it is declared as anomaly nodes. In a second way, we find out the quantization error (QE) in each instance from the cluster center. This is more like a two-step verification process just to make our algorithm more robust and to reduce the false-positive rates. Here we calculate QE value for the outliers data and then compare with inlying nodes QE value [14,15]. In any case, if values seem to vary drastically it is considered as the outliers data. On the close examination of the algorithm, it is noted that the time latency rate is reduced and error detection is improved. The data collected can also be used for cluster labeling as the normal data can be put up in one cluster based on MD values and the remaining in other clusters labeled as intrusive data. Even in any case if the new breached data looks similar to the inlier's data there comes QE value prediction. This helps us to avoid labeling the intrusive data as a genuine one, i.e., basically, it reduces the FP (False Positive) rate which is one of the most crucial constraints in the ICS systems like SCADA.

A.MSOM algorithm conjectures*:*

In the SCADA sensor network, the major threat is from adversaries. Adversaries are a group of persons who have a malicious intent to hack and down the system. The adversaries may be hackers, hacktivists or even cybercriminals [16].Their main motto is to create false alarms in the SCADA sensor network and to create a panic state which later leads to greater chaos. Here we make two conjectures: The attacker can't catch hold of all the incoming packets from the SCADA sensor. So that we can make sure that all the data that is been captured by our network analyzer (like Wireshark) need not be tampered data [17, 18]. So our initial assumption should be only to find out the captured and tampered data. If this assumption goes wrong by anyway, technically the hacker would subvert any protocol that is been implemented which would down the entire network. The next supposition is about data analysis. Usually, the data captured and influenced by the anomalies do statistically differ from the native network data. Hence to do this separation process we implement an anomaly detection model which is based on the SOM clustering principle [19]. This states how much the influenced data is abnormally varying from the native SCADA sensor data. Our proposed algorithm sheer lies in the data sequence of the SCADA sensor network. The data are examined in an order based on how good a wave can interfere with its stat but at a different point of time. By finding how monochromatic the data is we can find out the average correlation between the wave value and its constant phase relation [20]. The waves may be transverse longitudinal. But most of the sensor data we deal with comes under longitudinal. Hence we record the sequential longitudinal based sensor outputs. In this way, we can find out whether the data is intact or not and also detect the adversary though they are previously unseen from our attack library base.

A. 1.	*B. 1.21*	*C. 1.23*	*D. Occurrence2*	*E. Frequency 0:34*
F. 1	*G. 1:24*	*H. 1:23*	*I. Occurrence1*	*J. Frequency 0:18*
K. 1	*L. 1:23*	*M. 1:11*	*N. Occurrence1*	*O. Frequency 0.18*
P. 1	*Q. 1:11*	*R. 1:10*	*S. Occurrence1*	*T. Frequency 0.18*
U. 1	*V. 1:10*	*W. 1:25*	*X. Occurrence1*	*Y. Frequency 0.18*

Table no: 1-Linearity sequence Model

The proposed longitudinal model has certain measurements called n-grams (frequency in the given sequence).
The sequence goes by,
1:00; 1:20; 1:23; 1:22; 1:10; 1:00; 1:20; 1:23

As mentioned in Table no: 2. This model is used to calculate the linearity in the sequence which assures high mobility of the process. It also calculates pairwise anomaly detection based on clustering.
Figure 5. Shows the impact of the cloned outlier nodes Occurrence of the impact of the attack and the based on the redundancy of the nodes. Figure 5(a) shows how the sensor node redundancy and the impact of the attack are interdependent. The X-axis shows the total number of sensor nodes per every 100 outliers (n). Figure 5(b) reflects the system’s behaviors based on the network redundancy.

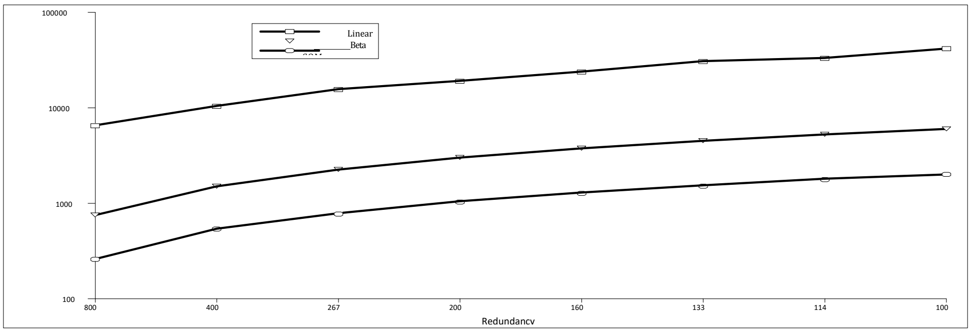

Fig 1(a) Impact of attack

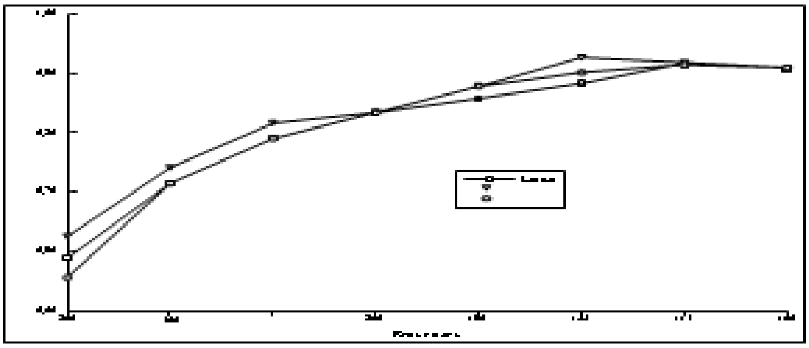

Fig 1(b) Degradation of the system

The graph comparison is between Linear, Beta, and SoM. Figure 6. Reputation occurrence for a clone attack. In (Figures 6(a) and 6(b) clone attack is not confined due to its low reputation. In Figure 6(c) it is evident that the MSOM algorithm finds out and filter the cloned attack nodes. This experiment consists of 180 foul nodes in a defined region and after a meanwhile, they get masqueraded as clone node with the same identity as of the inlier's node. There some basic constrains for reputation value such as reputation values are always ranged between [0, 1] where 0 represents the lowest trust value and 1 represents the maximum trust value. We give two more additional values, RQE and RMD which represent the reputation quantization error and Reputation Mean Distance.

$$\text{RM D} = \frac{(\text{MMDvalue} - \text{BMD}_{value})}{maxMD_{value}}$$

where MMD value is the maximum median distance and BMD value is the best matching unit's mean distance. In this way, RM chooses value in between the mentioned range (0,1) which have a higher reputation. Figure 2. Shows Reputation occurrence for a clone attack

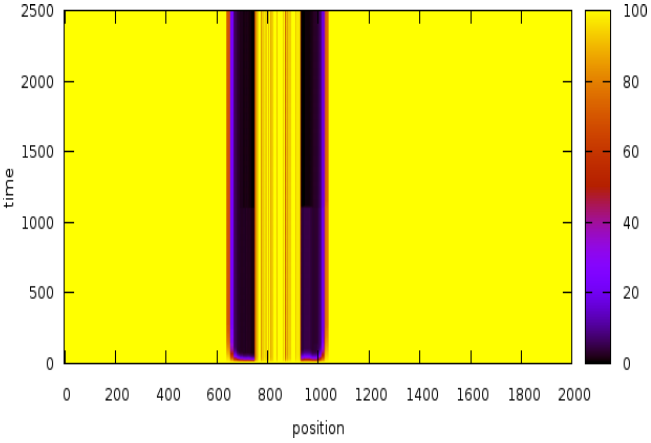

Fig 2(a) with linear algorithm

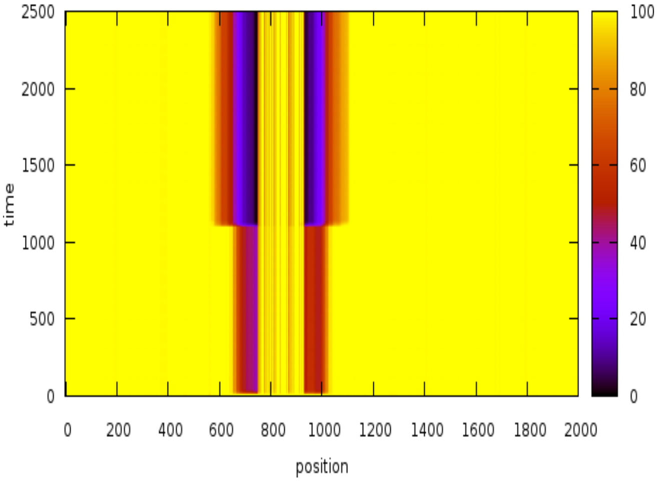

Fig 2(b) with beta algorithm

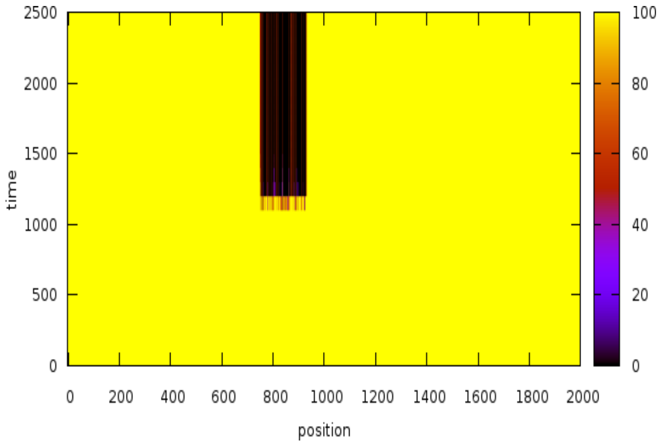

Fig 2(c) With SOM algorithm

Fig 3 shows Reputation occurance of clone attack

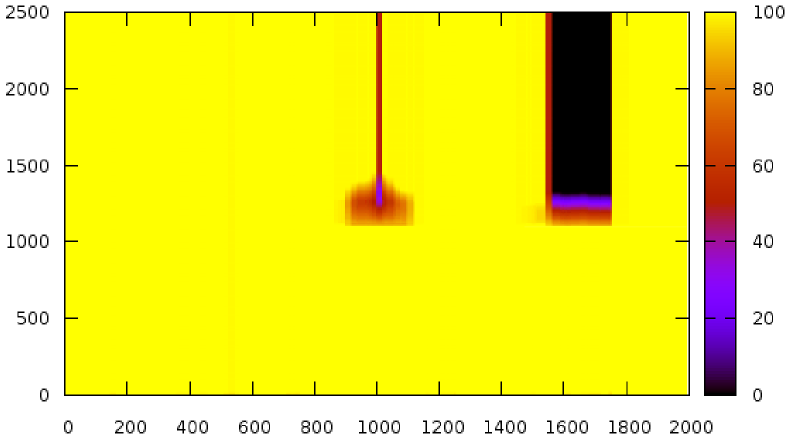

Fig 3(a) with a linear algorithm

Figure 3 shows the reputation of every node instance in the clone attack. The X-axis indicates space Y-axis represents time whereas, axis shows reputation values which are shown by color gradation.

This is the simulation result of 2000 nodes where the clone attack with 800 malicious clone nodes. The initial clone node is placed near the cluster center and the network is well and good until the 100th iteration and once the clone attack is launched the results are different. In all three algorithms, the detection time is almost the same. In linear algorithms (Figure 3(a)), the nodes close to the cloned ones have drastically decreased and needs time to recover. Figure 3(b) shows that the beta algorithm has the same effect but with a relatively higher impact. Finally in the MSOM algorithm (Figure 3(c)) the neighbor nodes are not affected and it leniently isolated the victim nodes.

IV. CONCLUSION

Our proposed algorithm modifies the neural network into MSOM algorithm which detects the attacker's behavioural response based on the traffic in the network. Moreover the conjectures is of nodes and neurons that examines every node's communication in the promiscuous mode to trace the symptoms of malicious behaviour. The SOM agents have a reputation system analysing the trust value of every node. This trust value is calculated based on the node behaviour. But in our proposed system, by finding how monochromatic the data is we can find out the average correlation between the wave value and its constant phase relation. The waves may be transverse longitudinal. But most of the sensor data we deal with comes under longitudinal. Hence we record the sequential longitudinal based sensor outputs. This results in a good way to find about the intactness of the data. In future research, MSOM algorithm is about to be tested in a real time test bed and attack simulation is made from different systems so that its efficiency and robustness is checked. This make sure that any further changes in this algorithm has to be done or not and will open doors for further research.

REFERENCES

[1] Ravi, S.; Raghunathan, A.; Kocher, P.; Hattangady, ‘S. Security in embedded systems: design challenges’. Trans. on Embed. Comput. Sys. 2004, 3, 461–491.

[2] Bar-El, H.; Choukri, H.; Naccache, D.; Tunstall, M.; Whelan, C. ,’The Sorcerer’s apprentice guide to fault attacks’. Proc. IEEE 2006, 94, 370–382.

[3] Butty, Hubaux, ‘Enforcing service availability in mobile ad-hoc WANs’ in Proceedings of the 1st ACM international symposium on Mobile ad hoc networking & computing; IEEE Press: Boston, MA, USA, 2000; pp. 87–96..

[4] Butty , Hubaux, J., ‘Stimulating cooperation in self-organizing mobile ad hoc networks.’ Mob.Netw. Appl. 2003, 8, 579–592

[5] Jose M.Moya,Alvaro Araujo,Zorana bankovic,juan-Mariano de Goveneche,Juan carlos Vallejo,Pedro Malagon,Daniel Villanueva,David Fraga,Elena Romero And Javior Blesa.’Improving Security for SCADA Sensor networks with Reputation Systems and Self organizing maps’ Madrid 2009,ISSN 1424-8220.

[6] Devi, S.Shitharth, ‘An Appraisal over Intrusion Detection systems in cloud computing security attacks’, 2nd International Conference on Innovative Mechanisms for Inddustry applications, ICIMIA -2020, Conference Proceedings, pp. 122

[7] Buchegger, S.; Boudec, J.L. ‘Performance analysis of the CONFIDANT protocol in Symposium on Mobile Ad hoc Networking & Computing;’ ACM: Lausanne, Switzerland, 2002; pp. 226–236.

[8] Marti, S.; Giuli, T.J.; Lai, K.; Baker, M., ‘Mitigating routing misbehavior in mobile ad hoc networks’ in Proceedings of the 6th annual international conference on Mobile computing and networking; ACM: Boston, MA, USA, 2000; pp. 255–266

[9] S.Shitharth, Masood Shaik, Sirajudeen, Sangeetha, ‘Integrated Probability relevancy classification (IPRC) for IDS in SCADA’, Design Framework for wireless network, Lecture notes in network and systems, Springer , vol. 82, Issue 1, 2019, pp. 41-64

[10] S.Shitharth, Masood Shaik, Sirajudeen, Sangeetha, ‘Mining of intrusion attack in SCADA network using clustering and genetically seeded florabased optimal classification algorithm’, Information Security, IET, vol. 14, Issue 1, 2019, pp. 1-1

[11] Wallenta, C.; Kim, J.; Bentley, P.; Hailes, S. ‘Detecting interest cache poisoning in sensor networks using an artificial immune algorithm.’ Appl. Intell. 2008, doi: 10.1007/s10489-008-0132-0.

[12] .Yu, Z.; Tsai, J.J.P. A ‘Framework of Machine Learning Based Intrusion Detection for Wireless Sensor Networks’. In Proceedings of the 2008 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC 2008)–Volume 00; IEEE Computer Society: Taichung, Taiwan, 2008; pp. 272–279

[13] Shitharth, &D.Prince Winston, D 2017, ‘An Enhanced Optimization algorithm for Intrusion Detection in SCADA Network’, Journal of Computers and Security, Elsevier, vol. 70, pp. 16-26.

[14] Shitharth, D. Prince Winston, “Comparison of PRC based RVM classification versus SVM classification in SCADA network”, Journal of Electrical Engineering, Vol.17 (1), Jan 2017, pp. 318-331.

[15] Yvtte.E,Gelago,Tai-hoon-kim,’Enhanced security mechanisms for securing SCADA wireless sensor networks’, .International Journal of Sensor and Its Applications for Control Systems Vol.2, No.1 (2014), pp.111-116

[16] Alexey.G.Finogeev,Anton.A.Finogeev.’Information attacks and security in wireless sensor networks of industrial SCADA system’, Journal of industrial information integration.Penza state university Russia .2017

[17] L. Aiping, et al., "A New Method of Data Preprocessing for Network Security Situational Awareness," in 2010 2nd International Workshop on Database Technology and Applications (DBTA), 2010, pp. 1-4

[18] R. R. Karthick, et al., "Adaptive network intrusion detection system using a hybrid approach," in Communication Systems and Networks (COMSNETS), 2012 Fourth International Conference on, 2012, pp. 1-7.

[19] S.-J. Horng, et al., "A novel intrusion detection system based on hierarchical clustering and support vector machines," Expert Systems with Applications, vol. 38, pp. 306-313, 2011.

[20] Fatima hussain,Rasheed hussain,Syed alihasan,Ekram hussain”Machine learning IoT security current problems and future chanllenges.”2019