TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 9497 - 9504

A Novel Unified Automata Integrated Intrusion Detection Model for Wireless Sensor Network

S. Prithi[1],S. Sumathi[2]

[1]*Assistant Professor, Department of CSE, Rajalakshmi Engineering College, Chennai, India*
[2]*Professor, Department of EEE, PSG College of Technology, Coimbatore, India*
[1]*prithi.s@rajalakshmi.edu.in,* [2]*ssi.eee@psgtech.ac.in*

Article Info
Volume 82
Page Number: 9497 - 9504
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 10 February 2020*

Abstract

To design an efficient intrusion detection system (IDS) various machine learning algorithms such as support vector machine, artificial neural networks, random forest, naïve Bayes and decision trees have been used. In this work, a Hybrid Support Vector Machine – Decision Tree / Random Forest (SVM-DT/RF) based IDS is integrated along with the automata with a view to ameliorate the utilization of energy and lifetime of WSN by detecting the malicious packets and discarding them before the node gets affected. By integrating the IDS with automata, the proposed model improves the detection rate, accuracy as well the energy is efficiently used among the nodes and network lifetime is extended. The proposed automata integrated Hybrid SVM-RF IDS shows an improvement in energy and network lifetime than automata integrated Hybrid SVM-DT, without automata integrated Hybrid SVM-DT IDS and cluster-based IDS.

Keywords; *Decision Trees, Intrusion Detection System, Automata, Random Forest, Support Vector Machine, Wireless Sensor Networks.*

I. INTRODUCTION

A large number of sensor nodes known as wireless sensor networks communicate and transmit the data to each other using multi-hop transmissions. The ultimate functionality of these node is to monitor and accumulate the data around a specific region in order to provide protection. The utmost sensor nodes energy are consumed by transmitting and receiving packets from nearby nodes. Therefore, designing an energy efficient scheme is an exciting topic for investigators. LEACH protocol [8] acts as a clustering approach which reduces energy dissipation in a sensor network. The authors have proven that LEACH outclasses the traditional clustering algorithms such as Energy Aware Data (EAD) centric routing [12], rumor routing [13], Geographic Energy Aware Routing (GEAR) [20]. Mahmoodet. al anticipated a cluster head replacement method called as MODLEACH [31] which is an energy efficient scheme. The cluster head is selected at each subsequent round using the threshold value. The amount of energy unexploited in packet routing and for forming the cluster and cluster head selection are preserved using this algorithm. Recent research on routing sensor networks indicates that researchers are more interested in providing energy-aware routing to increase network lifetime [32][33]. Therefore, energy efficient routing must be considered when selecting the clustering algorithm.

Secondly, the sensor nodes are susceptible to numerous forms of attacks. Intrusions are mainly precipitated by privileged attackers and legal users trying to misuse authorized rights [2]. Classification techniques can be used to detect attack such as Support Vector Machine (SVM) [5], Random Forest (RF) [4], Decision Tree (DT) [6], as well as combining one or more classifiers achieve an

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9497 - 9504

improved detection rate and accuracy and can successfully detect network intrusion.

Therefore, the core aim is to recognize the malicious activities that occur in the network and to classy these attacks into DoS, probe, R2L and U2R attacks. It also obtains an energy efficient optimal route for broadcasting the data beginning at sensor node to sink node thereby to improve the lifetime of the network, reduce energy conservation and throughput. To achieve this objective an automaton system has been implemented. Automata plays a foremost role in learning and continuously monitoring the network environment and can continuously keep track of the network activities. Based on the network environment and nodes density, the automaton can regulate the transmission energy level of the node. An automaton model has high adaptability to environmental changes, and hence, it is compatible to vastly dynamic WSN environments. A novel unified Automata is integrated with Hybrid SVM-RF/DT IDS which continuously monitors the network environment and detects the malicious activities that occurs in network. The intrusion detection module categorizes the attack into four main classes namely, Probe, U2R, R2L and DoS attack. This model improves the detection rate, accuracy, network lifetime as well as the energy is utilized efficiently.

A comprehensive related field of study is performed on the several clustering and routing algorithms and classifiers such as SVM, RF, DTetc are deliberated in section 2. The framework of the proposed novel unified automata integrated Hybrid SVM-RF/DT IDS, network model, anomaly and misuse detector module and the dataset used to assess the effectiveness of the system are discussed in section 3. The researches carried out by the proposed framework are discussed in section 4 and the closing comments are delivered in section 5.

II. LITERATURE SURVEY

Energy efficiency grabs the attention of various researchers since all the innovations in various technologies leads to the sustainable global energy system. Consequently, many routing algorithms based on efficient clustering have been developed by several researchers such as LEACH [8], EECS [21] and HEED [23]. From [8], it is found that the first hierarchical protocol developed for homogeneous WSNs based on clustering is LEACH. An extension of LEACH is the Hybrid Energy Efficient Distributed (HEED) clustering [23] wherein the selection of CH confides on the residual energy and transmission cost. The main benefit is that this protocol does not permit any node to participate in more than one cluster. The major threat to the HEED is that the hotspot problem that occurs during this protocol acts as a big threat to the network lifetime. Therefore, another work known as an Energy Efficient Clustering Scheme (EECS) [21] which combines the possibility of selection scheme from LEACH and the structure from HEED. Initially the node is chosen to exploit as a tentative CH and the other tentative heads participate in the competition to act as the final CH. The distribution of CH is improved in EECS. But there is no change in the energy consumptions in the nodes that possess the Euclidean distance from the sink and in addition the hot spot issue remains unsolved.

Greedy Load Balanced Clustering Algorithm (GLBCA) algorithm was proposed by Chor Ping Low et. al [19] and the author's intention was to bring the sensor nodes together to group as cluster. The foremost focus was to enhance the global expandability of the network. The process functioned by preserving the load among the gateways. An experimental scheme was presented by Wenkeet. al [25] to exhibit the various classification and clustering techniques to detect the doubtable actions. The data mining techniques was integrated with intrusion detection system by Hemalathaet. al [24] to categorize the relevant data competently. The proposed algorithm focused on the

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9497 - 9504

primary problems such as data classification, lack of labels in the data, human interaction level and the efficiency of various attacks. The technique was trained and tested using KDD Cup'99 dataset [9] and the outcomes exhibited with a high accuracy rate and there was reduction in false alarm rate.

Sulaimam et al. [26] reviewed the benefits and limitations of the developed data mining methods such as Bayesian Classifier, fuzzy logic, genetic algorithm, neural network and support vector machine in IDSs. Weller-Fahy et al [27] presented the usage of resemblance and the measure of distance within the network. A comprehensive background study has been performed on decision tree (DT) [17], fuzzy logic [29], support vector machines (SVM) [30], neural networks [28], random forest (RF) [16] and in order to design an efficient utilization of energy and an effective intrusion detection system. Various network classifiers were evolved with the help of the above algorithms to categorize the traffic of the network into two classes namely, normal class and attack class.

Finite automata (FA) have been recognized to accomplish well in many systems for anomaly detection such as fraudulence detection, liability detection, monitoring system health and event detection in sensor networks [22]. Deterministic Finite Automata (DFA) are designed to identify or receive member strings of a particular regular language. For instance, to recognize the language of various attacks such as sybil attack, wormhole attack, selective forwarding attack DFA can be designed. The author Joel W. Branch [11] developed a system that gives itself to the detection of DoS attacks by means of time dependent deterministic finite automata (TDFA). The authors Zong-Fen Han et.al [7] proposed an Adaptive Time dependent Finite Automata (ATFA) which is an extension of time dependent deterministic finite automata. The authors Gholipour and Meybodi proposed LA-Mobicast [1] which utilizes the automata to flexibly control the shape and the location of the progressing region. This model has implemented a complete distributed algorithm that needs lesser communication overhead to determine the forwarding zone.

The major emphasis of this work is on the hybridization of Support Vector Machine with Decision Tree and Random Forest for intrusion detection system in integrating Learning Dynamic Deterministic Finite Automata (LD^2FA) [34]. The core objective is to integrate Hybrid SVM-RF/DT IDS [14] with automata so that intrusions that occur in the network can be identified easily and automated to drop the malicious packets thereby the lifetime of the nodes can be extended and obtains almost cent percent accuracy and detection rate. The proposed framework concentrates on efficiently utilizing energy, improving the detection rate, true positive rate and overall accuracy of intrusion detection and also focusses on extending the network lifetime.

III. PROPOSED MODEL

3.1 Proposed Framework

Hybrid SVM-RF/DT IDS integrated with automata has been proposed in which the automata adapt to the responses from the network environment through a series of interactions about the sensor nodes. It dynamically learns the characteristics of the environment of the network such as node's position, its corresponding neighbor nodes, timestamp, cluster head information residual energy and when the node becomes dead node and identifies malicious activities that takes place in the environment. The automaton progressively studies the network environment through which the node, packet and route information are monitored and inspected. Hybrid Particle Swarm Grey Wolf Optimizer algorithm validates all the feasible route from the starting vertex to sink vertex and finds the optimal transmission route from origin vertex to sink vertex. In this optimal route the packets are traversed through the nodes and Hybrid SVM-RF/DT IDS agent is enabled which examines the

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9497 - 9504

packet samples and detects the malicious packets. The malicious packets are dropped thereby the energy consumption of the node is not detained. As the malicious packets are identified and dropped without affecting the node the network lifetime is also extended. The detection of malicious packets in the earlier stage makes the normal packets to move towards the destination at a faster rate. Thus, the packets are delivered more accurately with high detection rate and reaches maximum accuracy.

3.2 Automata integrated Hybrid SVM-RF/DT IDS

The main objective of automata integrated Hybrid SVM-RF/DT is to secure the transmission nodes while broadcasting the message so that the consumption of energy and lifetime of the network can be improved. The NSL-KDD datasets [10] and KDDCup'99 datasets [9] are used as the packet sample to measure the effectiveness of IDSs. Since SVM classification system cannot process the dataset in its present format each data sample is represented as numerical values so the categorical samples are converted into numerical samples. Feature 2, 3 and 4 contained strings and therefore it was converted into numeric data using factorize method [3]. In this proposed IDS, statistical normalization [15] is used which converts derived sample from any normal distribution to a standard normal distribution with mean 0 and 1 to achieve unity-based normalization. The samples are grouped into training and testing dataset and finally the samples are passed to the anomaly detector module. The anomaly detector module classifies the samples into normal and malicious samples. The malicious samples are further passed to the misuse detector module. The misuse detector module analyses the malicious packets, identifies the attacks and categorizes the attack into the attack classes.

IV. RESULTS AND DISCUSSIONS

The automata integrated model is experimented through simulation in MATLAB. NSL-KDD dataset and KDDCup'99 datasets are employed to train and test the automata integrated SVM-RF model and automata integrated SVM-DT model. The training phase is carried out with 90% of NSL-KDD dataset samples and testing phase are carried out with 10% of samples. The overall accuracy, true positive rate, precision rate and F1-Measure are measured and compared for NSL-KDD dataset and KDDCup'99 dataset with the data mining techniques such as linear-SVM [16], RBF-SVM [16], Random Forest [16] and Decision Tree [17]. The overall performance such as accuracy, true positive rate, precision rate and F1-score of the automata integrated SVM-RF model and automata integrated SVM-DT models are evaluated besides the benefits and limitations of each model are discussed in this section.

4.1 Evaluation of NSL-KDD Dataset

The overall accuracy of DT, RF, linear-SVM, RBF-SVM, automata integrated SVM-DT and automata integrated SVM-RF models on 10% testing and 90% training data samples for NSL-KDD dataset is depicted in Table 1. The overall accurateness of the automata integrated SVM-RF model achieves 0.10% improvement when compared with automata integrated Hybrid SVM-DT IDS, 5.07% better than DT, 2.12 % increase in accuracy than RF, 1.52% betterment than RBF-SVM and 0.10% improved accuracy than Linear-SVM.

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9497 - 9504

Evaluation Metric	DT	RF	Linear-SVM	RBF-SVM	Automata integrated SVM-DT	Automata integrated SVM-RF
Accuracy (%)	94.6	97.64	98.76	98.24	99.64	99.77
Precision (%)	99.06	97.92	97.45	97.78	99.69	99.76
True positive(%)	98.74	96.44	97.78	97.35	99.94	99.77

Table 1. Accuracy, Precision and True positive of NSL KDD Dataset

The detection rate and true positive rate of DT, RF, linear-SVM, RBF-SVM, the proposed automata integrated SVM-DT and the proposed automata integrated SVM-RF is depicted in Table.1. The rate of detection of the automata integrated SVM-RF achieves 0.16% improvement when compared to automata integrated SVM-DT, 0.4% improvement when compared with Decision Tree, 1.96% increase in comparison with SVM (RBF) and 2.28% better when compared to linear-SVM. The true positive rate of the automata integrated SVM-RF achieves 1.07% improvement when compared with Decision Tree, 3.32 % better true positive rate than Random Forest, 0.16% better than automata integrated SVM-DT, 2.37% better than RBF-SVM and 2.02% improvement compared with linear-SVM. The computational time of the automata integrated SVM-RF model and automata integrated SVM-DT model are 2.5392ms and 1.9720ms respectively.

The performance of lifetime of the network is depicted in Fig.1. The round at which the exhaustion of energy of the first node takes place is given in Fig.1 for the proposed and the existing algorithms such as automata integrated Hybrid SVM-DT/RF IDS, without automata integrated IDS [34], PSO [18] and GLBCA [19]

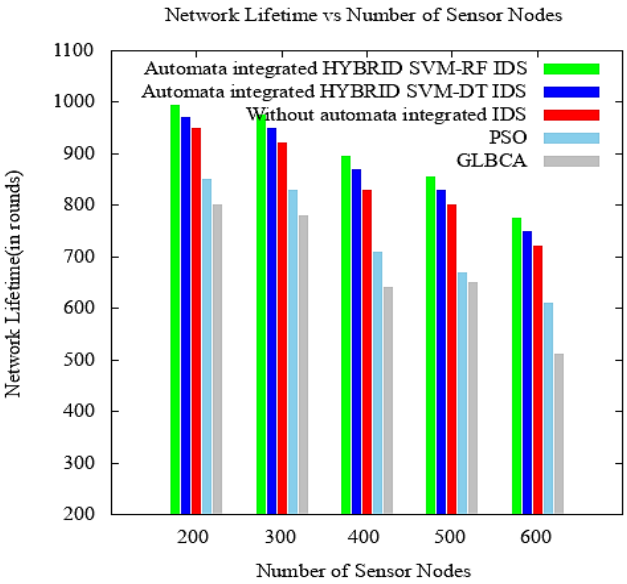

Fig.1. Network Lifetime

Fig.2 depicts the performance of consumption of energy for 200 - 2000 sensor nodes for each round. The automata integrated Hybrid SVM-RF exhausts lesser energy when compared with automata integrated Hybrid SVM-DT, without automata integrated IDS [34], PSO [18] and GLBCA [19]. The overall performance of the energy consumption indicated in Fig.2 shows that the proposed automata integrated Hybrid SVM-RF has obtained 6% lesser than automata integrated Hybrid SVM-DT, 11% lesser than without automata integrated IDS [34], 15% lesser energy than PSO [18], and 16% lesser than GLBCA [19].

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9497 - 9504

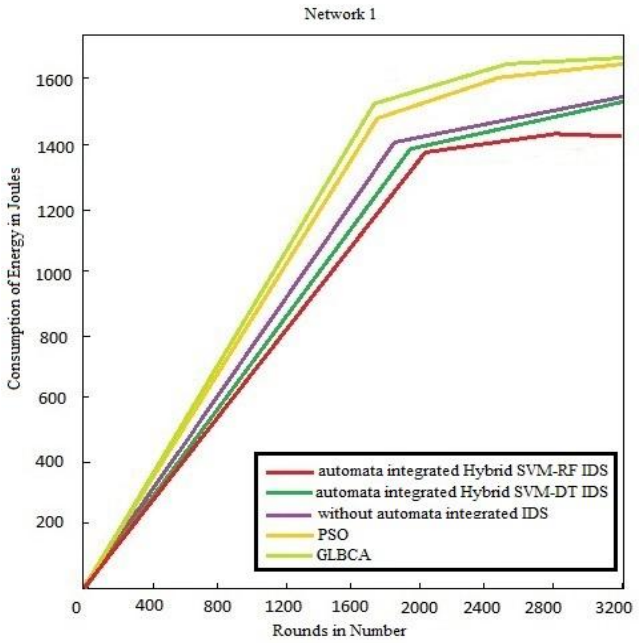

Fig.2. Energy Consumption

V. CONCLUSION

The proposed work has integrated an automata-based learning environment with Hybrid SVM-RF/DT IDS to dynamically learn and monitor the network environment to produce almost cent percent accuracy and detection rate besides it eliminates the intruders such that the data transmission can be successfully completed and the energy consumption can be reduced. The automata integrated Hybrid SVM-DT/RF IDS uses Hybrid PSO-GWO algorithm to persuade the legitimacy of all the feasible paths. The various performance measures such as lifetime of the network, consumption of energy, accuracy, true positive rate, detection rate and F1-score are considered for measuring the performance. The energy consumption of automata integrated Hybrid SVM-RF is almost 6%, 11%, 15% and 16% lesser than automata integrated Hybrid SVM-DT, without automata integrated IDS [34], PSO [18] and GLBCA [19]. The network lifetime of the proposed approach has been prolonged where the first node's energy exhaustion takes place after 1230^{th} round. From the inferences made on the two proposed algorithms, automata integrated Hybrid SVM-RF is considered as an efficient utilization of energy approach to route the packets in an optimal route and to detect malicious packets and to classify the attacks in WSN

REFERENCES

[1] Golipour, M., & Meybodi, M. R, "LA-mobicast: A learning automata based mobicast routing protocol for wireless sensor networks. Sensor Letters, 6(2), 305-311, 2008.

[2] Miao Xie, SongHan, BimingTian and, SaziaParvin, "Anomaly Detection in Wireless Sensor Networks: A Survey", Elsevier Journal of Network and Computer Applications, Vol.34, pp. 1302-1325, 2011.

[3] R. Belohlavek and V. Vychodil: Discovery of optimal factors in binary data via a novel method of matrix decomposition. J. Comput. System Sci 76(1)(2010), 3-20.

[4] J. Zhang, M. Zulkernine, A. Haque, "Random-Forests-Based Network Intrusion Detection Systems", IEEE Trans. Syst. Man Cybern. C, vol. 38, no. 5, pp. 649-659, 2008.

[5] Mukkamala, S., Janoski, G., Sung, A.: Intrusion detection: support vector machines and neural networks. In: Proceedings of the IEEE International Joint Conference on Neural Networks (ANNIE), pp. 1702–1707. St. Louis, MO (2002)

[6] J. R. Quinlan, "Induction of Decision Trees", Machine Learning, vol. 1, no. 1, pp. 81-106, March 1986.

[7] Zong-Fen Han, Jian-Ping Zou, Hai Jin, Yan-Ping Yang, Jun-Hwa Sun, "Intrusiqn Detection Using Adaptive Time-Dependent Finite Automata", Proceeding of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004

[8] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. "Energy Efficient Communication Protocols for Wireless Microsensor Networks". In Proceedings of Hawaiian International Conference on Systems Science, January 2000.

[9] http://kdd.ics.uci.edu/databases/kddcup99/ kddcup.

[10] University of North Brunswick, "NSL-KDD Dataset," 2016. [Online]. Available: https://web.archive.org/web/20150205070216/http://nsl.cs.unb.ca/NSL-KDD/. [Accessed: 03-Mar2016].

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9497 - 9504

[11] Joel W. Branch, "Extended Automata-Based Approaches To Intrusion Detection", Thesis Report, Rensselaer Polytechnic Institute Troy, New York, March 2003

[12] A. Boukerche, X. Cheng, and J. Linus, "Energy-aware data-centric routing in microsensor networks", Proceedings ACM MSWiM, in conjunction with ACM MobiCom, San Diego, CA, Sept. 2003, pp. 4249.

[13] D. Braginsky and D. Estrin, "Rumor routing algorithm in sensor networks", Proceedings ACM WSNA, in conjunction with ACM MobiCom'02, Atlanta, GA, Sept. 2002, pp. 22-31.

[14] S.Prithi, S.Sumathi, " Intrusion Detection System using Hybrid SVM-RF and SVM-DT in Wireless Sensor Networks", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S8, August 2019. https://doi:10.35940/ijrte.B1200.0882S819

[15] B. Etzkorn, "Data Normalization and Standardization," 2011. [Online]. Available: http://www.benetzkorn.com/2011/11/data-normalization-and-standardization/.

[16] Ahmad, I., Basheri, M., Iqbal, M.J., Rahim, A.: "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection". IEEE Access 6, 33789–33795 (2018)

[17] Vaishali Kosamkar, S Chaudhari Sangita, "Improved Intrusion Detection System using C4.5 Decision Tree and Support Vector Machine", International Journal of Computer Science and Information Technologies, vol. 5, no. 2, pp. 1463-1467, 2014.

[18] Pratyay Kuila, Prasanta K.Jana,"Energy Efficient Clustering and Routing Algorithms for Wireless Sensor Networks: Particle Swarm Optimization Approach", Engineering Applications of Artificial Intelligence 33 (2014) 127 – 140.

[19] Chor Ping Low, Can Fang, Jim Mee Ng, Yew Hock Ang, "Efficient Load-Balancing Clustering Algorithms for Wireless Sensor Networks", Computer Communications,2008,750–759.

[20] Y. Yu, R. Govindan, and D. Estrin, "Geographical and energy aware routing: A recursive data dissemination protocol for wireless sensor networks", Technical Report UCLA/CSD-TR-01-0023, UCLA Computer Science Department, May 2001.

[21] Mao Ye, ChengfaLi, GuiHaiChen, Jie Wu, "An Energy Efficient Clustering Scheme in Wireless Sensor Networks", Ad Hoc & Sensor Wireless Networks, Vol 3, pp. 99-119.

[22] V. Ramezani, S. Yang, and John Baras, "Finite Automata Models for Anomaly Detection", The Institute for Systems Research Technical Reports, TR 2002-42, CISS March 2003.

[23] K. O. Younis, S. Fahmy, "HEED: a hybrid, energy-efficient distributed clustering approach for ad hoc sensor networks," Mobile Computing, IEEE Transactions on, Vol. 3, No. 4., pp. 366-379,2004.

[24] G.V. Nadiammai, M. Hemalatha, "Effective approach toward Intrusion Detection System using data mining techniques", Egyptian Informatics Journal 2015, in proceeding Elsevier,37–50.

[25] Wenke Lee and Salvatore J. Stolfo "Data mining approaches for intrusion detection", In Proceedings of the 7th USENIX Security Symposium - Volume 7, SSYM'98, pages 6–6, Berkeley, CA, USA, 1998.

[26] A. S. Subaira and P. Anitha, "Efficient classification mechanism for network intrusion detection system based on data mining techniques: A survey," in 2014 IEEE 8th International Conference on Intelligent Systems and Control (ISCO), Jan 2014, pp. 274–280.

[27] D. J. Weller-Fahy, B. J. Borghetti, and A. A. Sodemann, "A survey of distance and similarity measures used within network intrusion anomaly detection," IEEE Communications Surveys Tutorials, vol. 17, no. 1, pp. 70–91, Firstquarter 2015.

[28] Srinivas Mockamole, Guadalupe Janoski, Andrew Sung, Intrusion Detection: Support Vector Machines and Neural Networks, In Proceedings of the IEEE International Joint Conference on Neural Networks, 2002, pp. 1702-1707.

[29] S. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review", Applied Soft Computing, vol.10, pp. 1-35, 2010.

[30] S. Mukkamala, G. Janoski and A. Sung, "Intrusion detection: support vector machines and neural networks", In proceedings of the IEEE International

Published by: The Mattingley Publishing Co., Inc.

January - February 2020
ISSN: 0193 - 4120 Page No. 9497 - 9504

Joint Conference on Neural Networks (ANNIE), St. Louis, MO, 2002, pp. 1702-1707

[31] D. Mahmood, N. Javaid, S. Mahmood, S. Qureshi, A. M. Memon, T. Zaman, "MODLEACH: A Variant of LEACH for WSNs", BWCCA '13 Proceedings of the 2013 Eighth International Conference on Broadband and Wireless Computing, Communication and Applications, 2013

[32] F. Kiani, "Maximizing Wireless Sensor Network Lifetime Based on Linear Programming Method", International Research Journal of Engineering and Technology, Vol. 3, No. 3, pp. 1354-1359, 2018.

[33] S. Jabbar et al., "Analysis of Factors Affecting Energy Aware Routing in Wireless Sensor Network", Wireless Communications and Mobile Computing, Vol. 2018, pp. 121, 2018.

[34] S. Prithi, S. Sumathi, "LD2FA-PSO: A novel Learning Dynamic Deterministic Finite Automata with PSO algorithm for secured energy efficient routing in Wireless Sensor Network", Ad Hoc Networks, 2019. https://doi.org/10.1016/j.adhoc.2019.102024

Published by: The Mattingley Publishing Co., Inc.