iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** II **Month of publication:** February 2022

DOI: https://doi.org/10.22214/ijraset.2022.40327

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue II Feb 2022- Available at www.ijraset.com

Review on Network Security Using Machine Learning Algorithms

K. Yasotha[1], Dr. K. Meenakshi Sundaram[2]
[1]*Research Scholar (PT), Department of Computer Science, Erode Arts and Science College, Erode, Tamilnadu-638009, India*
Assistant Professor in Computer Science, PPG College of Arts and Science, Coimbatore, Tamilnadu-641035, India
[2]*Associate Professor and Head, Department of Computer Science, Erode Arts and Science College, Erode, Tamilnadu-638009, India.*

Abstract: ***Today various real world applications network such as VANETs, MANETs, and WSNs supply a wide range of applications from security to infotainment. As the communication and network field has linked faraway corners of the world using advanced network technology, intruders, hackers or attackers have also raised attacks on networking infrastructure as in enough. Administrators can need to prevent such hacks by using advanced intrusion detection tools. The use of resources, loss of packs, and justice are common performance measures for network based applications, which are usually required by network safety applications to reduce transmission time and to achieve high reliability. The performance of the overall unit is decided by two logics, one is security and another one is privacy. To maintain such data secured architectures, Intrusion Detection System (IDS) is the most commonly used mechanism to detect the attacks on cloud. Hence, more number of researchers focused on this topic. To overcome the limitations of earlier IDS approaches and an elaborate comparison of machine learning based IDS in terms of factors efficiency, execution/detection time, flexibility and true/false rate.***
Keywords: Network security, machine learning algorithm, Intrusion Detection

I. INTRODUCTION

Due to the improvement of wireless sensor and embedded systems, network security is enhanced as one of the rising fields. The networks contain some applications such as event recognition, monitoring, tracking, disaster administration, examination, defensive reservation, etc. In WSNs, the routing can be one of the most research concerns (Rawat Priyanka et al., 2014). However, it can be a difficult work to expand a routing protocol that will be an effective process during its intrinsic distinctiveness of WSNs such as application precise, extremely dynamic network, limited energy, storage and processing ability, etc. In common, the network nodes contain lower computation and communication abilities compared to the full-featured computers contributing to ad hoc networks. In networks, the routing protocols have required to consider the difficulty of effective power resource utilization (Guleria, Kalpna and Anil Kumar Verma., 2019). Sensor networks contain more resourced base stations and resource controlled sensor nodes. In a network, every node will communicate with other sensor nodes or cloud storage using wireless links, wherever the cost of communication has been much higher compared to the cost of computation. Besides, the energy is wanted to broadcast a data packet that can be about double as immense as the energy required to accept similar data. As a result, every message route has intended to the base station can be critical in terms of the lifetime of the network such as applying short routes to the base station that having sensor nodes with exhausted batteries can give up a lessened lifetime of the network. The network delay has also been increased by a long route composed of several sensor nodes. In this network based communications authentication plays an important role in addressing all attack, which is able to check whether a car user and a legitimate user are registered or not before they can access applications. By using message authentication, the user can distinguish between false and reliable information. Authentication schemes are further classified as a one-by-one message verification and batch verification, taking into account the pattern of message verification (Gupta Surbhi et al., 2016).

The Data protection is a system used to protect applications admin and passengers 'sensitive and confidential information from attackers. As well as issues relating to security, the privacy of cars in the Vehicle area Network system should also be taken into account. Several research projects in recent years have been undertaken regarding the safety and privacy that ensures safety and improves the flow. Most recent works have in the past used a pseudonym approach that could be applied to protect users 'privacy and security. Users can improve and reliably preserve privacy by using pseudonym-based approaches. The trusted authority must change pseudonyms frequently in order to control attacks on privacy. In addition, privacy is divided into two types (Lu Rongxing et al., 2019):

1) Privacy of User Protection: It prevents the attackers from providing personal information to users.
2) User Location Protection: Such privacy protects the data of users such as the location or the area followed by the personal data of users such as the user identifier.

Authentication, security and privacy leverage for developing confidence between user and V2I communications in vehicle networks. The main purpose is to identify malicious nodes and fake messages. The authentication schemes make use of appropriate authentication schemes that makes it easy for trust authorities for identifying the attacker and the false message to secure communication (Manvi Sunilkumar and Shrikant Tangade., 2017). Several related works on authentication is proposed in this context, which seek to protect network applications against malicious users, fake messages and unregistered entities and address threats and attacks in all types. In this paper, the Machine learning algorithms for routing security was surveyed. In machine learning methods, the supervised and unsupervised learning methods for WSN will be reviewed. Also, the limitations, basics, and corresponding analysis and processing methods will be reported.

II. REVIEW

In this paper, the review is based on supervised and unsupervised based machine learning algorithm implementation in network security applications are explained and list out the advantages and drawbacks.

A. *Supervised Machine Learning Algorithm for Network Security*

Ponmaniraj et al. (2016) have presented an enhanced localization algorithm framework according to the Compressive security (CS) using the KNN classification method in WSN that has reduced the off-grid issues. In many research works, CS-based localization was presumed that every objective can be in the grid's center. Conversely, this kind of supposition is not possible. The targets have been located in this work wherever in the field. In the distinct spatial area, the target location was observed as an indefinite light signal. In common, the improved discrete signal can be not accurately as K-sparse. The targets have been localized in three distinct environments and the cluster region was selected depending on the power-sharing of targets to enhance the localization accurateness. After that, the final location has been attained using KNN for every classification process. For the process of monitoring in WSNs, Online Anomaly Detection (AD) can play a significant role as a technique that can be used to give the security to WSNs from random faults and cyber-attackers. K-Nearest Neighbor (kNN) algorithm is the most improved algorithm with a lot of awareness for its purposes in WSNs since a parameter-free unsupervised and a scalable and AD method.

Rassam Murad A et al. (2012) have presented two kinds of intelligent localization methods for WSNs. In this work, two methods were introduced to display range-free localization that has exploited the RSS from the anchor nodes in the network. In both methods, soft computing has played a significant role. Every anchor node's edge weight was considered in the first method independently and integrated them to calculate the sensor nodes' position. After that, by using a Fuzzy Logic System (FLS), the edge weights were modeled and the edge weights have been optimized using the GA. The localization was measured as a single difficulty in the second method and whole location mapping of the sensor was estimated from the anchor node signals using neural network In WSNs, the noisy distance estimations can be persistent trouble in the process of localization. In the localization, the NN methods can be not usually utilized, though, neural networks can be a feasible decision to solve the localization difficulties. Swarup and Britto Corthis (2002) compared the three various families of NN such as Multi-Layer Perceptron (MLP), Recurrent Neural Networks (RNN), and Radial Basis Function (RBF). This network performance has also been contrasted with Kalman Filter's two variants that can be conventionally exploited for the localization process. The computation and memory resource necessities were also compared.

Su Ming-Yang (2011) presented an improved kNN-based AD system according to the hyper-grid in this work considerably throughout redefining anomaly from a hypersphere finding area to a hypercube finding area. A hyper-grid structure was converted into a positive coordinate space by an attached coefficient to maintain the idleness for tailor an online update for bit process. Besides, bit operation has encoded the hypercube position by a small number of bits only. Kang Min-Joo and Je-Won Kang (2016) presented an Artificial Neural Network (ANN) according to the energy-efficient and robust routing system for wireless networks. The network was trained in this method over massive data set including every scenario to create the network with high reliability and flexibility to the environment. Besides, the group-based method has been utilized to enhance an entire network's e life-span, anywhere groups may be varied in size. Effective threshold values were given by an artificial neural network for the head node selection of each group and a cluster head has been chosen according to the back propagation method that was provided an efficient and robust group organization. In a genetic code, feed forward ANNs residents contain their structure that will be enhanced in twenty generations. Using the artificial neural network training process, every individual was estimated and additional computation of its root mean square error for every the testing set has also been performed. Here, the nodes were localized by the RSSI measurements that have been applied as the inputs of Artificial Neural Network (ANN).

Rai Kajal et al. (2016) have examined the difficulty of calculating sink mobility in event-driven and deadline-based appliances to attain an improved lifetime of the network. In this process, a deadline and visiting time was established according to the amount of captured data and the kind of occurrence while a sensor node can capture an occurrence. In singlehop broadcast, its trajectory has also been established to the crop data from active sensor nodes. Consequently, the lifetime of the network was enhanced. But NP-hard has occurred while there can be no predefined appearances such as an assignation point or virtual grid in the network. In this work, an algorithm was presented depending on the dynamic programming and decision tree to establish the best Deadline-Based Trajectory (ODT). The geographical active sensor nodes position and the captured events' elements were measured to attain the ODT. The cluster head assortment system according to the four major factors. The optimal node was chosen as the cluster head by using the decision tree method. In this work, an energy efficient and vulnerability-aware clustering algorithm has also been presented for wireless sensor networks. The four factors were utilized on which cluster head assortment can be used. The base station in every round has run the decision tree BS runs the decision tree algorithm and nodes have been chosen that appropriate for being cluster heads. The cluster Nodes must contain high energy, low mobility, and low susceptibility.

Hu Jingjing et al. (2019) have presented an enhanced localization method according to the learning idea of the SVM method. The limited size of grid cells has given the localization accurateness in the process of SVM classification. In this work, the localization error was estimated and contrasted with fuzzy logic and standard SVM. Aysa Mahdi Hassan et al (2020) have presented an Lagrangian Support Vector Machine (LSVM) that contains the following benefits. The network can be localized in LSVM according to the simple connectivity data and consequently. Though SVM is utilized for the classification process, the applicability of SVM will be utilized to find the localization difficulty and show that the localization error is upper bounded by any small threshold provided a proper size of training data. The border and coveragehole issues were found by LSVM efficiently. Lastly, LSVM fast localization was produced by the LSVM method in a distributed method with the effective utilization of communication and processing resources. A modified version of mass-spring optimization has also been presented in this work to the additional enhancement of the location assessment in LSVM. In this work, Area localization has been exploited as estimation metric. Two steps have been utilized in this area localization process: in the first step, establish whether the nodes have moved or not using Radio Frequency (RF)-based technique that has exploited the value modification in RSSI value relatively to range estimation. In the second step, the sensor nodes' are localization has been performed by SVM algorithm and connectivity information. In this process, black hole attacks were discovered and discerning forwarding attacks have also been identified with high accurateness with low energy consumption. The following table 1 explain the various Supervised Machine Learning algorithm implementations for network security

Table 1 Supervised Machine Learning Methods in network security

S.no	Author	Title	Advantages	Disadvantages
1	Ponmaniraj et al. (2016)	K-Nearest Neighbor (KNN)	The advantage of KNN is easy to implement. K-NN cannot be negatively affected while packet data are large and uncaring to conjunction root.	In KNN, it is needed to determine parameter K, the distances have to be estimated between the query instance and every training sample, distances are needed to sort and determine the nearest neighbors based on the Kth minimum 2distance, as well, the categories of the nearest neighbors should be determined.
2	Kang Min-Joo and Je-Won Kang (2016)	neural network	It must be noted that the RSSI values obtained are highly unstable and turn to vary under environmental noise and mobility of sensor nodes. A neural network can offer the advantage that prior knowledge of the environment and noise distribution is not necessary.	The neural networks that can be utilized in the wireless sensor network in the dispersed methods can be still not so persistent during the high computational necessities to studying the weights of network and also the high management overhead Process.
3	Rai Kajal et al. (2016)	decision tree	A decision tree has allowed for partitioning cluster node in a much deeper level	The decision tree often involves a higher time to select the cluster head for the data transmission
4	Hu Jingjing et al. (2019)	SVM	It is quite good in situations where the data samples provided are less as compared to the number of dimensions. SVMs are versatile in nature. To support the decision making more than one kennel methods can be introduced which improves the efficiency.	The performance of SVM decreases in the situations where the number of characteristics provided for classification exceeds the number of data samples provided.

B. Unsupervised Machine Learning Algorithm for Network Security

Tian Li and Wang Jianwen (2009) have presented an energy-efficient K-means clustering-based routing protocol and a best-fixed packet size has been considered depending on the channel conditions and radio parameters transceiver. In this work, the energy consumption of every node was decreased and the network lifetime has been enhanced. Besides, various power levels have been measured for data broadcast from the cluster head node to the cluster member node and BS. Experimental results of this method have demonstrated that this method algorithm has provided the low energy consumption compared to standard K-means based Energy-Aware Clustering (KEAC) under a lifetime of the network and entire network throughput has also been enhanced. In this presented system, a best-fixed packet size was measured for data broadcast to save power and network lifetime is also expanded. The total amount of energy has also been estimated that is required to broadcast the data packet that can be utilized all along with nodes' average distance to the center of the cluster to calculate an average weight. In a WSN, node clustering can be one of the established methods to attain enhanced network lifetime. Numerous methods were presented to attain this goal. An uneven clustering has been discussed with fewer nodes nearer to the BS that has attained the greater effectiveness compared to an identical number of nodes in every cluster and this has happened during the larger overheads for the nodes closer to the BS.

Tan Ling et al (2019) presented improved K-Means L Layer methods that have directed to the cluster generation with few nodes nearer to the BS as opposed to the ones distant away from it for arbitrarily arranged nodes. In this work, the K Means algorithm has been modified that has produced even clustering. Besides, the examination was conducted on the energy consumption of each node with regards to the optimization of the data packet. In WSNs, the sensor node might be lost during malicious attacks by an opponent. For many applications, the WSNs deployed containing military applications can be level to several attacks that can corrupt the performance of the network very quickly. The hybrid anomaly can be acted as one kind of anomaly that has several kinds of attacker nodes like misdirection, black hole, wormhole, etc. In the network, these kinds of attacks may be inserted by applying the hybrid anomaly. Thus, it can be very hard to discover what type of attacker nodes have been operated in the entire network. Consequently, a robust and capable secure IDS method has to be presented to expand the WSN's lifetime.

Sharma Sanjay Kumar et al (2012) have presented enhanced IDS for the hybrid anomaly that has utilized the existing data mining method known as the K-means clustering algorithm. By using the K-means clustering algorithm, intrusion patterns were constructed automatically on training data for the process of anomaly detection. Consequently, matching network activities have discovered the intrusions against these discovery patterns. Over a WSN dataset, this method was evaluated that has been made by applying the Opnetmodeler that having several metrics like end-to-end delay, traffic transferred and traffic received. In this estimation, the training dataset consists of the network parameters' standard values. In a good working manner, the testing dataset was made that contains normal and abnormal network parameters values. The black hole and misdirection nodes were discovered by this presented method.

Lima Moisés F et al. (2010) enhanced WSN lifetime given that sensor nodes can be only prepared with energy-limited batteries. In this work, sensor nodes were clustered and after that data broadcasting was routed to the BS through cluster heads. The issue of cluster head selection was analyzed that may rapidly weaken their energy. Thus, an improved k-Means-based Routing (KMR) protocol was proposed in this work wherever the assortment of CHs can be randomized hence as to improve allocate the broadcast load amongst the same cluster's sensor nodes.

Badis Hammi et al. (2014) utilized the PCA application as a solution of data aggregation in a power and the computationally inadequate environment in WSNs. While a lossy data compression strategy is applied, the optimality of PCA is performed according to the Mean Squared Error. In this work, projection basis performance has been examined that can be not the eigenvectors foundation. The signalling was reduced between the sink node and the Data Aggregation Node (DAN) that can be essential to coordinate matrices of the projection process. Finally, the utmost aggregation technique's computational load was decreased. Then, the number of projections was adjusted by this method that has needed to maintain a threshold value of user defined Normalized Mean Square Error (NMSE). For the large data sets' online recovery, the authors have presented a sparsity model that has permitted the utilization of Compressive Sensing (CS) in real WSN environments, in which the PCA method was applied to capture the temporal and spatial uniqueness of real signals. The principal components' statistical sharing has been estimated by Bayesian examination and it has exposed that the Laplacian distribution was provided a precise statistics illustration of real data. This presented framework was made by combining the PCA and CS called a SCoRe1: Sensing, Compression, and Recovery via On-line Evaluation for wireless sensor networks. The self-adaption is performed by SCoRe1 to random modifications in the sign statistics thanks to a feedback control loop that evaluation.

Mechtri Leila et al (2010) presented two completely distributed consensus-based methods that were definite to meet to the global results by applying local communications only amongst neighbor nodes, in spite of the data sharing or the sparsity of the network: Principal Component Analysis (PCA) is used depending on the discovering of local covariance matrices' eigenvectors, whereas PCA can be an Expectation-Maximization (ME) algorithm's distributed version. This method has offered a flexible trade-off between the attained approximation's tightness and the cost of connected communication. Labib Khaled and Rao Vemuri. (2004) presented a method that permitted the BS access clarification by presenting a distributed technique for evaluating the PCA method. According to the intermediate nodes' broadcast workload, this work was worked. The intermediate node's incoming data packets combined by the PCA method into one packet. Consequently, an intermediate node has transferred data packet easily instead of communicating every incoming packet. Thus, considerable reduction has been attained in the process of data transmission. By using the numerical simulations, the performance of this work was estimated. The transformations were discovered by the PCA method that was sparsified the signal that has been needed for CS to recover with the high-quality estimate and the original signal is obtained from a few examples. This method was adapted enthusiastically to real-world signals via the online evaluation of their association elements in time and space. Consequently, these were employed by the PCA method to obtain the conversion for the CS process. The following table 2 explain the various unsupervised Machine Learning algorithm implementation for network security and drawbacks.

Table 2: Unsupervised Machine Learning Methods in network security

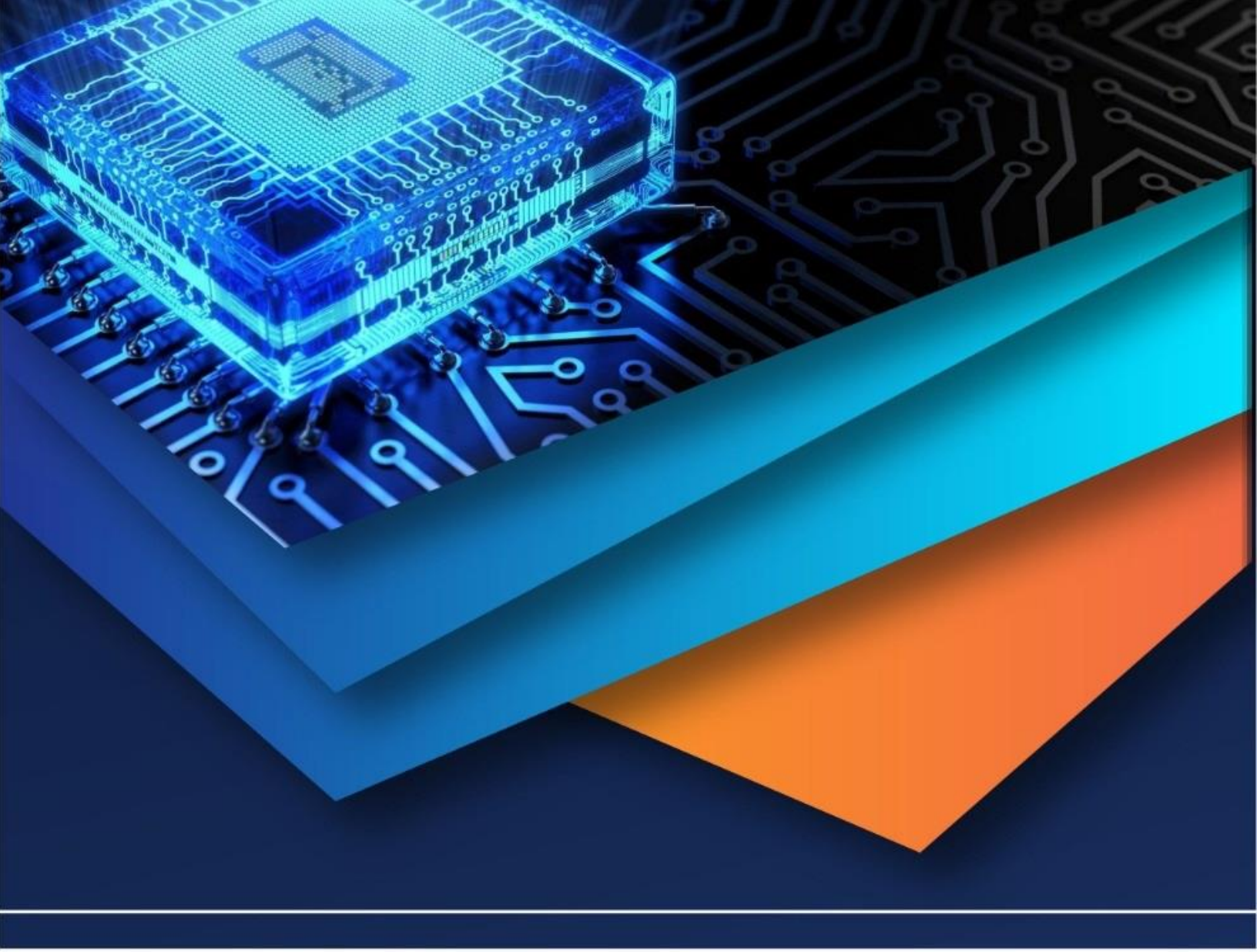
s.no	Author	Algorithm	Advantages	Disadvantages
1	Tian Li and Wang Jianwen (2009)	K-Means Clustering	K-means clustering can generalize to clusters of various shapes and sizes, such as elliptical clusters.	1) No guarantee for K-means can converge into an optimal solution for the clustering of sensor nodes. 2) Cluster number may not be determined precisely and automatically, therefore it is required to be set cluster number depending on the user input. 3) In some rare cases, K-means can give an empty cluster, because of the random initial centroids assortment.
2	Badis Hammi et al. (2014)	Principal Component Analysis	1) Existing PCA based methods can give dimensionality reduction for the data compression processing during the data transmission. 2) PCA reduces the amount of transmitted data among sensor nodes by finding a small set of uncorrelated linear combinations of original readings.	The existing methods are not suitable for highly dynamic environments as they require a long time to learn optimal routes.

III.CONCLUSION

The review of network security methods and algorithms in WSN was discussed in terms of machine learning algorithms. It has been discussed about a detailed review in the area of supervised learning methods such as neural network, decision tree, and SVM, unsupervised learning methods such as K-Means clustering and PCA. Finally, the review of the network security has been analyzed that many effective routing, clustering, classification, and query methods were utilized in this paper for localization and anomaly detection in WSN. In existing works, routing security and rate control. In recent years many techniques are available for intrusion detection. There is a much research scope involved for the research community in this field to find the right kind of methods of the IDS model on cloud computing environment. Optimization algorithm based intrusion detection is considered and suggested to improve its performance.

REFERENCES

[1] Aysa, M. H., Ibrahim, A. A., & Mohammed, A. H. (2020, October). IoT DDOS attack detection using machine learning. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-7). IEEE.
[2] Badis, H., Doyen, G., & Khatoun, R. (2014, May). Understanding botclouds from a system perspective: A principal component analysis. In 2014 IEEE Network Operations and Management Symposium (NOMS) (pp. 1-9). IEEE.
[3] Guleria, K., & Verma, A. K. (2019). Comprehensive review for energy efficient hierarchical routing protocols on wireless sensor networks. Wireless Networks, 25(3), 1159-1183.
[4] Gupta, S., Singhal, A., & Kapoor, A. (2016, April). A literature survey on social engineering attacks: Phishing attack. In 2016 international conference on computing, communication and automation (ICCCA) (pp. 537-540). IEEE.
[5] Hu, J., Ma, D., Liu, C., Shi, Z., Yan, H., & Hu, C. (2019). Network security situation prediction based on MR-SVM. IEEE Access, 7, 130937-130945.
[6] Kang, M. J., & Kang, J. W. (2016). Intrusion detection system using deep neural network for in-vehicle network security. PloS one, 11(6), e0155781.
[7] Labib, K., & Vemuri, V. R. (2004, June). Detecting and visualizing denial of-service and network probe attacks using principal component analysis. In Third Conference on Security and Network Architectures, La Londe,(France).
[8] Lima, M. F., Zarpelao, B. B., Sampaio, L. D., Rodrigues, J. J., Abrao, T., & Proença, M. L. (2010, September). Anomaly detection using baseline and k-means clustering. In SoftCOM 2010, 18th International Conference on Software, Telecommunications and Computer Networks (pp. 305-309). IEEE.
[9] Lu, R., Zhang, L., Ni, J., & Fang, Y. (2019). 5G vehicle-to-everything services: Gearing up for security and privacy. Proceedings of the IEEE, 108(2), 373-389.
[10] Manvi, S. S., & Tangade, S. (2017). A survey on authentication schemes in VANETs for secured communication. Vehicular Communications, 9, 19-30.
[11] Mechtri, L., Tolba, F. D., & Ghoualmi, N. (2010, March). Intrusion detection using principal component analysis. In 2010 Second International Conference on Engineering System Management and Applications (pp. 1-6). IEEE.
[12] Ponmaniraj, S., Rashmi, R., & Anand, M. V. (2018, September). IDS based network security architecture with TCP/IP parameters using machine learning. In 2018 International Conference on Computing, Power and Communication Technologies (GUCON) (pp. 111-114). IEEE.
[13] Rai, K., Devi, M. S., & Guleria, A. (2016). Decision tree based algorithm for intrusion detection. International Journal of Advanced Networking and Applications, 7(4), 2828.
[14] Rassam, M. A., Maarof, M. A., & Zainal, A. (2012). A survey of intrusion detection schemes in wireless sensor networks. American Journal of Applied Sciences, 9(10), 1636.
[15] Rawat, P., Singh, K. D., Chaouchi, H., & Bonnin, J. M. (2014). Wireless sensor networks: a survey on recent developments and potential synergies. The Journal of supercomputing, 68(1), 1-48.
[16] Sharma, S. K., Pandey, P., Tiwari, S. K., & Sisodia, M. S. (2012, March). An improved network intrusion detection technique based on k-means clustering via Naïve bayes classification. In IEEE-International Conference on Advances In Engineering, Science And Management (ICAESM-2012) (pp. 417-422). IEEE.
[17] Su, M. Y. (2011). Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification. Journal of Network and Computer Applications, 34(2), 722-730.
[18] Swarup, K. S., & Corthis, P. B. (2002). ANN approach assesses system security. IEEE Computer Applications in Power, 15(3), 32-38.
[19] Tan, L., Li, C., Xia, J., & Cao, J. (2019). Application of self-organizing feature map neural network based on K-means clustering in network intrusion detection. Computers Materials & Continua, 61(1), 275-288.
[20] Tian, L., & Jianwen, W. (2009, December). Research on network intrusion detection system based on improved k-means clustering algorithm. In 2009 International Forum on Computer Science-Technology and Applications (Vol. 1, pp. 76-79). IEEE.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY