Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey

Dr. E. Baraneetharan,
Associate Professor & Head,
Department of Electrical and Electronics Engineering,
Surya Engineering College,
Erode, India.

Abstract Machine Learning is capable of providing real-time solutions that maximize the utilization of resources in the network thereby increasing the lifetime of the network. It is able to process automatically without being externally programmed thus making the process more easy, efficient, cost-effective, and reliable. ML algorithms can handle complex data more quickly and accurately. Machine Learning is used to enhance the ability of the Wireless Sensor Network environment. Wireless Sensor Networks (WSN) is a combination of several networks and it is decentralized and distributed in nature. WSN consists of sensor nodes and sinks nodes which have a property of self-organizing and self-healing. WSN is used in other applications, such as biodiversity and ecosystem protection, surveillance, climate change tracking, and other military applications.Now-a-days, a huge development is seen in WSNs due to the advancement of electronics and wireless communication technologies, several drawbacks like low computational capacity, small memory, and limited energy resources infrastructure needs physical vulnerability to require source measures where privacy plays a key role.WSN is used to monitor the dynamic environments and to adapt to such situation sensor networks need Machine Learning techniques to avoid unnecessary redesign. Machine learning techniques survey for WSNs provide a wide range of applications in which security is given top priority. To secure data from attackers the WSNs system should be able to delete the instruction if any hackers/attackers are trying to steal data.

Keywords: Intrusion Detection System (IDS), Security, Wireless Sensor Network (WSN), Attacks, Reinforcement Learning (RL), Denial-of-Service (DoS), Networks, Machine Learning (ML)

1. Introduction

Wireless Sensor Networks (WSNs) are considered aslow power consumption candidates for processing and contemptible in features that are used in different fields to collect information on human activities and behavior, to supervise various natural activities, and so on. Security is the main issue in WSNs. The two major groups areActive and Passive. In Passive attacks, here they are unseen and tap the link above to store data; or remove the performance element of the internet. Broken node, tampering, traffic is some of the types in the passive attack. The essential attacks of the network which attack itself in active attack and the reason for attacks might be this and can also be detected [13]. For some time, the services may be stopped or corrupted because of these attacks. Many types of attacks are grouped as jamming, hole- attacks, Denial-of-Service (DoS), Sybil types, and flooding.The activity of the network is passively or actively achieved. “Intrusion Prevention,” Don’t avoid instruction, then “Intrusion Detection,” will take place.Intrusion Detection Systems (IDSs) gives a few information to other supportive Systems: detection and position of intruder, intrusion instance, type of intrusion, where this intrusion occurs. Such information can be useful in mitigating and remedying the cause of attacks, as more information on an intrusion is provided. So,the detection of intrusion systems is useful in network security.

Wireless Networks are considered as a non-trivial and complicated process for its performance and optimization. WSN has to fulfill all the set of operations. Figure 1 explains the design of WSN with gametheory which helps to counter a variety of intrusions on the network. WSN includes the network requirements as the power resources for a while in which WSN needs to communicate with a centralized or a remote base station to sense the data and for subsequent analysis. WSN mainly concentrates on Quality of Service (QoS) and Network Security to prevent the DoS attack (Denial of Service) with the addition of Intrusion detection. Packet Forwarding, data collection, and target tracking are considered to be integral objectives of the WSN.

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Information Technology & Digital World

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

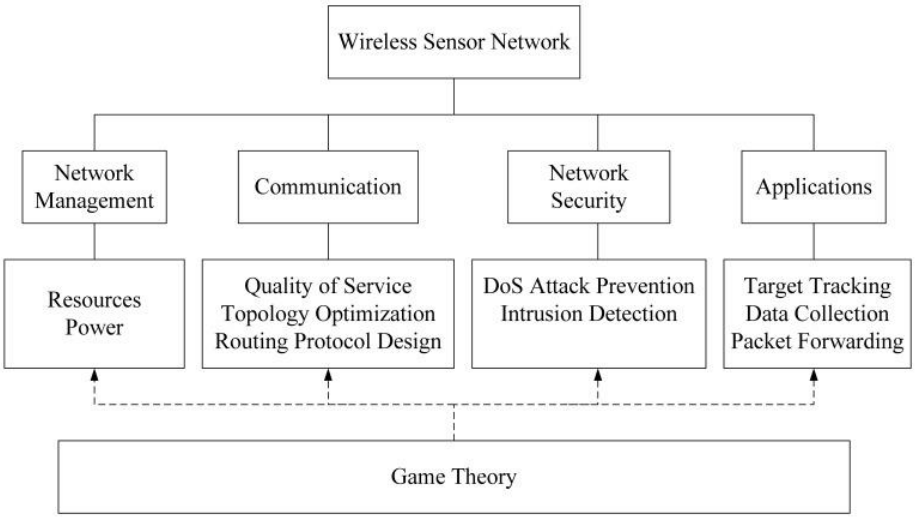

Fig.1 Design of WSN with Game Theory

WSNs have some common security goals[16]. Added to this, they have

(1) Forward secrecy: preventing leak of secret messages when it leaves the internet.
(2) Backward secrecy: prevent decryption of already transmitted messages.
(3) Survivability: Services of some level are in lack of failure.
(4) Freshness: Making sure that the data are new and no one can repeat the old/previous messages.
(5) Scalability: Handling a greater number of nodes.
(6) Efficiency: on sensor nodes like storage, communication border, and processing should not be measured.

Some kind of unapproved Internet / System activity would be known as intrusions. IDS is an identifying and distinguishing tool. A wide system of protection in a network region or unit is not a separate measure of safety. Intrusion is described as "any collection of acts aimed at compromising a resource's integrity, confidentiality or availability." Network intruders are of two types: intruders external and internal. (1) External intruders: An outsider intrusion comes to the network. (2) Internal intruder: A compromised node that is connected to the network. IDS detect both internal and external intruders, but internal intruders are difficult to detect, as they have the requisite keying tools to operate on any mechanism protections [20]. IDS havea partial response to the attacks. Yet, all devices should have complete IDS to detect all of the above intrusions. A deployment-based intrusion detection system (HIDS), network-based intrusion detection system (NIDS), and hybrid intrusion detection system are three types in IDS. A host-based intrusion detection system (HIDS) describes a distributed measure along with the host with some working intent and detects the following: on the host, they modify important data, several breakdowns access during host attempts, random allocation of memory, CPU activity is irregular or I/O operation. We accomplish this by researching log files. The Network-based Intrusion Detection System (NIDS) identifies a packet to be examined; the payload within, the IP address or ports are passive or active via network transmission.

Hybrid Intrusion Detection System describes the combination of NIDS and HIDS to form the hybrid IDS and is well structured by the use of the mobile agent. Mobile agents execute system log files and search anomalies network traffic [3]. IDS is classified based on methodologies of detection such as detection based on the anomaly, detection of misuse, and detection based on requirements. In the detection of misuse, the patterns are found and feed into the system, and behaviour or act of nodes is compared to most frequent patterns of attack. The downside is that they are asking to build up a pattern of attack and they are not having novel attacks. Drawbacks were designed to minimize system management performance, as the network provides IDS with the present database. Anomaly detection uses automated training to define suspicious behaviour and we don't try the perfect attack pattern here, but test node behaviour is usual or anomalous [26]. The IDS is sure that the node is malicious if the sensor node does not behave in a specific protocol; the detection accuracy is due to the false positive and false negative warning. The downside is that it masks the system's actions to be true, which shows the false alarm rate. Specification-based detection pays attention to the fact that there should be no deviations from standard behaviour

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

identified by machine learning techniques nor data mixing abuse of targets and detection mechanisms for anomalies. The downside here is the manual creation of specification, which is time-consuming for humans and cannot detect malicious activity where there is no concept of IDS protocol.

Machine Learning algorithms are mainly used for building accurate models that are specially designed for classification, clustering, and prediction [29]. In this paper, Machine Learning plays a vital role in Intrusion Detection in WSN using some of the Machine Learning algorithms such as Support Vector Machine, Logistic Regression, Random Forest, and Gaussian Naïve Bayes which are used to detect any attacks in the network. Machine Learning concept is very important for WSN applications for the specified reasons such as 1) All the sensor networks are used for monitoring the dynamic environments which change rapidly over time. 2) WSN is built on complicated environments where WSN can be defined using simple mathematical models but it needs a complex algorithm to solve. 3) WSN is used to collect new information about geographic locations. Due to some unexpected reasons, it does not operate properly. 4) New integrations are also possible with Cyber-Physical System (CPS), Machine-to-Machine (M2M) communications, and Internet of Things (IoT) which is used for decision-making. Machine Learning which is used as a different level of abstractions and used to perform AI-related task with human intervention. Machine Learning is classified based on the structure of the model and it is categorized into Supervised Learning, Reinforcement Learning, and Unsupervised Learning. At first, all the ML algorithms are considered as the labeled training data which specifies input, output, and some system parameters. Figure 2 describes the Machine Learning classifications. Supervised Learning is used as classification and regression models. Then the Unsupervised Learning algorithm is used for classifying the sample sets and grouped [24]. The Reinforcement Learning algorithm in which the agent learns by interacting with the environment. The combination of both Supervised and Unsupervised Learning is termed as Semi-Supervised Learning. It is also considered as hybrid algorithms which it inherits all the main functionalities from these categories.

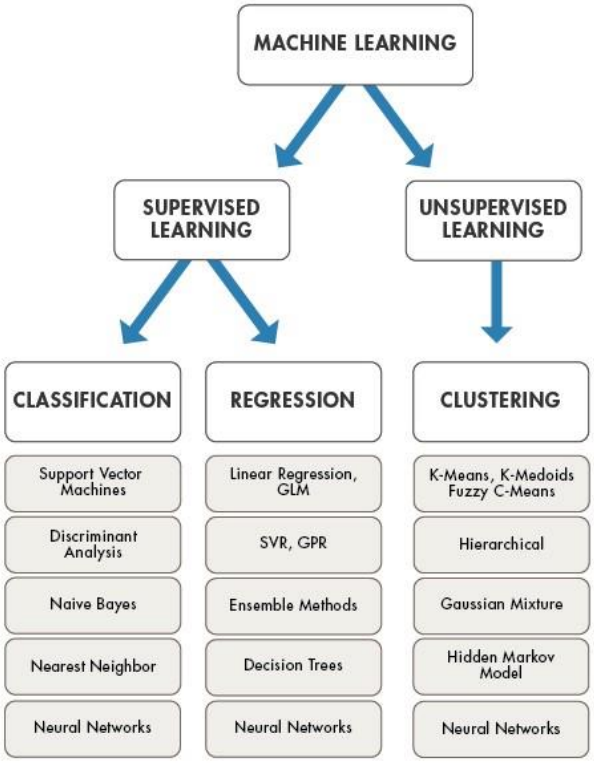

Fig.2 Machine Learning Algorithm Classifications [7]

Noureddine Assad et al propose a model that provides coverage and synchronization in K-sensing sensor network detection which is wireless and provides functionality where geometric analysis and probabilistic model are used. The quality of deployment is determined by rigorous analysis before deployment and quality of sensor deployment is an issue that reflects the cost and capability of the network [25]. However, determining deployment

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

is not that easy. It is necessary to measure the sensing range, transmission range, and node density performance of overall systems. The proposed model of intrusion detection in WSN is proposed to single-sensing / multi-sensing and k-sensing detection connectivity of a WSN. The result gives that the authors design and analyze is unique WSN, which helps to select parameters of the internet to meet WSN requirements.

2. Related Works

Machine Learning algorithms play a major role in different areas and there are many approaches related to IDS in Wireless Sensor Networks (WSN). Although many threats have been occurring in the wireless sensor networks it is necessary to protect the network resources at the early stages. WSN differs from other wired network threats and it is because of the WSN structure and some constraints which possess limited battery life [28]. All the classifications are based on the intruder and intrusion detection techniques, collected data sources and locations, frequency usage. In WSN, the intrusion is considered as stealing the data and creates false data by altering the system that leads to gain access to the system by using an energy-efficient method. In this source, data is considered as Network-Based Intrusion Detection, Hybrid Based Intrusion Detection, and Host-Based Intrusion Detection techniques. Figure 3 explains the IDS in WSN. Based on the location of data it is classified into two divisions such as distributed and centralized IDS. Some of the machine learning approaches are used in WSN are discussed below:

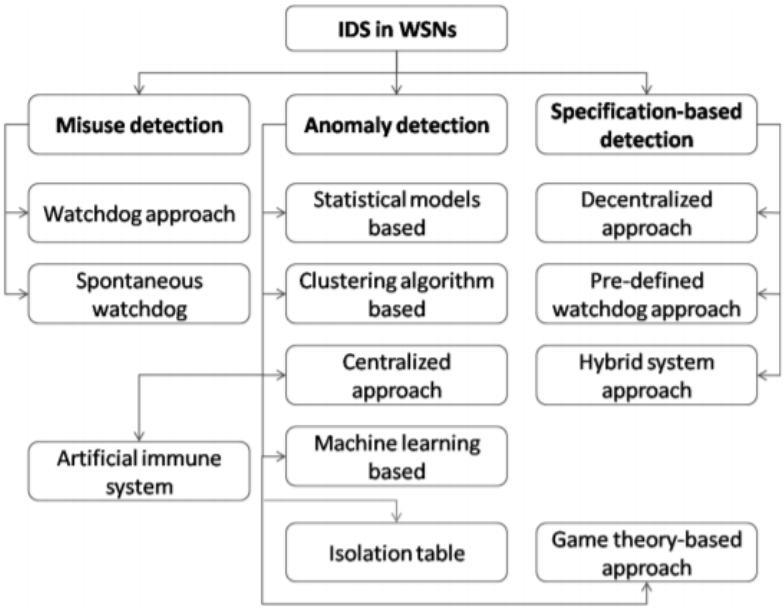

Fig.3 IDSin WSNs [9]

2.1 Anomaly Based Detection Approaches in WSN

The WSN anomalies are of different types so some of the anomalies are grouped as Node anomalies, Network anomalies, Data anomalies, and other anomalies. Mostly network anomalies deal with connection problems in WSN and there are an increment and decrement in the signal connectivity which decides if there is any loss in the network. Node Anomalies is about the software or hardware problems that occur in the sensors. This is mainly due to solar panels' failures and power issues [27]. Data Anomalies mainly occur due to the disorder of data sets and some irregularities caused by the sensor and environmental problems. Other anomalies explain that it is not fitted to any other types of anomalies. Table 1describes the various detection techniques of IDS in WSN.

It is used as a Game-Theoretic Framework for security purposes in WSN and to detect the intrusion detection problem in WSN. This game feature is considered as the main feature in WSN. Then, it uses an Ultra-Wide Band technology which provides low power consumption for wireless connectivity Table 1 discusses the comparison of various detection techniques. A rule-based approach is used to identify the intrusions and the

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

algorithm used is a round-based algorithm. FromData Mining Techniques to IDS in Wireless Sensor Networks (WSN) it uses both anomaly and misuse detection techniques. Then IDS consist of two agents namely,a central agent and several local agents. To identify the intrusion detection activities, it is fixed on the sensors.

Table 1. Comparison of various Detection Techniques [11] [14] [15] [19]

IDS	Statistical models based	Clustering algorithm based	Artificial Immune System	Isolation table	Game Theory based	Machine Learning
Accuracy	Medium	High	High/Medium	Low	High/Medium	High
Energy efficiency	No detail	Yes	No	No detail	No	Yes
Memory Requirement	No detail	High	No detail	Medium	Medium	High
Network Structure	Normal	Clustered	Normal	Clustered	Normal/Distributed	Normal

2.2 Misuse Based Detection Approaches in WSN

To detect the known attacks, misuse detection is used and it is also known as signature-based IDS. The biggest drawback of this approach is that it does not have predefined rules, so any new attacks cannot be identified. Using this technique for WSN is a tedious task, provides less effective and difficult [22]. The Watchdog Based Clonal Selection Algorithm is used for intrusion detection in WSN. It is used to check and detect the nodes whether it has any abnormal behavior during the data forwarding. In WSN, it is responsible for the continuous monitoring of neighboring nodes and transfers the information to the nodes. It mainly affects the WSN performance badly due to this behavior. This algorithm is used to detect the nodes of WSN whether it is malicious or selfish nodes. Then, it also uses distance- vector routing protocol (i.e.,) DSDV protocol for DoS detection and replay attacks. This is not only based on the accuracy but also the robust and the network performance with non- degradability. DSDV protocol has the regular updating process on the routing table and it not only decreases the energy on the nodes, but it consumes some part of the energy as valuable bandwidth.

2.3 Hybrid Based Detection Approaches in WSN

All the security protocols of WSN are defined by the administrator manually. The hybrid approach is a combination of misuse and anomaly detection techniques. Hybrid Detection Approaches is developed by a human with the development of protocol specifications. It can be used as the combined technique or not. This technique is also used for the clustered Wireless Sensor Networks to achieve an accurate intrusion detection system. It also uses a distributed learning algorithm to train the SVM algorithm and to solve the two-class problem for anomaly detection. The main aim of this approach is to save the energy used in the network.

2.4 Clustering Based IDS in WSN

The main usage of this method for making a global decision process and response. The purpose of the method is to save energy for the majority of the nodes by assigning the subordinates that come under the clustering.It is also known as Hierarchical based IDS. Clustering is considered as a single layer of promiscuous monitors. These are used for determining the misbehaviour routing by statistical anomaly detection. The cluster method is used to preserve the resources in which it continuously monitors using intrusion detection as a monitoring agent within each cluster [24]. So, this is implemented on each node for monitoring the local intrusions and also it investigates the response and intrusion sources. This scheme also uses MANET as a centralized and cooperative intrusion detection approach for the cluster method. This Clustering-based IDS is much more beneficial for WSN. It has more energy than other nodes because a greater number of batteries are to be placed on the clustering for a longer life. It is selected periodically for the node which has high energy among other nodes so that node is considered as the clustering node.

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

2.5 Trust Based IDS in WSN

The trust-based IDS is also called Reputation-based IDS which provides cooperation of nodes by monitoring the nodes and assigning grades to them with the respective results. This reputation approach is mainly used to examine the contribution of the member in the network. It is possible when they select the greater number of members with higher member reputation, then it is possible to get a greater number of selected connections that can be done with the other members in the network. So, this means that the member of the particular network can communicate with that particular node when compared to the node which has lower reputations. This made the members of the network encourage the members with an increase in reputations. It also consists of three types of reputations such as Functional Reputations, Indirect and subjective Reputations. In which, Subjective Reputations are used for evaluating the direct interactions between a subject and their corresponding neighbours. Then the Indirect Reputations are done by evaluating the non-neighbouring members of the community.

Functional Reputations consist of both the subjective and indirect reputations which are done by considering all the different functions. It also uses the reputation table for storing a data structure on each node and it is considered as a data structure that also includes the reputation data that is related to that node. DoS attack is a major drawback so they enforced cooperation between the nodes. Based on the analysis, using reputation mechanism DoS attacks can be prevented from selfish nodes. It also uses a DSR protocol which is a reputation-based system that rates the nodes based on malicious behaviour. It possesses the Watchdog mechanism to detect any suspicious activity on the node in the source route and it has alarm messages that come from the nodes and that is also evaluated then the reputation node is also under the investigation. It is updated only when the messages have arrived from the trusted nodes. Suppose when a node is malicious, it sends an alarm message to the other nodes from the list of trusted neighbours.

2.6 Zone-Based IDS in WSN

Zone-based IDS is classified into two types such as Gateway Zones and non-overlapping Zones in which every IDS agent will broadcast alerts inside the zones itself. Gateway Zones are mainly used for the correlation of locally generated alerts and aggregation purposes. Alerts are used to indicate the possible attacks in the zones and alarms are also used for detection. The usage of global aggregation and correlation engine in the gateway nodes used for the aggregate and correlate the detection results from the local nodes for making the final decisions. The Aggregation algorithm achieves lower false positives when compared to IDS. In this model, it is efficient to use Global Positioning System (GPS) receiver in MANET. But it is not more efficient for WSN because the sensor nodes are not installed with GPS due to the high cost and power restrictions.

2.7 Genetic Algorithm Based IDS in WSN

This Genetic Algorithm methodology is very useful for MANET and to run some energy-consuming algorithms. It also uses evolutionary computational techniques to deal with complex MANET environments. Applying both the grammatical evolution and genetic programming techniques to detect the route disruption attacks and ad hoc flooding in AODV. It has a good performance on all the stimulated networks with changes in mobility and some traffic patterns. But it is not feasible for WSN because the sensor nodes have very limited capacity for processing and storage of the data.

2.8 Existing methods used in WSN

Other methods are also used in WSN like watchdog mechanism which is used on the top of DSR protocol to check whether the node forwards the packet to the next node. This method is more efficient when it is used for WSN. Then Hybrid IDS which is used for both ad hoc networks and wired networks [23]. It compares the performance with the other nodes. This IDS requires end-to-end secure communication channels between the nodes for transmitting but it does not available in WSN.

MANET is very useful for handling some tactical networks such as vehicle convoys and infantry troops. Moreover, it aims for developing a solution for intrusion detection in MANET. Then using the data of time series data traffic model is used based on the AutoRegressive Moving Average (ARMA) it is used to predict the traffic

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

in the network. It comes up with the heavy load of traffic data and the data packets in the network are constantly monitored for this it needs a centralized unit for processing traffic data that does not exist in WSN.

3. Wireless sensor network intrusion detection system based on MK-ELM

The Multi Kernel-Extreme Learning Machine based system has been proposed by Weinjie Zhang et al. for the WSN intrusion detection system [5]. The prototype is named to be a hierarchical intrusion detection model. Algorithm grouping of kernel extreme learning machine to amalgamate multi-kernel functions following Mercer property. A multi-kernel extreme learning machine augmented for the WSN Intrusion Detection system by discerning linear combination optimally. The trail and appeal of multi-kernel functions are done, covenants accurate increased detection rate with the decrease in detection time. They are more convenient for resource-constrained wireless sensor networks.

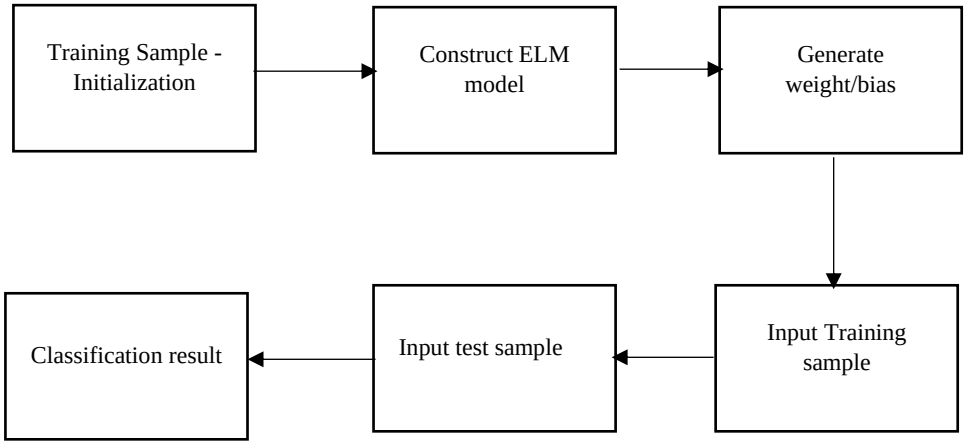

Fig.4 Flowchart of MK-ELM Algorithm

The extreme learning machine is neural networks of single-hidden-layer feedforward. The functions of kernel and output of activation function within the hidden layer are inaugurated into ELM, an overall solution with fast resolution speed [27]. KELM is a solution mixture of machine learning theory and method of caliber optimization due to weak constraints with superior performance. Kernel functions with various traits are correlated to attain the advancement of multi-kernel functions. Figure 4 defines the flowchart of the MK-ELM algorithm. One of such progress leads to a better conduct of mapping. These functions are most referred to as the use ofthe semi-positive specific symmetric function. The structured MK-ELM can be affected by different functions of the kernel. Kernel parameter selection and kernel type function are laboured by its performance classification.

An ideal system is erected by integrating distinct functions of a single kernel, optimal kernel function selection, directing the regulation parameter, and parameter values of the kernel. WSN consists of different nodes like a sensor, cluster head, sink, and management. To know the network operation stability, the wireless sensor network is congregated to undemanding management processes. The sensing nodes are stationed in the tracking area rally to form an assembled network, the information obtained is sent to the sink node by cluster head using relay technique extend to the management node through the Internet. An extreme learning machine with multi-kernel is established to sort out the demerits of single-kernel in terms of IDS. With the intrusion data, data processing is escorted and pre-owned as datasets. Acquired from an algorithm of KELM and the theory of Multi -kernel, the multi-function kernel, and the multiquadric kernel is adopted for the proposal of MK-ELM ID algorithm to WSN infrastructure by designing a hierarchical model.

4. Machine learning algorithms for wireless sensor networks

The Machine Learning techniques survey for WSN was suggested by D. Praveen Kumar et al. This survey presents various algorithms which are ML-based considering the merits, limitations with parameters [2].

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

It additionally debates about ML algorithms and ends the survey with statistical analysis. While system training, supervised learning finds the link between the input and output sets, and a foremost estimation output with its input has prevailed at the process end. One of the methods in supervised learning is regression and using the obtained feature sets, values are predicted. It's a simple methodology forecasts results accurately. A training model is created to augury a class or target centered on decision rules using decision trees [21]. A technique of supervised ML is the Random forest technique provides classification with the trees collection and every tree in the forest. It toils efficiently for huge datasets and even deviates the values which are missed.

A vast number of neurons are coupled with ANN to process information and fabricate accurate results. The deep learning representation method consists of multi-layer representations known to be deep learning approaches. They dispense modules that are simple and non - linear for transformation from the bottom to top layer to bring off the best solution. The support vector machine is a supervised classifier of machine learning, well suited for performing learning tasks relevant to its training instances for large numerals featured in.Using statistical methods, the datasets relationship is found by Bayesian learning. In K-nearest technique concerns the feature space input and classifies based on testing and training samples. The output is not associated with input only considers the data relationships in unsupervised learning. Some clustering techniques are used like K-means clustering, fuzzy-c-means clustering, and Hierarchical clustering [19]. A factorization method of matrix called singular value decomposition does not decrease dimensionality with the product of matrices. Principle and independent component analyses are contemplated for combining information and multivariate estimations. The semi-supervised learning goal is to foretell the labels from the data which are unlabelled. The algorithm of reinforcement learning accumulates information to get hold of actions by interconnecting with the environment.

5. Investigation of Computational Intelligence techniques for IDS in WSNs

The computational intelligence techniques were discussed by McDernott et al. Cyber-attacks are emerging in day to day life will add attacks notable threats of information in confidentiality, integrity, and availability. Dual computational techniques for IDS are examined. They are backpropagation neural networks juxtaposed with the machine classifier of support vector and NSL-KDD dataset for achieving detection rates [8]. The study shows that these techniques are suited for intrusion detection which affords low and high positive rates with the Boolean values. The SVM classifiers are fitted for anomaly detection by its capability illustration with sample sizes. The attacks of cyber in WSN are further classified into active and passive attacks. The passive attacks monitor the quarry for weaknesses and demand information from the network but not lessen the characteristics or alter data in the network. The active attacks ventures to modify the data or change route to target or even ingress to the network.

Hello flood, Sybil, Wormhole, Sinkhole, Selective forwarding, Misdirection, Desynchronization, Collision are some of the active attacks cladding wireless sensor networks. The Intrusion Detection System in scripts undeviatingly by making sure of data disclosure to unauthorized systems, guarantees that data is not modified and ensures by providing security to the system data are approachable to the authorized users [18]. Signature-based and anomaly-based are some types of ID methods. One such method for anomaly-based detection in machine learning is artificial neural networks. From provided samples of input, the output is derived approximately using interconnected neurons by information exchange. Feedforward and Feedback networks are the two ANN architectures. The algorithm of SVM supervised machine learning was used to resolve complex problems and support tasks of classification and regression by handling the listing and continuous variables.

6. Machine Learning in WSN

Machine Learning addresses the issues in WSN. The research was proposed by Mohammad Abu Alsheikh et al. in which the pros and cons are judged against the considered problems [1]. The supervised learning includes such different algorithms. Neighbors K-nearest catalogs the sample data based on the values of output. The decision tree is a classification procedure through the learning tree, input values are iterated for speculating data labels. To attain a specific category in course of this process each property is contrasted to decision conditions. Be one of the learning algorithms is neural networks, they are built by surging bonds of decision elements, and these functions are observed as composite and non – linear [15]. An approach for descrying malicious actions of a node to estimate the secular and structural data collections.

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

By adapting the probability distribution, Bayesian methods can effectually learn uncertain notions. The algorithm aim of unsupervised learning is to consign the dataset samples as various categories by evaluating the unique between them using K-means clustering and Principle component analysis. In reinforcement learning, the agent takes the movements from its occurrences by interacting with the environment [7]. The discussion is briefly about the functional and non - functional challenges in the design of WSNs. The challenge which is considered to be functional is routing and its enhancement using SOM and RL, clustering using neural networks and decision trees, data aggregation, data and query processing, localization, and event detection. Those non- functional challenges are security and anomaly ID using Bayesian networks, K-NN, and forwarding using SVM. In addition to QoS estimation, link quality, provisioning and also accessing accuracy and reliability.

7. ML techniques to solve WSN issues

Some of the WSN issues solved using Machine Learning techniques and the remarks are specified for the issues. There are twelve issues in WSN.They are classified as Localization, Routing, Mobile sink, Event detection, Congestion control, MAC, Coverage and Connectivity, Data aggregation, Energy harvesting, Target tracking, Anomaly and fault detection and Synchronization [2].The first issue of WSN deals with Localization and it uses Reinforcement Learning and k-NN. In Reinforcement Learning, having prior knowledge is not important and it also works for a dynamic environment. Then k-NN is used to calculate the efficient distance estimation for range-free localization. The second issue explains routing which uses Decision tree, Evolutionary computation, and Random forest. This is used to predict the optimal routing paths through dynamic alternative path selection for controlling the data traffic.The third issue explains about Mobile sink which consists of Evolutionary computation, Random forest, and Reinforcement Learning. Evolutionary computation is used to select the optimal sink path between the sensor nodes. Random forest is used to find the data forwarding routes and optimal route selection for large scale networks. In Reinforcement Learning, it is used to select the optimal points and tour selection. The fourth issue is Event detection which consists of PCA, Deep Learning, and ICA. In this, Deep Learning was used for efficient duty cycling management and PCA used for detecting an event from the sensor data.

The fifth issue is Congestion control which has Random Forest, Decision Tree, SVM, PCA, ICA, Evolutionary computation, and Reinforcement Learning. RL is used for predicting the congestion locations in the network and finding the alternate optimal routing paths. Random forest is used to classify the congestion nodes from the normal nodes in the large scale WSN. Evolutionary computation is used for finding the optimal dynamic alternative path selection for congestion avoidance. PCA is used for reducing the dimensions to control unnecessary data transmission. The sixth issue is MAC consists of the Decision tree, Deep Learning, and SVM [17]. Deep Learning is used for reconfiguring the new sensor nodes and predict the time slots. SVM is used for efficient channel assignment.The seventh issue deals with the Coverage and connectivity consists of Deep Learning, Decision tree, and Evolutionary computation. Decision tree used for efficient classification of connected nodes in the network and Deep Learning used for finding the minimum number of sensors.The eighth issueis Data aggregation which consists of SVM, Reinforcement Learning, and k-means. RL is used for identifying the optimal paths without prior knowledge in the network. k-mean used for finding the optimal cluster heads in the network.

Ninth issue Energy harvesting consists of Evolutionary computation, SVM, and Deep Learning. Evolutionary computation is used for predicting the amount of energy to be harvested and SVM used for forecasting the amount of energy to be harvested within the time slot [14]. The tenth issue is about Target tracking consists of Deep Learning, Decision tree, and SVM. Deep Learning was used for efficient multiple target tracking for WSN. SVM for classifying the targets in WSN. The eleventh issue is about anomaly and fault detection consists of PCA, ICA, Random forest,and Deep Learning. In which Random forest is used to classify the fault sensor node from the normal nodes. Deep Learning is used to detect an online anomaly or fault detection. PCA is used for detecting an anomaly in the network. Then the final issue is about Synchronization which consists only Deep Learning and it is used for predicting the efficient time slots for channel allocating and to resynchronize dynamically in the network.

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

8. Energy-efficient learning solution for intrusion detection in wireless sensor networks

It illustrates a new protocol that is simple, non-complex, and less energy-consuming with a self-instructed and differed manner. The dispensed nature comes in the role when an individual node is compromised in terms to circumvent all other nodes being oblate. The concept is about learning automata on the sampling of packets using the mechanism to reach energy apprised of IDS. The performance has been evaluated for the proposed system to obtain the result by mounting different experiments. In wireless sensor networks, the system model for the network is conferred [9]. The concerned protocol is more individualistic of the network structure. All the packets which are considered malicious that pass through the network is dredged by the node. The WSN is liable to various attacks due to the dearth of physical security and its open access to channels.

The main aim is spotting of malevolent data to extract the sway of the network. In the need for a learning system, the invader detects numerous harmful packets that are inoculated into the organization and removes those packets. The algorithm of rate control follows the energy-efficient method by updating the rate samples periodically. The learning automata-based intrusion detection model is presented in the wireless network [11]. The S-LAID works in a disseminated way that the node functions unknowingly about the adjacent nodes. The procedure of monitoring is done for every packet sent by the attacker using a rate control mechanism. The two named functions called reward and penalization are used either to decrease or increase the rate of sampling within the node. By evaluating the system performance in line with the false packets captured percentage or count of detected packets. The computing used for finding the sampling amount exhausted by predicting hostile packets is known to be sample efficiency.

9. Performance evaluation of Supervised machine learning for Intrusion Detection

Intrusion Detection Model is a Predictive model in network security. Machine Learning algorithms are used to build an intrusion detection system. These models are built by using machine learning classification algorithms such as Logistic Regression, Gaussian Naive Bayes, Support Vector Machine, and Random Forest. The Internet is becoming the lifeblood of the modern-day lifestyle. IDS observe the data traffic & identify the inclusion. This is also convenient with advanced network technologies including wireless devices [10]. IDS identify whether data traffic is normal or unusual. Previous data usage has been maintained to identify the Intrusion. Machine Learning-Based Intrusion Detection comes under two categories anomaly & misuse. IDS usea misuse-based method. This method identifies the attacks made & does not identify new attacks. Anomaly-based IDS can detect a new attack. It proposes an integrated for IDS such as Artificial neural networks, Support Vector Machine, and Naive Bayes.

A network intrusion detection system can be designed with different supervised machine learning classifiers. LR is used for the classification of problems. It uses both binary classification, multiclass classification. A support vector machine (SVM), is one of the popular machine learning algorithms. SVM is commonly used for classification and regression analysis of data.SVM is an advanced learning method that categorizes the data based on the nature of the data. In SVM the data is sorted & the maximum differentiation is identified. SVMs are used in text categorization, image classification, handwriting recognition, and in the sciences. A support vector machine is also known as a support vector network (SVN) [12]. Naive Bayes classifiers are based on Bayes' Theorem which is a collection of classification algorithms. It is not a single algorithm but a family of algorithms. "Every pair of features being classified is independent of each other" is the common principle that applies to all the algorithms. Breiman proposed the random forest machine classifier. It works with proximity search. Its main principle is strong learner group is created by a group of weak learners. The standard intrusion detection data set KDDCUP99 didn't give desire results. The advanced version of the same was better in performance with 42 features and 4 stimulated attacks.

The Denial-of-service attack (DoS attack) is a kind of cyber-attack. Its main purpose of the attack is to make the machine or network unavailable temporarily or permanently for the users. This is the disruption of the connection between the host and the end-users on the internet [6].This attack is carried on by flooding the targeted machine or resource with surplus requests. This makes the system or network to overload prevent some or all legitimate requests from being proceeding. These attacks are based on the exploitation of vulnerable users of a network. The hacker starts a normal user account in the network & tries to collect information regarding the

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

system and its users. This information is used to abuse and exploits the vulnerabilities in the system. This gives them a superuser privilege e.g. Perl, xterm. The Probe-response attacks target one or a small number of clients. These kinds of attacks create a unique mark on the intrusion alert of the system or the network. In Remote to local attack (r2l) the attacker tries to gain unauthorized access to a victim machine system in the entire network. Similarly, the user to root attack (u2r) is usually launched for illegally obtaining the root's privileges when legally accessing a local machine.

The Categorical data is converted into numerical form. This pre-processed data divided into testing & training data. Different models are used to predict the labels of test data. Actual label & predicted labels are compared [4]. True and false rates are calculated. Based on the parameter performance of the models are compared. In this attempt, supervised machine learning classifiers are used compared to intrusion detection. Based on the observation random forest classifiers outperform others. This classifier has shown up a promising result.

10. Future Applications of Machine Learning in WSN

Machine Learning techniques have been applied to many research areas [1]. Some of the future applications of Machine Learning in WSN are:

1) Compressive Sensing and Sparse Coding
2) Distributed and Adaptive Machine Learning Techniques for WSN
3) Resource Management Technique using Machine Learning
4) Detecting Data Spatial and Temporal Correlations using Hierarchical Clustering

1) Compressive Sensing and Sparse Coding

To maintain the detection accuracy, sensor measurements are required. It is stated that 80% percent of energy is spent on sending and receiving the data. Two techniques are used to reduce the transmission and extent the lifetime of the network such as data compression and dimensionality reduction techniques.

2) Distributed and Adaptive Machine Learning Techniques for WSN

Machine Learning techniques are well-suited for the limited resources devices like WSN. Though it requires less computational power to process the data [5]. This decentralized learning techniques which enable the nodes to predict and adapt to future behavior with the current environment. Eg. (Adaptive Regularization of Weights).

3) Resource Management Technique using Machine Learning

The main issue is energy-saving in WSN. It is done with two techniques such as the first technique deals with the layers of OSI (Physical, MAC, network layer). The second technique defines decreasing the consumed energy in non-functional and minor requirements.

4) Detecting Data Spatial and Temporal Correlations using Hierarchical Clustering

One of the unsupervised learning algorithms is Hierarchical clustering that designed to build a hierarchy of clusters. This algorithm generates the decomposition of the set of objects (i.e.,) set of sensor nodes. This is an emerging technique in WSN using clustering criteria such as temporal correlations of readings and spatial readings. Hierarchical Clustering provides energy-saving methods. So, only one node from each cluster is activated at a particular time then it covers and monitors the whole cluster area in the network.

References

1. Alsheikh, M. A., Lin, S., Niyato, D., & Tan, H. P. (2014). Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials*, *16*(4), 1996-2018.
2. Kumar, D. P., Amgoth, T., & Annavarapu, C. S. R. (2019). Machine learning algorithms for wireless sensor networks: A survey. *Information Fusion*, *49*, 1-25.

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

3. Maleh, Y., Ezzati, A., Qasmaoui, Y., & Mbida, M. (2015). A global hybrid intrusion detection system for wireless sensor networks. *Procedia Computer Science*, *52*, 1047-1052.
4. Ioannis, K., Dimitriou, T., & Freiling, F. C. (2007, April). Towards intrusion detection in wireless sensor networks. In *Proc. of the 13th European Wireless Conference* (pp. 1-10). Citeseer.
5. Zhang, W., Han, D., Li, K. C., & Massetto, F. I. (2020). Wireless sensor network intrusion detection system based on MK-ELM. *Soft Computing*, 1-14.
6. Alrajeh, N. A., Khan, S., & Shams, B. (2013). Intrusion detection systems in wireless sensor networks: a review. *International Journal of Distributed Sensor Networks*, *9*(5), 167575.
7. Butun, I., Morgera, S. D., & Sankar, R. (2013). A survey of intrusion detection systems in wireless sensor networks. *IEEE communications surveys & tutorials*, *16*(1), 266-282.
8. McDermott, C. D., & Petrovski, A. (2017). Investigation of computational intelligence techniques for intrusion detection in wireless sensor networks. *International journal of computer networks and communications*, *9*(4).
9. Misra, S., Krishna, P. V., & Abraham, K. I. (2010, January). Energy efficient learning solution for intrusion detection in wireless sensor networks. In *2010 Second International Conference on COMmunication Systems and NETworks (COMSNETS 2010)* (pp. 1-6). IEEE.
10. Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. *Procedia Computer Science*, *89*(2016), 117-123.
11. Tan, X., Su, S., Huang, Z., Guo, X., Zuo, Z., Sun, X., & Li, L. (2019). Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm. *Sensors*, *19*(1), 203.
12. Amouri, A., Alaparthy, V. T., & Morgera, S. D. (2020). A Machine Learning Based Intrusion Detection System for Mobile Internet of Things. *Sensors*, *20*(2), 461.
13. Yu, Z., & Tsai, J. J. (2008, June). A framework of machine learning based intrusion detection for wireless sensor networks. In *2008 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (sutc 2008)* (pp. 272-279). IEEE.
14. Abduvaliyev, A., Pathan, A. S. K., Zhou, J., Roman, R., & Wong, W. C. (2013). On the vital areas of intrusion detection systems in wireless sensor networks. *IEEE Communications Surveys & Tutorials*, *15*(3), 1223-1237.
15. Soliman, H. H., Hikal, N. A., & Sakr, N. A. (2012). A comparative performance evaluation of intrusion detection techniques for hierarchical wireless sensor networks. *Egyptian Informatics Journal*, *13*(3), 225-238.
16. Roman, R., Zhou, J., & Lopez, J. (2006). Applying intrusion detection systems to wireless sensor networks. In *IEEE Consumer Communications & Networking Conference (CCNC 2006)*.
17. Loo, C. E., Ng, M. Y., Leckie, C., & Palaniswami, M. (2006). Intrusion detection for routing attacks in sensor networks. *International Journal of Distributed Sensor Networks*, *2*(4), 313-332.
18. Li, G., He, J., & Fu, Y. (2008). Group-based intrusion detection system in wireless sensor networks. *Computer Communications*, *31*(18), 4324-4332.
19. Chen, R. C., Hsieh, C. F., & Huang, Y. F. (2009, February). A new method for intrusion detection on hierarchical wireless sensor networks. In *Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication* (pp. 238-245).
20. Loo, C. E., Ng, M. Y., Leckie, C., & Palaniswami, M. (2006). Intrusion detection for routing attacks in sensor networks. *International Journal of Distributed Sensor Networks*, *2*(4), 313-332.
21. Onat, I., & Miri, A. (2005, August). A real-time node-based traffic anomaly detection algorithm for wireless sensor networks. In *2005 Systems Communications (ICW'05, ICHSN'05, ICMCS'05, SENET'05)* (pp. 422-427). IEEE.
22. Hai, T. H., Khan, F., & Huh, E. N. (2007, August). Hybrid intrusion detection system for wireless sensor networks. In *International Conference on Computational Science and Its Applications* (pp. 383-396). Springer, Berlin, Heidelberg.
23. Yan, K. Q., Wang, S. C., & Liu, C. W. (2009, March). A hybrid intrusion detection system of cluster-based wireless sensor networks. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 1, pp. 18-20).
24. Wang, Y., Attebury, G., & Ramamurthy, B. (2006). A survey of security issues in wireless sensor networks.
25. Raymond, D. R., & Midkiff, S. F. (2008). Denial-of-service in wireless sensor networks: Attacks and defenses. *IEEE Pervasive Computing*, *7*(1), 74-81.

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020

Journal of Information Technology and Digital World (2020)
Vol.02/ No. 03
Pages: 161-173
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2020.3.004

26. Mubarak, T. M., Sattar, S. A., Rao, G. A., & Sajitha, M. (2011, March). Intrusion detection: An energy efficient approach in heterogeneous WSN. In *2011 International Conference on Emerging Trends in Electrical and Computer Technology* (pp. 1092-1096). IEEE.
27. Islam, M. S., & Rahman, S. A. (2011). Anomaly intrusion detection system in wireless sensor networks: security threats and existing approaches. *International Journal of Advanced Science and Technology*, *36*(1), 1-8.
28. Singh, S. K., Singh, M. P., & Singh, D. K. (2011). Intrusion detection- based security solution for cluster-based wireless sensor networks. *International Journal of Advanced Science and Technology*, *30*(83).
29. Jadidoleslamy, H. (2011). A high-level architecture for intrusion detection on heterogeneous wireless sensor networks: hierarchical, scalable and dynamic reconfigurable. *Wireless Sensor Network*, *3*(07), 241.
30. Zhang, Y., Meratnia, N., & Havinga, P. (2010). Outlier detection techniques for wireless sensor networks: A survey. *IEEE communications surveys & tutorials*, *12*(2), 159-170.

ISSN: 2582-418X
Submitted: 01.08.2020
Accepted: 22.08.2020
Published: 29.08.2020