ISSN 2278-3091

Volume 9, No.3, May - June 2020

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcs236932020.pdf

https://doi.org/10.30534/ijatcse/2020/236932020

Support Vector Based Regression Model to Detect Sybil Attacks in WSN

R Vijaya Saraswathi[1], L Padma Sree[2], K Anuradha[3]

[1]Assistant Professor, VNRVJIET, Hyderabad, India. E Mail: vijayasaraswathi_r@vnrvjiet.in

[2] Professor, VNRVJIET, Hyderabad, India. E Mail: padmasree_l@vnrvjiet.in

[3] Professor, GRIET, Hyderabad, India. E Mail: kodali.anuradha@yahoo.com

ABSTRACT

A Wireless Sensor Network (WSN) is a wireless network that includes minute sensor nodes. Sensors monitor physical and environmental conditions. WSNs are well used in military and civilian applications.Also, WSNs being deployed in an unattended area, may get inclined to different types of attacks,leading to harmful effects for the nodes.Sybil attack is one such type of a node that claims illegitimately multiple identities. Legitimate node shares data to the malicious node leading to loss of data. Hence, the proposed work attempts to secure the network with the use of Machine Learning Techniques. The proposed methodology results in a quantitative based machine learning Sybil attack detection methodology to analyze the performance of nodes .

Key words : WSN, Sybil Attack, Machine Learning, SVM, Regression.

1. INTRODUCTION

Wireless Sensor Networks (WSNs) comprise of an enormous number of sensor nodes[15], indistinctly sent over an area, cooperating to monitor and gain information about an environment. Sensor nodes are capable of collaborating with one another with the use of surrounding environment factors such as light, temperature, sound, vibration. The sensed measurements are then converted into digital signals and processed to reveal some properties of the phenomena around sensors. WSN is a wireless network that consists of base stations and many sensor nodes. Following Figure 1. shows the architecture of WSN. Sensor nodes have limited resources regarding energy, transmission range, computation, available bandwidth and memory.

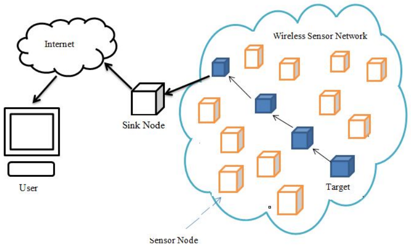

Figure 1: Wireless Sensor Network

They are typically installed in a distant or hostile location and are left unattended to monitor and report.

Therefore, limited resources of nodes need to be used efficiently in order to extend network lifetime and acquire better throughput. These networks will set up in a wide range of applications like health care, environmental monitoring and military surveillance are few to mention. Hence the usage of wireless sensor networks demand the prominence on ensuring security.

The deployment and use of sensor nodes in an unattended area, with the limited resources[17] has let the nodes prone to many malicious attacks[16]. Also, in WSNs, data dissimination has to be done often, and sensor nodes need to be deployed in an arbitrarily surroundings, There is a scope that an attacker opponent can be easily attacked to a WSN. There is a possibility that an attacker may eavesdrop messages, compromise a sensor node, alter the integrity of the data, inject fake messages, and misutilize network resources.

1.1 Sybil Attack

Sybil Attack [14] is one of the most vulnerable attack among many attacks of WSN[13], in which malicious node creates a huge number of fake identities in order to gain an excessively high advantage through a byzantine method. A Sybil node using only one physical device that generates a random number of additional node ,that will identify the disruption of normal functioning of the WSN with the use of multipath routing of multiple disjoint paths between source and destination pairs. Following Figure2 describes an overview of Sybil Attack.

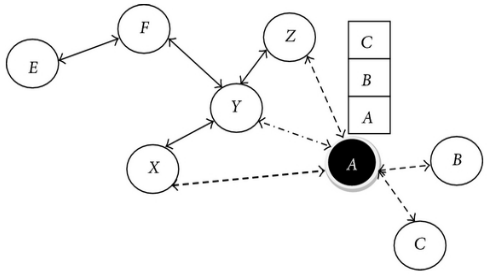

Figure 2: An overview of Sybil Attack

Sybil attack[21] pretences a serious hazard to geographic routing. In this attack, a malicious node attempts to broadcast the following incorrect information regarding its identity, location, and secret key information. Sybil attacks[24] are classified into many types based on their category,usage,resource utility and environment conditions as

i. Direct and In-Direct.
ii. Busy and Idle.
iii. Concurrent and Non Concurrent.
iv. Business executive and Outsider.

i. In the first type, honest nodes are influenced directly by the Sybil nodes or attacked by a corresponding nodes that communicate with preceeding Sybil nodes. In In-Direct case, the attacker communicate with the honest node by using his genuine identity, and then divert the Sybil data to the honest node via the genuine one.

ii. few Sybil nodes will participate in the network while the remaining are in idle.

iii. In a Concurrent attack, all the Sybil identities involve simultaneously or a single physical node can pretend its identities in fixed time slots to appear like all the identities are involved concurrently. An attacker will bring all his identities into the network not simultaneously but slowly over a period of time involving few identities each time in Non- Concurrent attack .

iv. If the attacker is a node in the network and holds at least one genuine identity, then the attacker is called a Business executive, otherwise he is an outsider.

Moreover, the middle node in Sybil attack is to compromised as it's under malicious influence of Sybil node(s).A Sybil node[22] tampers its neighboring nodes for the purpose of converting them as malicious. The network traffic will seriously affect, as the amount of Sybil nodes increase in the network, and the data packets will never reach to their destinations. To address the network traffic issues of Sybil attack[23], many researchers have proposed several schemes to identify Sybil attacks. However, most of proposed works provide an expensive setup of encryption mechanisms to validate the location information.This paper tries an attempt to resolve network traffic issues and results in detection of sybil attacks.

The paper is organised as the following: Section 2 outlines the use and importance of machine learning to handle the WSNs. Section 3 desceibes about related work, 4 illustrates the mathematical model and architecture used for the analysis. Section 5 and 6 illustrates about the algorithmic approach, data set and metrics used to perform the result analysis followed by conclusion.

2. APPLICATIONS OF MACHINE LEARNING IN WSN

Machine Learning[19] is a technique ,that is able learn and improve from pevious experiences without being explicitly programmed. The extensive use of computational statistics,ability to predict based on results and the trained samples has led the tremendous demand of machine learning even in the security domain related research works. Machine learning is used extensively used and well popularised starting from supervised based classification, regression, density estimation, prediction or unsupervised categorisation[11]. Machine learning will also sense and perform better to tackle the performance of sensor networks.

In the domain of security and WSN,Machine learning plays a vital role leading to the protection of nodes against attacks by identifying malicious behaviour not only based on rules but also by identifying abnormal data movement and helps to spot outliers. The use of functions and training samples lets the user to eliminate the need of unnecessary design. The use of Machine Learning maximizes the resource utilization and provides security through out the lifespan of Network.

Machine learning can be characterized as a collection of tools and algorithms that are used to generate prediction models by sensor network designers. The proposed work tries to infer the model out of very large themes and patterns. Existing machine learning algorithms infer the proposed structure of the model. Furthermost, machine learning algorithms as shown Figure 3. broadly categorised into following:

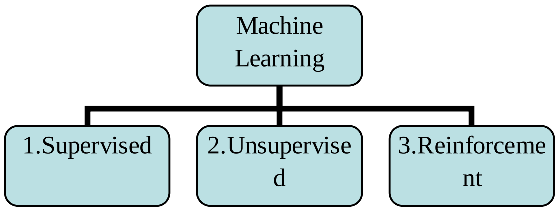

Figure 3: Classification of Machine Based Learning techniques.

1. Supervised Algorithms includes labeled training data sets and system model. It provides relation between the input, output and system parameters, where as **2.*Unsupervised Learning Algorithms*** have no output vector or a specific training sets, but result to group the sample data sets into clusters for a given input data samples. **3.*Reinforcement Learning*** demands the need for maximizing the notion of cumulative reward of indirect function mappings, there by used to solve complex problems. There are many algorithms in machine learning that can be used based on our usage of the problem defined and the type of the analyis, which we do require.However, some algorithms such as Support Vector

Machines[18], Naive Bayes Classifier[25], Clustering techniques, Neural Networks have been well known for their mathematical analysis and their performance.

3. RELATED WORK

Much of the work has been explored in the field of secuirty and cyber attacks with the use of machine learning. Some of the highlighted and benchmark related works are :Mohammad Abu Alsheikh etc.,[1] discuss about issues in WSN's like localization,data aggregation, and enhances performance of WSN .[2] proposed a detection approach to identify DDOS based attack approach by using neural networks. Yair Meidan, Michael Bohadana etc.,[3] classified the infected IoTs like Heterogeneity tolerance, open world, efficiency. The Methods and materials of Sybil Udaya Surya etc.,[5] succeeded in classifying network Model for Sybil attack to configure, energy efficient, promising nodes in the network. Palak etc.,[6] resulted in classification of various Sybil attacks. Ehsan Ullah Warriach etc.,[8] proposed Hidden Markov Models to handle sybil attacks[].Sunil Ghildiyal etc.,[12] discussed about limited processing capability of nodes in WSN. [4] R.Amuthavalli et al uses random password comparision method to detect and prevent Sybil attack. Salah Zidi etc., [9] proposed support vector machines (SVMs) classification method to fault detection based on statistical learning theory. Sona Malhotra et al., [7] proposed a key management protocol that prevent the Sybil attack using cryptography. [20] M. Al-Qurishi, et al., proposed a model to predict a Sybil attack using deep-regression model.

We propose a approach of machine learning to detect sybil attacks from genuine network traffic in centralized wireless sensor networks.When a possibly new infected data of network traffic triggers, our proposed methodogly will be able to detect anomalies.

4. MATHEMATICAL ANALYSIS AND ARCHITECTURE

The main philosophical approach used in machine learning is getting trained to understand sequence and pattern of continuous network packets, between attackers_traffic and valid_nodes. The current session illustrates the mathematical analysis in real world scenarios where faults [10] suchs as offset_faults, stuck-at faults etc., are predicted. The training sample results in the capability of a sensor node to get differentiate between a gain_fault, g_f and a fire node, f_n. Mathematically, For any organization 'O', following name conventions may be used as A_u: {Set of n' unconventional signals}

Xi: $\{X_1,X_2,...X_n\}$--[1]

and a set of 'm' attack types as

$A_t = \{A_1,A_2,...A_m\}$--[2]

The objective of proposed model has to forecasts the occurrences of future cyber incidents against 'O', by performing binary prediction for an attack A_u. Hence, diagnosis of WSN plays an important role to determine the efficientcy of network, there by restricting the loss.

The current session illustrates mathematical analysis in the following steps. For a given attack type 'A_t',

i)if the j^{th} observation of a signal X_i is missing,

ii)Previous (j−1) observations of the remaining signals in '$X_{i'}$ and the entity ground truth for 'A_t' will be considered as features to train our proposed model with SVM, to estimate the value for X_{ji}.

iii)The proposed algorithm calculates and imputes the missing signals for a given data set{'D'}.

iv)Next, model 'g_p' is predicted.

v)The target function 'y' is then deployed in the cluster head to classify the data nodes.

$Y = g_p(W^Tx+b)$-------------------------------------[3]

where, $g_p(z) = 1$ if $z \geq 0$,

-1 otherwise,where 'y' is the output,'$g_p(z)$' is the function,used to predict the nodes.

vi)The class labels in SVM are denoted by,

$y \epsilon \{-1,1\}$---[4]

-1 for negative class

1 for positive class

The proposed system is modeled as a result of linear function and computations, analysed in a high dimensional feature area covered by the neighboring nodes. The use of algorithmic rules, bias function b_f has let our prediction model to successfully identify and classify the faults as persistent faults and transient faults. Also the use of parameter values like baisoffset, gain ratio, stuck-at, out of bounds, spike value, and data loss fucntions had made our model to classify genuine nodes and sybil nodes for a particular span of time.

5. ALGORITHM AND RESULT ANALYSIS

One such approach named SVR has been into use, to predict temporal and spatial correlations for the data nodes in WSN. To take it further, observations in WSN are considered as dimensional points in the feature space. Support vectors are data points that are bounded by a hyperplane and a marginal area, covering and influencing position and orientation of the regression line that is observed for a given data points.The possible margins, also called as separation gaps, along with a new reading will be enough to classify our solution, as shown in the following Figure 4.

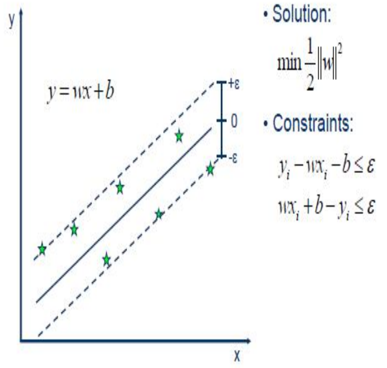
y
x
$y = wx + b$
+ε
0
-ε
• Solution:
$\min \frac{1}{2}\|w\|^2$
• Constraints:
$y_i - wx_i - b \leq \varepsilon$
$wx_i + b - y_i \leq \varepsilon$

Figure 4: Lay out of the Support Vector Regression Model

Next, moving towards the SVR algorithm, our model makes use of optimised quadratic function with linear constraints. Our main intention is to reduce the residue error, which maximizes the hyperplane margin area. Following loss_function is used to determine the gain_fault and loss_function as

$$L_F(X,A) = f(X) + A_i.g_f(x) \text{------------------------} [5]$$

a function of (n+ m) variables 'n' for the X's, 'm' for the A_i. The 'n' equations are differentiated with respect to each X_i to give the gradient conditions, g_{ci}. Next, 'm' equations are differentiated with each A_i to recover the constraints g_c .
Gradient min of function 'f' constraint condition 'g_f', is given as:

$$f(x): ½\| Wi\|^2 \text{-------------------------------------} [6]$$

and there by

$$g_f(x) : Y_i (W•X_i +bias)–1 = 0 \text{------------------------------} [7]$$

Inorder to minimize the error, We need to find the minimum lossfunction_rate as

$$M_{Lfr} = ½\|W_i\|^2–ΣA_i[Y_i(W_i•X_i+b)–1] \text{----------------------} [8]$$

with respect to w_i, bias. Hence, we are using the min_function in our algorithm. Finally,We expand the above equation to get the following final_L form:

$$F_L=½\|W_i\|^2–ΣA_{i.}Y_{i.}(W_i•X_i+bias)+ΣA_i \text{---------------------} [9]$$

with respect to w_i, bias.

Algorithm*:* For the proposed methodology

```
#For Inputs and initailization purpose:

1. Include Data_Set, '$D_s$'
2. Partition the Data_Set $D_s$ into :
   Training_Set($t_s$) ,Suppot_Set($s_s$) ,
   Error_Set($e_s$) and Remaining_Set($r_s$)(for
   testing purpose)
3.Training_Set { $x_i$,  i=1..n}
4.Initialize Weights $q_i$, i=1..n
5.Specify Bias 'b'
6.Include params: e , c , kernel type k and
  kernelparams kp
7. Specify the matrix  $W_m$.

   #For Computation:
8. $S_{new}$ = ($x_{new}$, $y_{new}$), #The model is the pair
   wise of input 'x', output 'y'
   #For Training and Testing:
9.Compute f($x_{new}$) and h($x_{new}$)
10.1 If (|h($x_{new}$)| < e) #Error_function_value
   10.1.1 Add New_Sample_Set, ($S_{new}$) To
          the Remaining_Set($R_s$) and Exit.
11.Compute h($x_i$), i=1..n
12.While ($S_{new}$  is not added into a set)
12.1. Update the values b and g
12.2. find least variations ($l_{c1}$, $l_{c2}$, $l_s$, $l_e$, $l_r$)
12.3. find min variation min_var = min($l_{c1}$,
      $l_{c2}$, $l_s$, $l_e$, $l_r$)
12.4. update h($x_i$)# ,h is a function
      #For Output:
13. Prediction_model g($W^T$x+b)
14. New_training_Set { $x_i$, $y_i$, i=1..n+1}
15. New_coefficients $q_i$, i=1..n+1
16. New_bias $b_{new}$
17. New_training_Set_partition
18. Update New_matrix,$Q_{new}$
```

SVR results in a linear regression model, predicting a target value testing set data instances. Proposed regression methodology determines loss function based on the relationship between variables obtained from the training datasets and thereby mapping it to the final testing sample set.
The use of loss function is explained in Section IV.With the use such marginal support vectors, we can there by maximize the margin of the classifier resulting in the increase of accuracy in the classification of the nodes. The kernel approach deploys the course of dimensionality feature space. Follwing are the main stages, we need to follow:

i. Collection of Data
ii. Extraction of features
iii. Data normalization
iv. Training an anomaly detector
v. Model deployment
vi. Testing for a new attack
vii. Continuous monitoring.

To start with, raw network traffic data is taken, preprocessed and considered for the algorithm traning. Next relevant features are extracted from the network traffic such as packet rate, protocol type, packet id. Next, data normalization is done to define a common framework for the data set. Algorithmic approach and mathematical analysis is explained in the session 4 and 5. The use of gradient functionand partial derivative functions and operations on matrices made our work easy to analyse the predictable output function, 'y'. The work is analysed by considering bench marking data set "N-BaIoT: Data for network based detection of IoT botnet attacks".

Table 1: Metric Assessment Obtained from the Proposed Model

Type of attack name	Metric assessment obtained from the model		
	precision	recall	f1-score
Fake nodes	0.8501	0.9801	0.9101
Genuine nodes	0.9803	0.8503	0.9102
Micro average	0.9102	0.9104	0.9111
Macro average	0.9200	0.9112	0.9121
Weighted average	0.92001	0.9124	0.9123

Pseudo codes of Normalization and feature extraction of the attributes gave a good essence for a clean outcome. Some of the attributes are Node_ID, Time, Is_CH, CH_ID, RSSI, Avg_Dist_CH, Dist_CH, Energy_Consp etc.The experimental results and analysis is done in Python, version3 platform.The easy accessible imported files, packages and library files of python ,implementation of the algorithm and the analysis seems to be satisfactory in terms of accuracy.

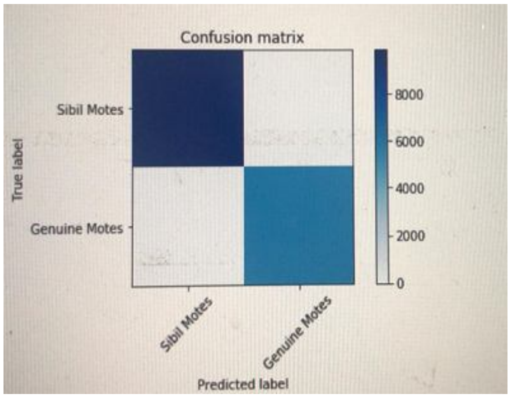

Figure :5.a Confusion Matrix

Further, to document and display the results assessment metrics like mean, standard_deviation, median, micro, macro, weighted average, recall, precission are determined. The above Table I illustrates a sample scenario of the results. Further , confusion matrix is determined, to describe the possible outcomes of the sybil nodes and genuine nodes detection. Following Figure 5.a. is the sample illustration of the confusion matrix.

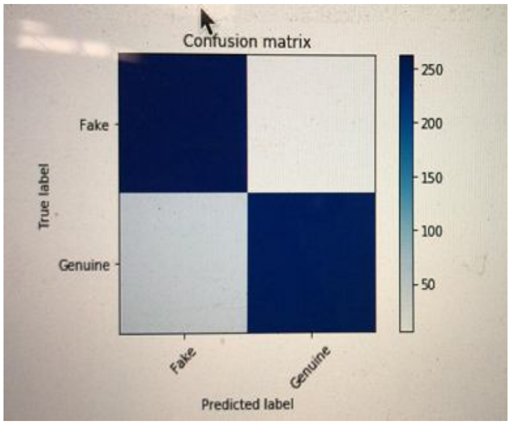

Figure : 5.b Normalized Confusion matrix

.

Figure 6 shows the illustration of the accuracy score obtained from the training samples and the est score.

Finally, Receiver Operating Characteristic, ROC Curve is shown in Figure 7 is determined as a result of the analysis.

Illustration of ROC with an accuracy of 91 percentage. Results observed out of experimental setup, clearly indicates that SVR based approach performs orders of magnitude better than classical Sybil algorithms. The proposed model proves to be more flexible in handling noise levels generated in nodes as well in the network nodes.

Further, to document and display the results assessment metrics like mean, standard_deviation, median, micro, macro, weighted average, recall, precission are determined. The above Table I illustrates a sample scenario of the results. Further , confusion matrix is determined, to describe the possible outcomes of the sybil nodes and genuine nodes detection. Following Figure 5.a. is the sample illustration of the confusion matrix.

false positives and true positives for given nodes are shown in Figure 5.b.

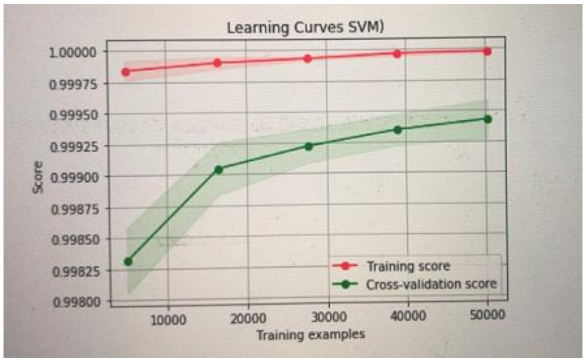

Figure : 6.Accuracy Score

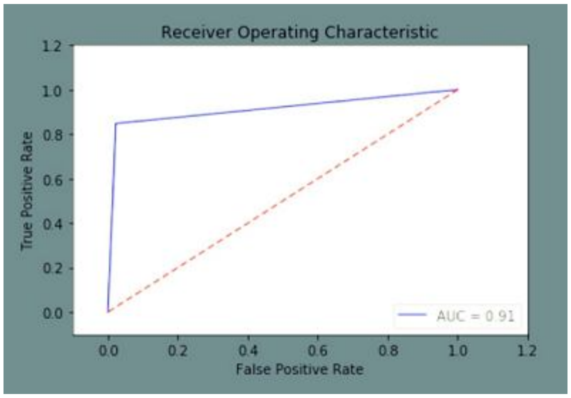

Figure :7 ROC Curve

6. CONCLUSION

A well performed and rigid Sybil attack detection scheme to predict the status information of the nodes is proposed with the use of support vector regression model. One of the differentiation factors of the proposed work results in the detection of Sybil attacks that rely on the trained on statistical features extracted from network traffic analysis data. Regression provides the prediction of network packets, compare them with the normal ones over bidirectional links, and availability of beacon nodes.

REFERENCES

[1] Mohammad Abu Alsheikh, Shaowei Lin, Dusit Niyato and Hwee-Pink Tan, ”**Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications”,**in *IEEE communication surveys & tutorials, vol. 16, no. 4, fourth quarter,2014.*

[2] Xiaoyong Yuan, Chuanhuang Li, Xiaolin Li, ”**DeepDefense: Identifying DDoS Attack via Deep Learning”**, in *IEEE International Conference,2017.*

[3] Yair Meidan, Michael Bohadana, Yael Mathov, Yisroel Mirsky**,”N-BaIoT: Network-based Detectionof IoT Botnet Attacks Using Deep Autoencoders”,***Published by the IEEE Computer Society,2018.*

[4] R.AMUTHAVALLI,DR.R.S.BHUVANESWARAN, “ **Detection and prevention of sybil attack in wireless sensor network employing random password comparison method ”**, in *Journal of Theoretical and Applied Information Technology,2014.*

[5] Udaya Suriya Raj Kumar Dhamodharan and Rajamani Vayanaperumal,” **Detecting and Preventing Sybil Attacks in Wireless Sensor Networks Using Message Authentication and Passing Method”**, in *Hindawi Publishing Corporation e Scientific World Journal, 2015.*

[6] Palak, **”Review on the various Sybil Attack Detection Techniques in Wireless Sensor Network”**, in *International Journal of Computer Applicati.ons,Volume 164 – No 1, April* ,2017.

[7] Sona Malhotra, Pankaj Rathee, **” Prevention of Sybil Attack using Cryptography in Wireless Sensor Networks”,** IJIRST –International Journal for Innovative Research in Science & Technology,2015.

[8] Ehsan Ullah Warriach, Kenji Tei, **”Fault Detection in Wireless Sensor Networks: A Machine Learning Approach”**, 2013, published in 2013 IEEE 16th International Conference on Computational Science and Engineering.

[9] Salah Zidi, Tarek Moulahi, and Bechir Alaya**,”Fault Detection in Wireless Sensor Networks Through SVM Classifier”**, published in IEEE SENSORS JOURNAL, VOL. 18, NO. 1, JANUARY 1, 2018.

[10] Thaha Muhammed, Riaz Ahmed Shaikh **,”An analysis of fault detection strategies in wireless sensor networks”** published in Journal of Network and Computer Applications.

[11] Uday Shankar Shanthamallu, Andreas Spanias, Cihan Tepedelenlioglu, and Mike Stanley, **“A Brief Survey of Machine Learning Methods and their Sensor and IoT Applications”,** published in 8th International Conference on Information, Intelligence, Systems & Applications,2017.

[12] Sunil Ghildiyal1, Ashish Gupta, Nitesh Tomar, AnupamSemwal, **“Analysis of Sybil Attack in Wireless Sensor Networks”**, International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 5,pp.845-848. May – 2014.

[13] R Vijaya Saraswathi, Dr. L Padma Sree , Dr. K Anuradha **“Multi-stage Key Management Scheme for Cluster based WSN”** in International Journal of Communication Networks and Information Security (IJCNIS) Vol. 10, No. 3, December 2018.

[14] Neil Zhenqiang Gong **“SybilBelief: A Semi-Supervised Learning Approach for Structure-Based Sybil Detection”** IEEE Transactions on Information Forensics and Security Volume: 9 , Issue: 6 , June 2014 .

[15] R. Vijaya Saraswathi, L. Padma Sree and K. Anuradha **“Secured Cluster-Based Distributed Dynamic Group Key Management for Wireless Sensor Networks”** © Springer Nature Singapore Pte Ltd. 2020 M. Pant et al. (eds.), Computational Network Application Tools for Performance Management, Asset Analytics.

[16] R Vijaya Saraswathi, L Padma Sree and K. Anuradha **“Key Management Schemes in Wireless Sensor Networks: A Survey”** in CiiT International Journal of Wireless Communication, Vol 8, No 05, May 2016.

[17] R Vijaya Saraswathi, L Padma Sree and K. Anuradha **“Dynamic and probabilistic key management for distributed wireless sensor networks”** in 2016 IEEE International Conference on Computational Intelligence and Computing Research(ICCIC 2016).

[18] Taeshik Shon’, Yongdue Kim, Cheolwon Lee’, and Jongsub **Moon “A Machine Learning Framework for Network Anomaly Detection using SVM and GA”** Proceedings of the 2005 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY.

[19] Pengwenlong Gu, Rida Khatoun, Youcef Begriche, Ahmed Serhrouchni **“Support Vector Machine (SVM) Based Sybil Attack Detection in Vehicular Networks”** IEEE International Conference, 2017978-1-5090-4183-1/17/$31.00 ©2017 IEEE.

[20] M. Al-Qurishi, et al., **“A prediction system of Sybil attack in social network using deep-regression model”** in Future Generation Computer Systems (2017).

[21] S. T. Patel and N. H. Mistry, **"A review: Sybil attack detection techniques in WSN,"** 2017 4th International Conference on Electronics and Communication Systems (ICECS), Coimbatore, 2017, pp. 184-188, doi: 10.1109/ECS.2017.8067865.

[22] A. B. Karuppiah, J. Dalfiah, K. Yuvashri, S. Rajaram and A. K. Pathan, **"A Novel Energy-Efficient Sybil Node Detection Algorithm for Intrusion Detection System in Wireless Sensor Networks,"** 2014 3rd International Conference on Eco-friendly Computing and Communication Systems, Mangalore, 2014, pp. 95-98, doi: 10.1109/Eco-friendly.2014.94.

[23] Joseph Kamel, Farah Haidar, Ines Jemaa, Arnaud Kaiser, Brigitte Lonc, et al.. **“A Misbehavior Authority System for Sybil Attack Detection in C-ITS”.** In The IEEE 10th Annual Ubiquitous

Computing, Electronics & Mobile Communication Conference – IEEE UEMCON 2019, Oct 2019.

[24] Rincy Medayil John Jacob P. Cherian Dr. Jubilant J Kizhakkethottam **“A Survey of Techniques to Prevent Sybil Attacks”** in 2015 International Conference on Computer *Communication and Informatics (ICCCI -2015), Jan. 08 – 10, 2015, Coimbatore, INDIA.*

[25] Saurabh Mukherjeea , Neelam Sharma **“Intrusion Detection using Naive Bayes Classifier with Feature Reduction”** 2212-0173 © 2012 Published by Elsevier Ltd.