International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 11, Issue 12, December 2020, pp. 3348-3356, Article ID: IJARET_11_12_315
Available online at https://iaeme.com/Home/issue/IJARET?Volume=11&Issue=12
Journal Impact Factor (2020): 10.9475 (Calculated by GISI) www.jifactor.com
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: 10.34218/IJARET.11.12.2020.315

CLASSIFICATION APPROACH-BASED SYBIL NODE DETECTION IN MOBILE AD HOC NETWORKS

Dr. S. Rethinavalli
Assistant Professor of Computer Science, Shrimati Indira Gandhi College, Tiruchirappalli, Tamil Nadu, India

Dr. R. Gopinath
D.Litt. (Business Administration) - Researcher, Madurai Kamaraj University, Madurai, Tamil Nadu, India

ABSTRACT

Mobile Ad Hoc Network (MANET) is an auto-configuring network that is designed spontaneously by a mix of mobile nodes without the intervention of a centralised administration or fixed infrastructure, thanks to technological advancements. In MANET, Intrusion Detection Systems (IDS) must system blocks of packets with a variety of attributes that prevent anomalies from being detected. Sampling and Feature Selection can be used to reduce computing time and hence reduce the time it takes to detect intrusions. Those that create assaults on network, data link, application layer, and physical layer functioning are selfish and nasty. In this paper, three feature selection strategies and a suggested Artificial Neural Network (ANN) classification model are proposed to improve the classification of Sybil nodes in the MANET. The accuracy, precision, and recall of the Sybil node identification architecture, as measured by numerous evaluation measures. Other classification techniques, such as Support Vector Machine, are used to evaluate the proposed ANN architecture (SVM).

Key words: Mobile Ad Hoc Network, Intrusion Detection System, Artificial Neural Network, Feature Selection, Classification, Support Vector Machine

Cite this Article: S. Rethinavalli and R. Gopinath, Classification Approach-Based Sybil Node Detection in Mobile Ad Hoc Networks, *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 11(12), 2020, pp. 3348-3356.
https://iaeme.com/Home/issue/IJARET?Volume=11&Issue=12

1. INTRODUCTION

When a framework does not exist or is difficult to set up, mobile ad hoc networks [1] are vital. They're useful for disaster recovery, hunting and saving in remote areas, battlegrounds, patient

monitoring, Bluetooth, sensors, typhoon evolution analysis, attrition rate earthquake identification, interactive museums or toys, providing security in public buildings, and identifying an object. Clients can communicate across mobile ad hoc networks without the need for a physical infrastructure. It is required to design an impermanent system that does not require wires, a communication framework, or administrative intervention [2][3] [17].

To expedite data packets in MANETS, each node acts as a router and host. The nodes can move in a haphazard manner. Self-organizing competence for a healthy version of active conditions and interoperating proficiency amongst the nodes [4] [5] are distinctive and significant networking elements in an expanded size of administrative condition.

1.1 Sybil Attack (SA) in Mobile Ad Hoc Networks

The Sybil attack was first introduced by J. R. Douceur. The Sybil assault, according to Douceur, is an attack in which a single individual manages a critical section of the organisation by assuming several identities. Furthermore, a Sybil assault can be thwarted by a simple display of several attributes intended for a separate dynamic node. The SA can occur in a distributed network without a central administration to investigate the uniqueness of each cooperating object. The proportions of Sybil attacks are described in the table below.

- ***Communication:*** The types of communication are as follows.
  - Direct communication – Sybil Node (SN) was able to engage with legal nodes in the network thanks to Malicious Node (MN). (Malicious node to Sybil node to Legitimate node).
  - Indirect communication – The malicious node prevents SN from passing freely across the network's genuine nodes. (Malicious node->sybil node->malicious node).
- **Participation**
  - Simultaneous participation - At the same time, the MN lifts off all of the bogus identities (Sybil node).
  - Non-simultaneous participation — After a fixed or variable interval of time, the MN reveals the SN one by one.
- **Identify**
  - Stolen Identity - SN can snip the singularity of an authentic node by replicating it in this process.
  - Fabricated identity refers to the fabrication of a new false identity.

2. RELATED WORKS

Umar *et al*. [6] In cluster-based MANETs, a new mechanism using cluster heads and cryptographic keys for each node was hosted. Cluster heads are chosen based on a set of principles shared by the nodes in the clusters [18]. Cluster chiefs hold the isolated keys for decryption, while each member node is given a unique public key to use for encryption. The strategy is particularly flexible since it allows nodes to migrate to different clusters and assign regularity from one cluster head to another.

Borkar, Gautam M., and A. R. Mahajan [7] To evaluate this technique, a standard ad hoc on-demand multi-path distance vector protocol is used as the base routing protocol. The Dolphin Echolocation Algorithm is used in the proposed mesh-based multipath routing method to identify all possible secure paths with secure adjacent position trust verification protocol and healthy link optimal path catch for competent communication in MANET.

Ramakrishnan *et al.* [8] In the VANET, a cluster-based solution for emergency message distribution and pile-up avoidance has been proposed. In this study, the cluster is first created

in such a way that any form of collision is avoided. The responsibility of intra-cluster organisation follows the designation of the cluster head to avoid encroachment between the clusters. Two MAC layer transmission protocols are used to improve consistency during the distribution of emergency messages.

Arain *et al*. [9] introduced a clustering-based energy efficient and communication protocol (CEECP) for a variety of mix-zones throughout road networks, which is expected to close loopholes in existing clustering methods. Furthermore, we propose an unique CEECP for chain situations that connects with roadside devices to provide V2V and V2I communication for Cooperative Traffic Information Systems. The system's coverage possessions are responsible for the constancy time. It examines the rises in the consistency period as well as the network's reliability and application of Machine Learning Techniques [26].

Das *et al.* [10] An optimised energy efficient routing (OE2R) approach was intended. Artificial intelligence technologies such as geometric programming, multi objective optimization, ambition level, and tolerance limit accelerate this process. The integration of these quantified artificial intelligence technologies produces an operational device that can supplement a variety of conflict goals and estimate erroneous system restrictions in real time. LINGO optimization programme replicates the proposed approach OE2R.

Khan *et al.* [11] The researcher proposed and evaluates different IDS strategies for IoT Networks that are suitable for modest procedures. They developed a trust management system that certifies the technologies used to get status information about their neighbours. This technology makes it possible to identify maliciously performing components in a cost-effective and energy-efficient manner.

Zhang *et al.*[12] a cloud-based trust evaluation technique for clustered wireless sensor networks is proposed and described, which gears the transformation between quantitative and qualitative trust metrics of sensor nodes with the goal of achieving a healthier trust assessment. To begin, the strategy considers multi-factors such as message factor, communication factor, and energy factor, then builds a computational method for each trust factor to produce a component trust cloud. Second, the immediate trust cloud is created by combining the updated weights for each factor trust cloud. Finally, in order to obtain a final trust cloud, commendation trust cloud and instantaneous trust cloud are joined with a time sensitive component. Furthermore, trust cloud decision-making transforms the final trust cloud of the sensor node into a trust grade [19].

Li *et al.* [13] suggested a clustering-based routing method that allows for intervention and load matching routing metrics with the goal of decreasing network disputes. In this study, we propose a structure that reduces end-to-end delays while simultaneously paying close attention to both the quality of the entire route to the destination and the likely longevity of nodes with bottlenecks due to increased traffic in IoT.

Gupta *et al.* [14] proposed an improved cuckoo search-based energy balanced node clustering methodology that incorporates a new unbiased job for cluster head uniform distribution. In addition, an augmented harmony search-based routing protocol is planned for data packet direction-finding between cluster heads and sinks.

Meng, Weizhi, *et al* [15] Devices may generate disproportionate facts and data in the age of big data, which could reduce the efficiency of trust calculation. The authors focus on this challenge in this study, proposing a method for combining Bayesian-based trust management with traffic sampling for wireless intrusion detection in a categorised structure.

Gaber, Tarek, *et al* [16] Clustering would aid in clarifying the high communication overhead disputes in Intelligent Transportation Systems (ITS). We provide a bio-inspired and trust-based cluster head selection strategy for WSN that has been approved for use in ITS

applications in this study. The Bat Optimization Algorithm (BOA) is used to select cluster heads based on three metrics: trust value, residual energy, and number of neighbours.

3. PROBLEM IDENTIFICATION

Due to attacks and incursion, MANET suffers the following problems:

- Every system in the network has a one-to-one mapping between a node's logical and physical identity, but malicious nodes break this rule and inject fake logical identities into the network.
- Large network nodes appear to be incapable of validating and verifying the uniqueness of another network node.
- Network routing is affected by malicious nodes, which generate a false path and interrupt network functionality.
- Using several identities, Sybil node acquires a disproportionate amount of network resource.
- Sybil node conceals the attacker's identity, making it difficult to track them down.

4. FEATURE SELECTION TECHNIQUES

4.1 Chi-Square Analysis

Another widely used method is feature selection using the chi-square[2] test [1]. The importance of the chi-squared statistic to the class is used to measure the goodness of a feature in CS attribute evaluation. The first hypothesis, H (0), is that the two traits are unrelated, and it is tested using the chi-squared formula:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left(\frac{O_{ij}-E_{ij}}{E_{ij}}\right)^2 \quad (1.1)$$

The null hypothesis H_0 [2] confirms that O_{ij} is the observed frequency and E_{ij} expected (theoretical) frequency. The higher the value of 2, the more likely the hypothesis H0 is to be true [20].

4.2 Symmetrical Uncertainty

To find the optimum characteristics for classification, the symmetrical uncertainty (SU) [3] between the target idea and features is used. The elements with the highest SU values carry more weight. Based on the information theory, SU calculates the relationship between A and B variables. This is how it was calculated:

$$SU\ (A,)\ B = 2\ \frac{I\ (A,B)}{H(B)A + H\ (B)}$$

Computing I (A, B) as the MI between A and B features, and H(..) as an entropy function for A and B features. As the correction factor value is 2, the SU displays the normalised range value [0,1]. If the SU value is 1, one feature's information is predictable. If the SU value is 0, A and B are not linked.

4.3 Information Gain

The information theory metric, which defines the purity of an absolute collection of examples, commonly uses entropy. Gain Ratio, Information Gain, and Similarity Uncertainty (SU) [3] are all built on this basis. The entropy measure is used to determine the unpredictability of a system. Y has an entropy of

$$H(Y) = \sum_{y\in Y} p(y)\log_2(p(y)) \quad (1.2)$$

The marginal probability density function for the random variable Y is denoted by p(y). There is a link between features Y and X if the observed values of Y in the training data set S have partitioned according to the values of a second feature X, and the entropy of Y for the partitions produced by X is less than the entropy of Y before partitioning. After seeing X, the entropy of Y is:

$$H(Y|X) = \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2\big(p(y|x)\big) \quad (1.3)$$

The conditional probability of y given x is p(y |x). We can develop a measure representing extra information about Y provided by X that indicates the amount by which the entropy of Y reduces, given that entropy is an impurity criterion in a training set S. IG is the abbreviation for this measurement. It is provided by

$$IG = H(Y) - H(Y|X) = H(X) - H(X|Y) \quad (1.4)$$

IG [4] is a symmetrical measure that may be calculated using equation (1.4). After viewing X, the knowledge gained about Y is identical to the information gained about X after observing Y. The IG criterion has a flaw in that it favours features with higher values even when they aren't more informative.

5. PROPOSED ARTIFICIAL NEURAL NETWORK ARCHITECTURE

ANN is a powerful calculating system whose main theme is based on the biological NN. "Parallel Distributed Processing Systems" is another name for ANNs. ANN obtains a large number of units that are interconnected in some way, allowing the units to communicate with one another [21]. Those units, also known as neurons or nodes, are nothing more than parallel-processing CPUs. The ANN has been used in this study to classify Intrusion Detection in the system. The steps of the MLP-NN training algorithm are depicted in the following approach [22].

Step 1: *Initialize Bias, Learning rate* α, *weights, to begin the training of Multi-Layered Perceptron Neural Network. For simplicity and calculation, need to set weight =0 and bias* $\alpha = 1$.

Step 2: *Proceed step 3-8 at the terminating condition is true.*

Step 3: *Proceed step 4-6 for all training vector a.*

Step 4: *Initiate each input as follows:*

$$r_j = s_j \ (j = 1 \text{ to } m)$$

Step 5: *Get the net input with the next relations*

$s_{jn} = b + \sum_j^m r_j w_{jk}$ *Here bias is given as b, and the whole amount of input neuron is given by 'n'.*

Step 6: *Apply the activation function to obtain the final output for each input unit k=1 to n*

$$f(s_{jm}) = \begin{cases} 1 & if\ s_{jmk} > \theta \\ 0\ if - \theta \le s_{jmk} \le \theta \\ -1\ if\ s_{jmk} < -\theta \end{cases}$$

Step 7: *Adjust the weight and bias for r=1 to m and k=1 to n as follows:*

Step 7.1: *Case 1: if* $s_k \neq t_k$ *then*

$$w_{jk}(new) = w_{jk}(old) + \propto t_k r_j$$
$$s_k(new) = s_k(old) + \propto t_k$$

Step 7.2: *Case 2: if* $s_k = t_k$ *then*

$$w_{jk}(new) = w_{jk}(old)$$
$$s_k(new) = s_k(old)$$

Here 's' is the exact output, and 't' is the desired/target output.

Step 8: *Testing for terminating condition, which will occur while there is no variation in weight.*

6. RESULT AND DISCUSSION

6.1 Performance Metrics

The following table 1 depicts the performance metrics used to evaluate the performance of the proposed Sybil node detection framework with feature selection and classification methods [23].

Table 1 Performance Metrics

Metrics	Equation
Accuracy	$\frac{TP+TN}{TP+FN+TN+FP}$
True Positive Rate (TPR)	$\frac{TP}{TP+FN}$
False Positive Rate (FPR)	$\frac{FP}{FP+TN}$
Precision	$\frac{TP}{TP+FP}$

6.2 Dataset Description

KDD cup dataset is used to find the malicious node in the network. Since malicious node itself acts as a Sybil node in the MANET. KDD training dataset consists of relatively 4,900,000 single connection vectors where each single connection vectors consists of 41 features and is marked as either normal or an attack, with exactly one particular attack type.

6.3 Number of Features Obtained

The following table 2 gives the number of features obtained by the feature selection techniques and the combination of the considered three feature selection techniques. From the table 2, it is clear that the combined feature selection gives reduced number of features than the existing feature selection techniques.

Table 2 Number of Features Obtained

Feature Selection Techniques	Number of Features obtained
Original Dataset	41
Chi-Square (CS)	34
Symmetrical Uncertainty (SU)	32
Information Gain (IG)	31
Combined Feature Selection Method	25

6.4 Performance Analysis

Table 3 depict the classification accuracy (in %) obtained by the existing feature selection technique and combined feature selection method. From the table 3, it is clear that the proposed Feature Selection method with ANN more accuracy than other classifier.

Table 3 Classification Accuracy (in %) obtained by Feature Selection techniques

Original Dataset and Feature Selection processed datasets	Classification Accuracy (in %) obtained	
	ANN	SVM
Original Dataset	69.333	64.111
Chi-Square	85.58	66.51
Symmetrical Uncertainty	92	69.667
Information Gain	94.667	74.471
Combined Feature Selection Method	98	91.778

Table 4 depict the True Positive Rate (in %) obtained by the existing feature selection technique and combined feature selection method. From the table 4, it is clear that the proposed Feature Selection method with ANN more TPR than other classifier.

Table 4 True Positive Rate (in %) obtained by Feature Selection techniques

Original Dataset and Feature Selection processed datasets	True Positive Rate (in %) obtained	
	ANN	SVM
Original Dataset	68.24	63.224
Chi-Square	84.47	65.62
Symmetrical Uncertainty	91.78	68.556
Information Gain	93.776	73.362
Combined Feature Selection Method	97.75	90.667

Table 5 depict the False Positive Rate (in %) obtained by the existing feature selection technique and combined feature selection method. From the table 5, it is clear that the proposed Feature Selection method with ANN reduced FPR than another classifier.

Table 5 False Positive Rate (in %) obtained by Feature Selection techniques

Original Dataset and Feature Selection processed datasets	False Positive Rate (in %) obtained	
	ANN	SVM
Original Dataset	49.35	53.32
Chi-Square	35.56	38.51
Symmetrical Uncertainty	33.97	35.47
Information Gain	29.12	31.45
Combined Feature Selection Method	24.64	29.88

Table 6 depict the Precision (in %) obtained by the existing feature selection technique and combined feature selection method. From the table 6, it is clear that the proposed Feature Selection method with ANN more TPR than other classifier.

Table 6 Precision (in %) obtained by Feature Selection techniques

Original Dataset and Feature Selection processed datasets	Precision (in %) obtained	
	ANN	SVM
Original Dataset	55.3	49.94
Chi-Square	85.33	78.87
Symmetrical Uncertainty	93.89	70.69
Information Gain	95.57	80.71
Combined Feature Selection Method	98.84	92.77

7. CONCLUSION

In the second work, Malicious Node Identification Scheme for the detection of Identity-based Sybil node has proposed to detect the malicious node among the trusted node from the above method. Again, it improves the checking of Sybil node and malicious node in the network [24]. Like the above method, the Sybil nodes have discarded, and malicious nodes were considered for further identification. The false positive rate has decreased by using the method. The packet delivery ratio, detection rates are increased when the number of nodes as well as some Sybil nodes and malicious node presented in the network [25].

REFERENCES

[1] Li, Wenchao, et al. "A new intrusion detection system based on KNN classification algorithm in wireless sensor network." Journal of Electrical and Computer Engineering (2014).

[2] Creech, Gideon, and Jiankun Hu. "A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns." IEEE Transactions on Computers 63.4 (2014): 807-819.

[3] Nadiammai, G. V., and M. Hemalatha. "Effective approach toward Intrusion Detection System using data mining techniques." Egyptian Informatics Journal 15.1 (2014): 37-50.

[4] Feng, Wenying, et al. "Mining network data for intrusion detection through combining SVMs with ant colony networks." Future Generation Computer Systems 37 (2014): 127-140.

[5] Luo, Bin, and Jingbo Xia. "A novel intrusion detection system based on feature generation with visualization strategy." Expert Systems with Applications 41.9 (2014): 4139-4147.

[6] Umar, Muhammad Muneer, Amjad Mehmood, and Houbing Song. "SeCRoP: secure cluster head centered multi-hop routing protocol for mobile ad hoc networks." Security and Communication Networks 9.16 (2016): 3378-3387.

[7] Borkar, Gautam M., and A. R. Mahajan. "A secure and trust based on-demand multipath routing scheme for self-organized mobile ad-hoc networks." Wireless Networks 23.8 (2017): 2455-2472.

[8] Ramakrishnan, B., et al. "Cluster based emergency message broadcasting technique for vehicular ad hoc network." Wireless Networks 23.1 (2017): 233-248.

[9] Arain, Qasim Ali, et al. "Clustering based energy efficient and communication protocol for multiple mix-zones over road networks." Wireless Personal Communications 95.2 (2017): 411-428.

[10] Das, Santosh Kumar, and Sachin Tripathi. "Energy efficient routing formation technique for hybrid ad hoc network using fusion of artificial intelligence techniques." International Journal of Communication Systems 30.16 (2017): e3340.

[11] Khan, Zeeshan Ali, and Peter Herrmann. "A trust based distributed intrusion detection mechanism for internet of things." Advanced Information Networking and Applications (AINA), 2017 IEEE 31st International Conference on. IEEE, 2017.

[12] Zhang, Tong, Lisha Yan, and Yuan Yang. "Trust evaluation method for clustered wireless sensor networks based on cloud model." Wireless Networks 24.3 (2018): 777-797.

[13] Li, Jilong, et al. "A clustering-based routing algorithm in IoT aware Wireless Mesh Networks." Sustainable cities and society 40 (2018): 657-666.

[14] Gupta, Govind P., and Sonu Jha. "Integrated clustering and routing protocol for wireless sensor networks using Cuckoo and Harmony Search based meta heuristic techniques." Engineering Applications of Artificial Intelligence 68 (2018): 101-109.

[15] Meng, Weizhi, et al. "Enhancing Trust Management for Wireless Intrusion Detection via Traffic Sampling in the Era of Big Data." IEEE Access 6 (2018): 7234-7243.

[16] Gaber, Tarek, et al. "Trust-based secure clustering in WSN-based intelligent transportation systems." Computer Networks146 (2018): 151-158.

[17] Subhashini, M., & Gopinath, R. "Employee Attrition Prediction in Industry using Machine Learning Techniques", International Journal of Advanced Research in Engineering and Technology 11.12 (2020): 3329-3341.

[18] Subhashini, M., & Gopinath, R. "Mapreduce Methodology for Elliptical Curve Discrete Logarithmic Problems – Securing Telecom Networks", International Journal of Electrical Engineering and Technology 11.9 (2020): 261-273.

[19] Upendran, V., & Gopinath, R. "Feature Selection Based on Multi criteria Decision Making for Intrusion Detection System". International Journal of Electrical Engineering and Technology 11.5 (2020): 217-226.

[20] Upendran, V., & Gopinath, R. "Optimization Based Classification Technique for Intrusion Detection System". International Journal of Advanced Research in Engineering and Technology 11.9 (2020): 1255-1262.

[21] Kalaiarasi, K., & Gopinath, R. "Fuzzy Inventory EOQ Optimization Mathematical Model", International Journal of Electrical Engineering and Technology 11.8 (2020): 169-174.

[22] Shanmugavadivu, S. A., & Gopinath, R. "On the Homogeneous Five Degree Equation with five unknowns 〖2(x〗^5-y^5)+2xy(x^3-y^3)=〖37(x+y)(z〗^2-w^2)P^2," International Journal of Advanced Research in Engineering and Technology 11.11 (2020): 2399-2404.

[23] Shanmugavadivu, S. A., & Gopinath, R. "On the Non-homogeneous Ternary Five Degrees Equation with three unknowns x^2-xy+y^2=52z^5," International Journal of Advanced Research in Engineering and Technology 11.10 (2020): 1992-1996.

[24] Kalaiarasi, K., & Gopinath, R. "Stochastic Lead Time Reduction for Replenishment Python-Based Fuzzy Inventory Order EOQ Model with Machine Learning Support", International Journal of Advanced Research in Engineering and Technology 11.10 (2020): 1982-1991.

[25] Priyadharshini, D., Gopinath, R., & Poornapriya, T.S. "A fuzzy MCDM approach for measuring the business impact of employee selection", International Journal of Management 11.7 (2020):1769-1775.

[26] Poornapriya, T.S. & Gopinath, R. "Application of Machine Learning Techniques for Improving Learning Disabilities", International Journal of Electrical Engineering and Technology 11.10 (2020): 403-411.