International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

Sybil Attack Detection in Iot Environmental Sensors Using Optimized Fuzzy C-Means Clustering Algorithm

R. Premkumar[1], Dr. R. Manikandan[2], Dr.N. Palanivel[3]
[1]Research Scholar, Dept of Computer science and Engineering, Annamalai University, Chidambaram, Tamilnadu, premkumar.phd.16@gmail.com.
[2]Associate Professor, Dept of Computer science and Engineering, Annamalai University. Chidambaram, Tamilnadu, rmkmanikandan1111@gmail.com.
[3]Professor and Head, Dept. of Computer science and Engineering, Manakula Vinayagar Institute of Technology, Puducherry, npalani76@gmail.com.

Abstract
IoT environmental sensors are devices that use the Internet of Things (IoT) to monitor and measure various environmental conditions like temperature, humidity, air quality, and light levels. A Sybil attack happens when a malicious node in a network falsifies multiple fake identities to look like many distinct nodes. It affects an IoT sensor network where a malicious node generates multiple fake identities to betray and interrupt network operations. The focus of this work is to use of Machine Learning (ML) algorithm to detect Sybil attack instances in IoT wireless networks. Hence an optimized Fuzzy C-Means (FCM) clustering algorithm is proposed for Sybil attack detection in IoT networks. During the training phase, both the signal feature (SF) and frequency offset feature (FOF) from each IoT device are trained to classify the Sybil attack nodes, by means of FCM clustering technique. For optimizing the clustering performance, a metaheuristic Improved Gradient-based Optimizer (IGBO) algorithm is employed. Experimental results show that the optimized FCM clustering algorithm provides higher detection accuracy with reduced affected packets and computational overhead, when compared to the existing techniques.
Keywords: *Internet of Things (IoT), Sybil attack, Fuzzy C-Means clustering, Gradient-based Optimizer (IGBO)*

1. INTRODUCTION

Because of the Internet's quick development, smooth communication has become essential in today's digital environment. Improvements in Internet architecture led directly to the emergence of the Internet of Things (IoT), which allows a vast number of objects to communicate, interact, and exchange data on their own. IoT environmental sensors are used to monitor various parameters like air and water quality, temperature, humidity, and soil conditions, providing real-time data for environmental management. These sensors, connected to the internet, transmit data to a central platform for analysis and informed decision-making [1]. Since IoT networks gain popularity, they are becoming progressively susceptible to security threats. Consequently, safeguarding IoT devices and designing intrusion-resistant networks is vital for protecting sensitive data [2].

A Sybil attack happens when a malicious node in a network falsifies multiple fake identities to look like many distinct nodes. This weakens trust and interrupts network functions like voting, routing, or resource allocation. It can result in data manipulation, routing misdirection, and degraded network performance. Sybil attack is extensively recognized as one of the most critical threats contributing to network failure [3]. This attack affecting an IoT network includes a malicious node generating multiple fake identities to betray and interrupt network operations [4]. Consequently, Sybil attacks can strictly compromise network services like resource allocation, routing, data aggregation, and voting systems. Hence, recognizing and mitigating Sybil attacks is vital for maintaining the integrity and security of a network [5].

The use of localization-based detection methods, especially those based on Received Signal Strength Indicator (RSSI), is one of the most promising ways to combat Sybil assaults. Comparing the RSSI values of many identities can assist in identifying Sybil nodes because a single device claiming multiple identities will emit signals from the same physical location [6]. Time Difference of Arrival and Angle of Arrival are two more localization-based techniques that use the direction or duration of signal transit to determine the position of nodes [7]. The channel feature-based detection methods mainly utilize the variation in the channel environment between illegal and legitimate devices due to their different locations for identification. Swarm intelligence, artificial immune systems, and neural networks are examples of bio-inspired algorithms that can learn from novel situations, adapt to shifting network settings, and withstand errors. These cooperative, decentralized systems mimic natural systems to identify anomalous patterns of activity in the network and make wise decisions [8].

2. Related Works

An unsupervised Sybil attack detection method using signal frequency bias distribution features is proposed [8]. It estimates the signal frequency bias of the emitted signal of each wireless device and then the signal frequency offset distribution characteristics. It uses the DBSCAN clustering method to perform cluster analysis based on the distribution features.

An intelligent Sybil attack detection approach for FANETs-based Internet of Flying things (IoFT) has been proposed [9] using physical layer characteristics of the radio signals emitted from the UAVs. A supervised machine learning approach is employed based on received signal strength difference (RSSD) and the time difference of arrival (TDoA).

A distributive and lightweight Sybil attack detection scheme for the mobile IoT is proposed [10]. The scheme consists of two rounds in which the Identity information is sent from member nodes to edge nodes. In the first round, edge nodes calculate the possible RSSI interval for each member node and in the second round, they check the RSSI value of member nodes to detect Sybil attacks. A position prediction algorithm is also proposed using LSTM networks.

The authors in [11] have proposed a multi-factor authentication model for detecting malicious activities in UWSN. The authors proposed algorithms for both the monitoring, detection, and mitigation of Sybil attacks. The multi-factor authentication model involves updating the packets' header information by including an identifier based on MAC address (IMAC), direction of arrival, and hop count (HC) for validating incoming packets.

3. PROPOSED METHODOLOGY

3.1Overview

The focus of the architecture is on the use of Machine Learning (ML) algorithm to detect Sybil attack instances in IoT wireless networks. In the Sybil attack scenario, since signal frames of Sybil devices are created by the same device of the attacker, they will share similar signal features which provide the possibility to detect the Sybil attack. Using this principle, the proposed method detects the Sybil attack in the surrounding area and distinguishes the Sybil devices from other legitimate devices. The TDoA_ratio which captures the relationship between the two signals, was selected as one of the attributes. Similarly, the RSSI_diff is another attribute which represents the difference between the RSSI values of the signal measured at the two different monitoring nodes.

3.2 Training Phase

Signals from each IoT sensor would be sampled by the monitoring nodes at certain intervals until the end of the training period. The two monitoring nodes would detect and collect the RSSI and ToA of each signal being sampled. Subsequently the corresponding RSSI_diff and TDoA_ratio of each signal would be calculated as

$$TDoA\_ratio_j = ToA_i(n_1) - ToA_i(n_2) \quad (1)$$

$$RSSI\_diff_i = RSSI_i(n_1) - RSSI_i(n_2) \quad (2)$$

Where $ToA_i(n1)$ and $ToA_i(n2)$ represent the ToA of radio signal i obtained by monitoring stations n_1 and n_2, respectively .$RSSI_i(n1)$ and $RSSI_i(n2)$ represent the RSSI of radio signal i obtained by monitoring stations n_1 and n_2, respectively.

Then a signal feature (SF) for device j is derived from RSSI_diff and TDoA_ratio as

$$SF_j = \{ TDoA\_ratio_j, RSSI\_diff_i \} \quad (3)$$

The frequency offset (FO) of i^{th} frames emitted by the k^{th} IoT device in the acquisition environment is denoted as $f_{k,j}$, where j=1,2....N . Then, the mean and variance for k^{th} device frequency offset are given by

$$MeanFO_k = \frac{1}{N}\sum_{j=1}^{N} f_{k,j} \quad (4)$$

$$VarFO_k = \frac{1}{N}\sum_{j=1}^{N}(f_{k,j} - Mean(f_k))^2 \quad (5)$$

Then a frequency offset feature (FOF) for device k is derived from MeanFO and VarFO as

$$FOF_k = \{MeanFO_k, VarFO_k\} \quad (6)$$

During the training phase, both the signal feature (SF) and frequency offset feature (FOF) from each device are trained to classify the Sybil attack nodes.

3.3 FCM clustering

FCM is the most popular fuzzy-based clustering algorithm. In FCM, a dataset is clustered into k groups in which each data component may be mapped to each cluster with some degree of membership, which ranges between 0 and 1. FCM repeatedly executes two phases to attain the best solution. In first phase, each data component will be mapped with a membership value for each cluster. In phase two, the data component is assigned to a cluster having highest membership value. The basic FCM clustering algorithm [12] is presented below:

FCM Clustering Algorithm

Step 1: The process to arbitrarily choose 'c' centres off cluster.

Step 2: The procedure of estimation in fuzzy membership φ^{ij} using

$$\varphi^{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{D^{ij}}{D^{ik}}\right)^{\left(\frac{2}{m}-1\right)}} \quad (7)$$

Step 3: The process of fuzzy centers computation 'v_j' using:

$$V^{j} = \frac{\left(\sum_{j=1}^{N}\left(\varphi^{ij}\right)^{m} x^{i}\right)}{\left(\sum_{j=1}^{N}(\varphi^{ij})^{m}\right)}, \forall^{j=1,2,\ldots c} \quad (8)$$

Step 4: Duplicate walk 2) in addition to 3) minimum value till 'J' is attained otherwise$||U(k+1) - U(k)|| < \beta$.

Where, 'k' is iteration stages,'β' is the termination criterion linking [0, 1],'$U = (\varphi^{ij})n * c$' is the matrix membership fuzzy and'J' is the function objective.

3.4 Improved Gradient-Based Optimizer

For optimizing the clustering performance, a metaheuristic Improved Gradient-based Optimizer (IGBO) [13] algorithm was employed.

Algorithm 2:Pseudo code of IGBO

---

1. Initialization

Assign values to g_1, g_2, M, ω_{min}, ω_{max}, pr, and ε.

An initial population X0 = [$x_{0,1}$, $x_{0,2}$,..., $x_{0,D}$] should be created.

Determine the value of the objective function f(X_0), where n = 1,..., N

List the worst and greatest options. X^{m}_{worst} and X^{m}_{best}

2. Main Loop

While m<M (iterations) **Do**

for n = 1: N (particles) do

for i = 1: D (dimensions) do

Select randomly r1 ≠ r2 ≠ r3 ≠r4 ≠ n in the range of [1, N]

Calculate the operator G using the following equation:

$G = g_1 + ((g_2 - g_1) * (rand))$ (9)

where g_1 and g_2 are real numbers

Compute the direction of movement (DM) as

$DM = G \times \rho_2 \times (x_{best} - x_n)$ (10)

3. Gradient search Rule

Calculate the parameters r_a and r_b using the following eqns.

$r_a = \frac{Itr}{MaxItr}$

$r_b = 1 - \left(\frac{Itr}{MaxItr}\right)$ (11)

Calculate the position $x^{m+1}_{n,i}$ as

$$x_n^{m+1} = w(itr) * r_a * (r_b * x1_n^m + (1-r_b) * x2_n^m) + (1-r_a) * x3_n^m \quad (12)$$

end for

4. Local escaping Rule

Calculate the parameters γ_1, γ_2 and γ_3 using the equations (13)-(15)

$$\gamma_1 = \frac{\text{Itr}}{\text{MaxItr}} \quad (13)$$

$$\gamma_2 = 1 - \left(\frac{\text{Itr}}{\text{MaxItr}}\right) \quad (14)$$

$$\gamma_3 = 1 - \left(\frac{\text{Itr}}{\text{MaxItr}}\right) \quad (15)$$

Compute random numbers u_1, u_2, and u_3 using Equations (16)–(18)

$$u1 = L_1 * 2 * \gamma_1 + (1-L_1) \quad (16)$$

$$u2 = L_1 * \gamma_2 + (1-L_1) \quad (17)$$

$$u3 = L_1 * \gamma_3 + (1-L_1) \quad (18)$$

if rand < pr, then

if rand < 0.5 , then

Calculate the position x_m^{LEO} as

$$x_m^{LEO} = x_n^{m+1} + f_1 * (u_1 * x_{best} - u_2 * x_k^m) + f_2 * \rho_1 * (u_3 * (x2_n^m - x1_n^m) + u_2 * (x_{r1}^m - x_{r2}^m))/2 \quad (19)$$

At the next iteration calculate the new coming position X^{m+1}_n

$$x_n^{m+1} = w(itr) * r_a * (r_b * x1_n^m + (1-r_b) * x2_n^m) + (1-r_a) * x3_n^m \quad (20)$$

Where the inertia weight w(itr) is given by

$$w(itr) = \left(\frac{\max itr - itr}{\max itr}\right)^2 * (w_{\max} - w_{\min}) + w_{\min} \quad (21)$$

where w_{max} and w_{min} are the maximum and minimum inertia weights.

end if

end if

5. Update the positions x^m_{best} and x^m_{worst}

end for

m=m+1

end

return x^m_{best}

---

4. Experimental Results

The proposed Optimized FCM clustering technique for Sybil attack detection is implemented in Python 3.6.

4.1 Classification Results

The accuracy and F1-score metrics are used to measure the performance of optimized-FCM clustering technique with existing FCM, K-Means and DBSCAN clustering techniques.

Table 1 and Figure 1 show the comparison results of all techniques for the metrics accuracy, recall and F1-score.

Metrics	**Optimized-FCM** (%)	**FCM** (%)	**K-Means** (%)	**DBSCAN** (%)
Accuracy	0.96	0.95	0.93	0.94
F1-score	0.90	0.85	0.84	0.87

Table 1 Classification results

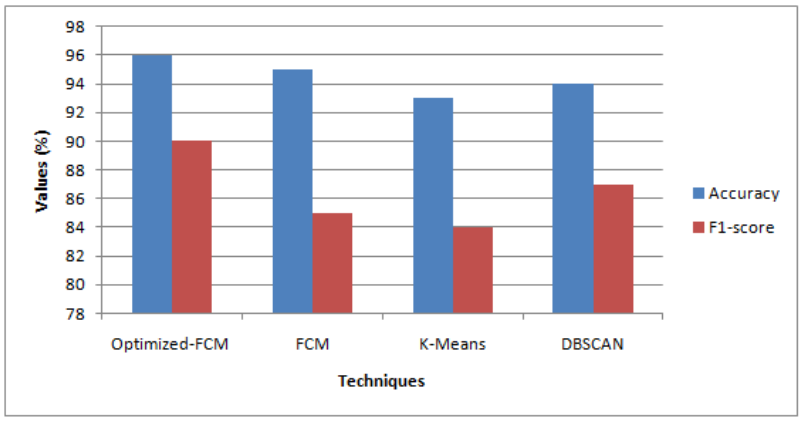

Figure 1 Detection results of all techniques

As seen from Figure 1, the proposed optimized-FCM attains highest accuracy around 96%, which is 1% greater than that of FCM, 3% greater than that of K-Means and 2% greater than that of DBSCAN. The optimized-FCM attains highest F-score around 90%, which is 5% greater than that of FCM, 6% greater than that of K-Means and 3% greater than that of DBSCAN.

4.2 Results of Varying the Attackers

To analyze the impact of attackers over the network size, the number of attackers is varied from 2 to 10 out of 100 nodes. The performance metrics fraction of affected packets and computational overhead are measured for the optimized-FCM and unsupervised Sybil attack detection [8] techniques.

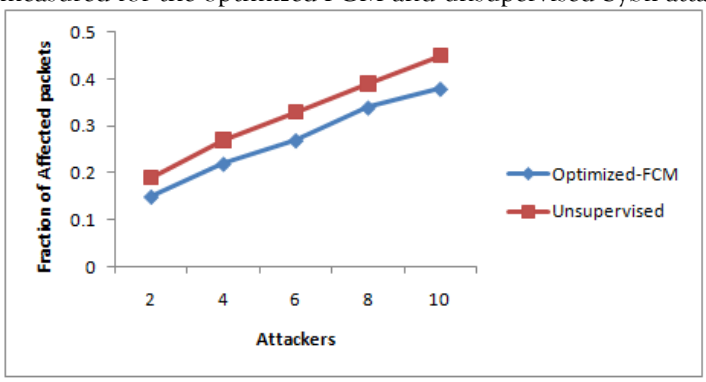

Figure 2 Fraction of affected packets Vs attackers

The results of fraction of affected packets for both the techniques are presented in Figure 2. As optimized-FCM detects unauthorized attacks and Sybil attacks, the fraction of affected packets is 17% lesser than that of unsupervised detection technique.

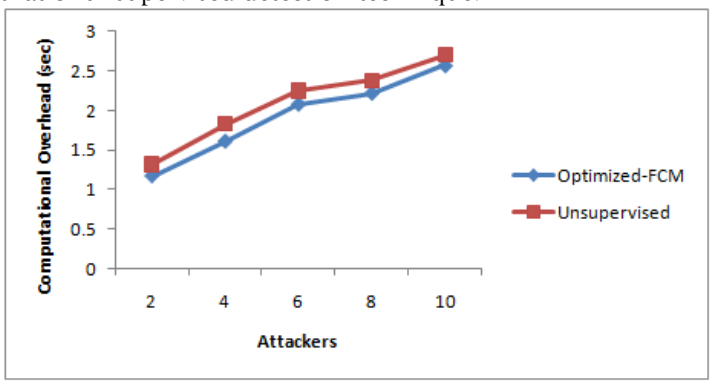

Figure 3 Computational overhead Vs attackers

The computational overhead in terms of time for both the techniques is shown in Figure.3. Since optimized-FCM does not use any trace data, it has 8% lesser overhead than unsupervised detection technique.

5. CONCLUSION

In this paper, an optimized FCM clustering algorithm is proposed for Sybil attack detection in IoT environmental sensors. During the training phase, both the signal feature (SF) and frequency offset feature (FOF) from each IoT device are trained to classify the Sybil attack nodes, by means of FCM

clustering technique. For optimizing the clustering performance, IGBO algorithm is employed. The detection accuracy, F1-score, fraction of affected packets and computational overhead metrics are used to measure the performance the optimized-FCM clustering technique with existing FCM, K-Means and DBSCAN clustering techniques. Experimental results show that the optimized FCM clustering algorithm provides higher detection accuracy with reduced affected packets and computational overhead, when compared to the existing techniques.

REFERENCES

1. Madhu, B., Chari, M. V. G., Vankdothu, R., Silivery, A. K., & Aerranagula, V. (2023b). Intrusion detection models for IOT networks via deep learning approaches. Measurement Sensors, 25, 100641. https://doi.org/10.1016/j.measen.2022.100641
2. Sohail, S., Fan, Z., Gu, X., & Sabrina, F. (2022b). Multi-tiered Artificial Neural Networks model for intrusion detection in smart homes. Intelligent Systems With Applications, 16, 200152. https://doi.org/10.1016/j.iswa.2022.200152
3 A. Vasudeva and M. Sood, "On the Vulnerability of the Mobile Ad Hoc Network to Transmission Power Controlled Sybil Attack: Adopting the Mobility-Based Clustering," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 9, 2022.
4. A. Paul and S. Sinha, "Modeling Different Forms of Sybil Attack in MANET," Grenze International Journal of Engineering and Technology, 2016.
5 P. V. Rao, S. Murthy, V. G. Krishnan and D. V., "Detection of Sybil Attack in MANET Environment Using ANFIS with Bloom Filter Algorithm," Indian Journal of Computer Science and Engineering, vol. 13, no. 1, pp. 82–92, Feb. 2022.
6 P. Muthusamy and Sheela, "Sybil Attack Detection Based on Authentication Process Using Digital Security Certificate Procedure for Data Transmission in MANET," International Journal of Engineering & Technology, vol. 7, no. 3.27, pp. 270–276, 2018.
7.A. Angappan, T. P. Saravanabava, P. Sakthivel and K. S. Vishvaksenan, "Novel Sybil attack detection using RSSI and neighbour information to ensure secure communication in WSN," Journal of Ambient Intelligence and Humanized Computing, vol. 12, p. 6567–6578, 2021.
8. Yinghua Tian et al ,"Unsupervised Detection of Sybil Attack in Wireless Networks", 8th Annual International Conference on Geo-Spatial Knowledge and Intelligence, IOP Conference Series: Earth Environmental Science,2021, 693 012114
9. Donpiti (Mick) Chulerttiyawong and Abbas Jamalipour," Sybil Attack Detection in Internet of Flying Things-IoFT: A Machine Learning Approach",IEEE Internet of Things Journal,25364-2022.
10. Junwei Yan, Tao Jiang, Liwei Lin, Zhengyu Wu, Xiucai Ye, Mengke Tian and Yong Wang,"A novel Sybil attack detection scheme in mobile IoT based on collaborate edge computing",EURASIP Journal on Wireless Communications and Networking,(2023) 2023:25,https://doi.org/10.1186/s13638-023-02233-8
11. A. Al Guqhaiman, O. Akanbi, A. Aljaedi, and C. E. Chow, ''Lightweight multi-factor authentication for underwater wireless sensor networks,'' in Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI), Dec. 2020, pp. 188–194.
12. J. Miao, X. Zhou and T.-Z. Huang, "Local segmentation of images using an improved fuzzy C-means clustering algorithm based on self-adaptive dictionary learning" Applied Soft Computing Journal (2020), doi: https://doi.org/10.1016/j.asoc.2020.10620
13. Altbawi, S.M.A., Khalid,S.B.A., Mokhtar, A.S.B., Shareef, H., Husain, N., Yahya, A., Haider, S.A., Moin, L., Alsisi, R.H, "An Improved Gradient-Based Optimization Algorithm for Solving Complex Optimization Problems", Processes 2023, 11, 498. https://doi.org/ 10.3390/pr11020498