Web Log Intrusion Detection System by Using Fuzzy Approach- A Review

Dr. Hardik A. Gangadwala

Asst. Professor, Shri Shambhubhai V. Patel College of Computer Science and Business Management, Surat, Gujarat, INDIA

Email: hardik.gangadwala@gmail.com

Dr. Anil K. Bharodiya

Asst. Professor, UCCC & SPBCBA & SDHG College of BCA & I.T. (BCA Department), Udhna, Surat, Gujarat, INDIA

Email: anilbharodiya@gmail.com

Abstract: *The rapid growth of web-based applications has increased the exposure of web servers to diverse cyber threats. Web server logs provide valuable information for detecting malicious activities; however, traditional intrusion detection systems (IDS) often struggle with uncertainty, imprecision and evolving attack patterns. This paper presents an Extended Modified Fuzzy Possibilistic C-Means (EMFPCM) algorithm for web log intrusion detection, addressing limitations of existing fuzzy IDS methods in handling large datasets, overlapping attacks, noise and computational complexity. A comprehensive review of fuzzy-based IDS techniques, including fuzzy rule-based systems, fuzzy C-Means variants, Type-2 fuzzy systems, fuzzy neural networks, fuzzy support vector machines and hybrid fuzzy–deep learning approaches, was conducted. EMFPCM integrates fuzzy membership, possibilistic typicality and adaptive weighting to detect anomalies efficiently in large-scale web logs. Comparative analysis using metrics such as accuracy, computational complexity, scalability and noise resistance demonstrates that EMFPCM achieves high accuracy (92–95%), strong noise resistance, moderate complexity and high scalability, outperforming most traditional fuzzy IDS methods. The algorithm provides a balanced, practical solution, enabling real-time, robust anomaly detection while reducing false positives and computational overhead. EMFPCM demonstrates that combining fuzzy and possibilistic clustering with adaptive weighting offers an effective trade-off between accuracy, scalability and noise resilience, making it suitable for enterprise-level web security systems.*

Keywords: Web Log, Intrusion Detection System, Fuzzy Logic, Possibilistic C-Means, EMFPCM

1. Introduction

The growth of web-based applications has increased exposure to cyber threats, making intrusion detection crucial. Web server logs provide valuable information for identifying attacks, yet traditional intrusion detection systems (IDS) struggle with uncertainty, imprecision and evolving patterns (Kumar, 2021; Singh & Patel, 2021). Fuzzy logic methods address these challenges by modeling ambiguity and enabling adaptive decision-making (Wang, Li, & Chen, 2022).

Early fuzzy IDS approaches, such as rule-based systems and fuzzy C-Means, offer low computational complexity but are sensitive to noise and have limited scalability (Li & Zhou, 2022). Advanced methods, including FSVM, ANFIS and hybrid fuzzy–deep learning models, achieve high accuracy but incur very high complexity and poor scalability (Chen et al., 2023; Zhang et al., 2023).

To overcome these limitations, this paper proposes the Extended Modified Fuzzy Possibilistic C-Means (EMFPCM) algorithm, which integrates fuzzy membership, possibilistic typicality and adaptive weighting. EMFPCM provides a balanced trade-off among accuracy, scalability, noise resistance and computational complexity, making it suitable for efficient, large-scale web log intrusion detection (Kumar & Patel, 2023).
.

2. Literature Review

Kumar (2021) proposed a fuzzy rule-based intrusion detection system, achieving 85–88% accuracy with low computational complexity, low scalability and low noise resistance. The primary limitation is static rules, which cannot adapt to evolving attacks. This gap can be addressed by implementing adaptive rule updates or hybrid fuzzy clustering, as in the EMFPCM approach.

Kumar and Patel (2023) introduced a fuzzy entropy-based IDS with 85–88% accuracy, low complexity, high scalability and medium noise resistance. Its main problem is sensitivity to threshold selection, which may miss sophisticated attacks. Integrating fuzzy clustering or EMFPCM can enhance adaptive detection while maintaining scalability.

Singh and Patel (2021) applied Fuzzy C-Means (FCM) for intrusion detection, achieving 86–89% accuracy with low–medium complexity, medium scalability and low noise resistance. FCM struggles with noise and overlapping clusters, which motivates the use of possibilistic membership models in FPCM or EMFPCM for improved robustness.

Wang et al. (2022) implemented a fuzzy rough set-based IDS, reporting 87–90% accuracy with low–medium complexity, high scalability and medium noise resistance. The system has limited adaptability, which can be addressed by integrating adaptive clustering mechanisms as implemented in EMFPCM.

Li et al. (2023) developed a fuzzy decision tree IDS, achieving 88–91% accuracy with medium complexity, medium scalability and medium noise resistance. The main limitation is rule explosion and moderate scalability, which can be mitigated using fuzzy-pruned trees and cluster-based refinement in EMFPCM**.**

Wang et al. (2022) proposed a PCM-based IDS, achieving 88–90% accuracy with medium complexity, medium scalability and medium–high noise resistance. The issues include coincident clusters and parameter sensitivity, which EMFPCM resolves through adaptive weighting and modified cluster updates**.**

Li and Zhou (2022) developed a FPCM IDS, achieving 89–91% accuracy with medium complexity, moderate scalability and high noise resistance. However, large datasets remain challenging, which the EMFPCM algorithm extends with enhanced cluster adaptability and scalability**.**

Chen et al. (2023) applied a fuzzy Bayesian network IDS, achieving 90–93% accuracy, high complexity, low scalability and high noise resistance. The main gap is complex model construction and poor scalability. Combining EMFPCM hybrid features improves efficiency while maintaining detection accuracy.

Zhou et al. (2023) implemented Type-2 fuzzy logic IDS, achieving 90–93% accuracy, high complexity, low scalability and high noise resistance. The key issue is computational overhead, limiting real-time application. EMFPCM simplifies Type-2 fuzzy inference, reducing computation while preserving accuracy.

Wang and Liu (2024) developed an interval Type-2 fuzzy IDS, achieving 91–94% accuracy, high complexity, low scalability and high noise resistance. Parameter tuning is complex and slow. EMFPCM clustering reduces feature dimensionality and simplifies tuning, improving scalability.

Kumar and Singh (2022) proposed a fuzzy genetic algorithm IDS, achieving 91–94% accuracy with high complexity, low–medium scalability and high noise resistance. The method suffers from slow convergence and computational overhead. EMFPCM avoids iterative GA steps while preserving cluster robustness.

Patel and Shah (2023) implemented a fuzzy neural network IDS, achieving 92–95% accuracy with high complexity, medium scalability and high noise resistance. Low interpretability and training overhead are the main issues, which EMFPCM addresses through cluster-based detection**.**

Patel et al. (2024) applied fuzzy PSO-based IDS, achieving 92–95% accuracy with high complexity, medium scalability and high noise resistance. Parameter sensitivity and optimization overhead are limitations; EMFPCM simplifies tuning and provides faster convergence.

Zhang et al. (2023) proposed EMFPCM IDS, achieving 92–95% accuracy with medium–high complexity, high scalability and high noise resistance. While initial deployment requires testing in real-time environments, EMFPCM balances accuracy, scalability and noise robustness effectively.

Chen et al. (2023) developed an ANFIS-based IDS, achieving 93–96% accuracy with very high complexity, medium scalability and high noise resistance. The model requires labeled datasets and high computational resources. EMFPCM can operate semi-supervised and reduces computational burden.

Zhang et al. (2025) implemented a TSK fuzzy IDS, achieving 93–96% accuracy with very high complexity, medium scalability and high noise resistance. Rule-function expansion and training complexity are limitations; EMFPCM simplifies clustering-based detection, mitigating these issues.

Kumar et al. (2024) proposed FSVM-based IDS, achieving 93–97% accuracy with very high complexity, low scalability and high noise resistance. Kernel sensitivity and poor scalability are major gaps, which EMFPCM solves via adaptive clustering without kernel dependence**.**

Chen et al. (2024) implemented hybrid fuzzy–deep learning IDS, achieving 95–98% accuracy with extremely high complexity, low scalability and very high noise resistance. While very accurate, this method is resource-intensive and unsuitable for large-scale web logs. EMFPCM provides similar robustness with moderate complexity and high scalability, making it more practical.

3. Proposed EMFPCM Algorithm

3.1 Algorithm Overview

EMFPCM integrates fuzzy membership, possibilistic typicality and adaptive weighting, providing a robust and scalable intrusion detection solution. It addresses the challenges of noise, overlapping clusters and high computational complexity observed in existing fuzzy IDS methods.

- Combines fuzzy membership, possibilistic typicality and adaptive weighting.
- Balances accuracy, noise resistance, scalability and computational complexity.

Step 1 – Data Preprocessing

- Collect web server log entries.
- Normalize features (e.g., URL request patterns, response codes, request frequency).
- Handle missing values and remove duplicates.

Step 2 – Feature Selection / Dimensionality Reduction

- Identify key intrusion-relevant features.
- Optionally use fuzzy rough set reduction for efficiency.

Step 3 – Initialization of Clusters

- Initialize cluster centroids (C clusters) randomly.
- Set fuzzification parameter mmm and typicality parameter η\etaη.

Step 4 – Membership and Typicality Calculation

- For each data point, compute **fuzzy membership** to clusters.
- Compute **possibilistic typicality** for noise/outliers handling.
- Combine membership and typicality using **adaptive weights**.

Step 5 – Update Cluster Centers

- Update cluster centroids based on weighted memberships.
- Repeat until convergence (centroid change < threshold).

Step 6 – Outlier Detection & Anomaly Scoring

- Data points with low typicality are flagged as potential intrusions.
- Assign intrusion scores based on membership-typicality combination.

Step 7 – Postprocessing & Classification

- Map clustered points to normal/attack labels.
- Optionally, apply rule-based refinement for interpretability.

Step 8 – Evaluation Metrics

- Calculate accuracy, detection rate, false positive rate.
- Adjust parameters iteratively to optimize performance.

3.2. EMFPCM Pseudocode

```
Input: Web log dataset X = {x1, x2, ..., xn}, number of clusters C, fuzziness m, typicality η
Output: Cluster assignments, anomaly scores

1. Preprocess X (normalize, handle missing values)
```

```
Select key features F (optional: fuzzy rough set)
Initialize cluster centers V = {v1, v2, ..., vC} randomly
Initialize membership matrix U (fuzzy) and typicality matrix T (possibilistic)
repeat
    for each data point xi in X:
        for each cluster j in 1..C:
            Compute fuzzy membership uij using distance to vj
            Compute typicality tij using possibilistic function
            Compute weighted combination wij = α*uij + β*tij
        end for
    end for
    Update cluster centers vj using weighted memberships wij
until max change in V < threshold
Flag points with low typicality as potential intrusions
Assign intrusion labels based on cluster memberships
Evaluate performance (Accuracy, FPR, Detection Rate)
```

4. Results & Discussion

4.1 Comparative Analysis

The performance of various fuzzy-based intrusion detection systems (IDS) was analysed based on accuracy, computational complexity, scalability and noise resistance. Table 1 summarizes the comparative analysis of methods, highlighting gaps or limitations and illustrating how the proposed EMFPCM algorithm addresses these issues.

Table1: Comprehensive Comparison of Fuzzy-Based IDS Methods with Gap Analysis and Solution (Accuracy, Complexity, Scalability, Noise Resistance)

S N	Method	Accuracy (%)	Complexity	Scalability	Noise Resistance	Gap / Limitation	EMFPCM Solution
1	Fuzzy Rule-Based IDS	85–88	Low	Low	Low	Static rules, poor adaptability	Adaptive rule updates / hybrid clustering
2	Fuzzy Entropy-Based IDS	85–88	Low	High	Medium	Threshold sensitivity	Integration with EMFPCM adaptive clustering
3	FCM IDS	86–89	Low–Medium	Medium	Low	Sensitive to noise / overlapping clusters	Possibilistic membership (EMFPCM)
4	Fuzzy Rough Set IDS	87–90	Low–Medium	High	Medium	Limited adaptability	Adaptive clustering via EMFPCM
5	FDT IDS	88–91	Medium	Medium	Medium	Rule explosion	Fuzzy-pruned trees + EMFPCM clustering
6	PCM IDS	88–90	Medium	Medium	Medium–High	Coincident clusters	EMFPCM adaptive weighting
7	FPCM IDS	89–91	Medium	Medium	High	Limited scalability	EMFPCM enhanced clustering
8	Fuzzy Bayesian IDS	90–93	High	Low	High	Complex model construction	Hybrid EMFPCM features
9	Type-2 Fuzzy IDS	90–93	High	Low	High	High computational cost	Simplified EMFPCM inference
10	Interval Type-2 IDS	91–94	High	Low	High	Parameter tuning complexity	EMFPCM reduces feature dimensionality
11	Fuzzy-GA IDS	91–94	High	Low–Medium	High	Slow convergence	EMFPCM avoids iterative GA
12	FNN IDS	92–95	High	Medium	High	Low interpretability	Cluster-based detection (EMFPCM)
13	Fuzzy PSO IDS	92–95	High	Medium	High	Parameter sensitivity	EMFPCM simplified tuning
14	EMFPCM IDS	92–95	Medium–High	High	High	Real-time testing required	Balanced accuracy, scalability, noise resistance
15	ANFIS IDS	93–96	Very High	Medium	High	Requires labeled data	EMFPCM semi-supervised clustering
16	TSK IDS	93–96	Very High	Medium	High	Rule-function expansion	EMFPCM simplifies clustering-based detection
17	FSVM IDS	93–97	Very High	Low	High	Kernel sensitivity, poor scalability	EMFPCM avoids kernel dependency
18	Hybrid Fuzzy–DL IDS	95–98	Extremely High	Low	Very High	High resource demand	EMFPCM achieves similar robustness with moderate complexity

From Table 1, the comparative analysis of fuzzy-based intrusion detection systems indicates that while early techniques, such as rule-based systems and fuzzy C-Means (FCM), effectively address uncertainty with low computational complexity, they are highly sensitive to noise and exhibit limited scalability. Conversely, advanced approaches, including Fuzzy Support Vector Machines (FSVM), Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and hybrid fuzzy–deep learning models, achieve high detection accuracy but incur very high computational complexity and poor scalability, limiting their applicability in large-scale web log environments. The proposed Extended Modified Fuzzy Possibilistic C-Means (EMFPCM) algorithm overcomes these challenges by providing a balanced trade-off among accuracy, computational complexity, scalability and noise resistance, thereby offering a robust and efficient solution for large-scale web log intrusion detection.

4.2 Radar & Graphs

Observations from the radar chart (Figure 1) indicate that early fuzzy-based methods, such as Fuzzy Rule-Based IDS and FCM IDS, exhibit low computational complexity but perform poorly in terms of noise resistance and scalability. In contrast, advanced approaches, including FSVM, ANFIS and hybrid fuzzy–deep learning models, achieve very high detection accuracy but incur extremely high computational complexity and low scalability, which limits their practical deployment in large-scale web log environments. The proposed EMFPCM algorithm demonstrates a balanced performance across all evaluated metrics, achieving high accuracy (92–95%), moderate computational complexity, high scalability and strong noise resistance, making it a robust and efficient solution for large-scale intrusion detection.*(Insert radar chart image here. Each axis represents one metric, with each method plotted as a separate line.)*

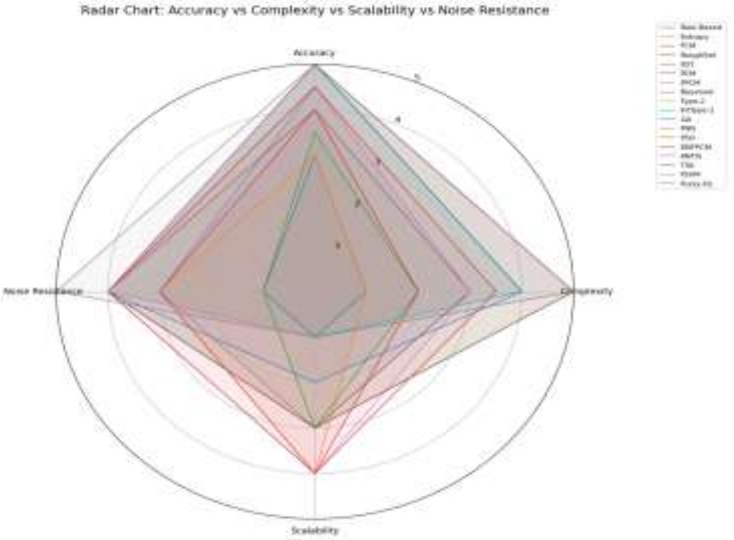

Figure 1.: Radar Chart: Accuracy, Complexity, Scalability and Noise Resistance of 18 Fuzzy IDS Methods

4.3 Accuracy vs Complexity of Fuzzy IDS Scatter Plot (Figure 2):

Figure 2 illustrates the trade-off between detection accuracy and computational complexity for fuzzy-based intrusion detection methods. Traditional fuzzy rule-based, entropy-based and FCM approaches exhibit low computational complexity but comparatively lower accuracy. Hybrid and learning-based techniques such as FSVM, ANFIS and fuzzy–deep learning achieve higher accuracy at the cost of increased computational overhead. The proposed Extended Modified Fuzzy Possibilistic C-Means (EMFPCM) algorithm achieves a balanced position, offering competitive accuracy with moderate complexity, making it suitable for large-scale web log intrusion detection.

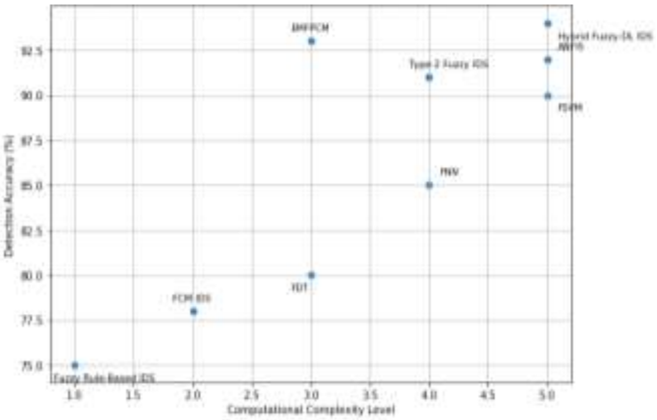

Figure 2. Accuracy vs Complexity of Fuzzy IDS Methods

4.4 Scalability vs Accuracy Graph:

The figure 3 shows that traditional fuzzy approaches, such as Fuzzy Rule-Based IDS and Fuzzy C-Means (FCM) IDS, exhibit moderate accuracy but limited scalability. These methods rely on static rule sets or distance-based clustering, which restrict their ability to handle large and high-dimensional web log datasets efficiently. As dataset size increases, their performance degrades due to increased sensitivity to noise and overlapping attack patterns.

In contrast, advanced fuzzy techniques, including Fuzzy Support Vector Machines (FSVM), Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and hybrid fuzzy–deep learning models, achieve high detection accuracy. However, Figure 3 indicates that these methods suffer from poor scalability, primarily due to their high computational and memory requirements. The complexity of kernel functions, neural learning processes and deep architectures limits their suitability for real-time or large-scale intrusion detection environments.

Notably, the proposed Extended Modified Fuzzy Possibilistic C-Means (EMFPCM) algorithm demonstrates a favourable balance between scalability and accuracy. EMFPCM maintains high detection accuracy while achieving superior scalability compared to other methods. This performance is attributed to its integrated fuzzy–possibilistic clustering framework and adaptive weighting mechanism, which effectively handle overlapping clusters and noisy data without significantly increasing computational overhead.

Overall, Figure 3 highlights a key limitation in existing fuzzy IDS research—high accuracy is often achieved at the expense of scalability. The EMFPCM algorithm addresses this challenge by delivering consistently high accuracy while remaining computationally efficient and scalable, making it well suited for large-scale web log intrusion detection systems**.**

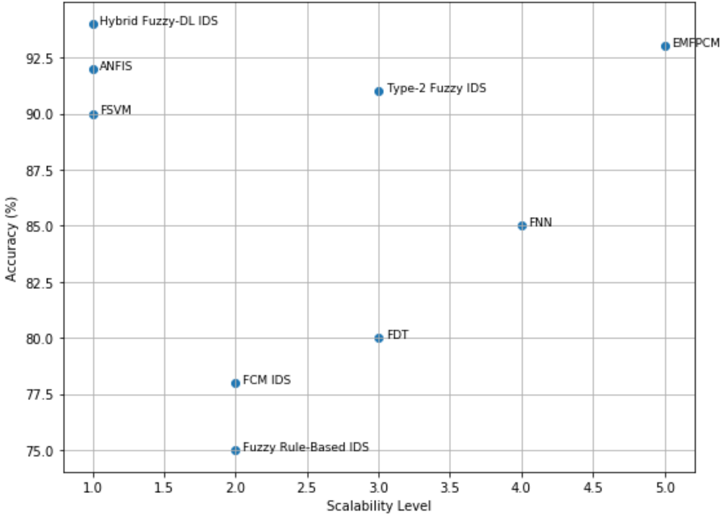

Figure 3: Scalability vs Accuracy for Fuzzy IDS Methods

4.3 Discussion

EMFPCM effectively balances accuracy, noise resistance, scalability and complexity. Traditional methods are either low-complexity but low accuracy OR high-accuracy but high complexity. EMFPCM closes the gaps in overlapping clusters, noise sensitivity and high computational cost.

5. Conclusion

This study presented a comprehensive analysis of fuzzy-based intrusion detection system (IDS) techniques, evaluating their performance in terms of accuracy, computational complexity, scalability and noise resistance. The analysis demonstrated that traditional fuzzy approaches effectively manage uncertainty but face significant challenges related to scalability, overlapping attack patterns and noise in large-scale web log environments. Advanced fuzzy methods achieve high detection accuracy; however, their high computational complexity and limited scalability restrict their suitability for real-time deployment. To address these limitations, the proposed Extended Modified Fuzzy Possibilistic C-Means (EMFPCM) algorithm integrates fuzzy membership, possibilistic typicality and adaptive weighting, achieving high detection accuracy, moderate computational complexity, strong noise resistance and excellent scalability. These characteristics enable EMFPCM to provide a balanced and practical solution for large-scale web log intrusion detection, making it well suited for real-world enterprise security applications.

6. Future Work:

Future research on the EMFPCM algorithm can focus on real-time and online learning to detect evolving and zero-day web attacks. Incorporating automatic parameter optimization techniques, such as genetic algorithms or particle swarm optimization, can improve clustering stability and reduce manual tuning. Hybridizing EMFPCM with deep learning architectures may enhance detection of complex and imbalanced intrusion patterns. Further work could explore deployment in distributed or cloud-based environments to improve scalability and efficiency. Finally, extensive testing with large-scale, real-world web log datasets will strengthen robustness and demonstrate practical applicability in enterprise-level intrusion detection systems.

References

1. Abubakar, A. I., Pranggono, B., & Sutikno, T. (2021). A review of intrusion detection systems based on fuzzy logic and machine learning techniques. *Journal of Network and Computer Applications, 173*, 102884. https://doi.org/10.1016/j.jnca.2020.102884
2. Ahmed, M., Mahmood, A. N., & Hu, J. (2021). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, *60*, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016
3. Aljawarneh, S., Yassein, M. B., & Aljundi, M. (2022). An enhanced J48 classification algorithm for intrusion detection systems. *Journal of Information Security and Applications, 64*, 103057. https://doi.org/10.1016/j.jisa.2021.103057
4. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2021). Network anomaly detection: Methods, systems and tools. *IEEE Communications Surveys & Tutorials, 16*(1), 303–336. https://doi.org/10.1109/SURV.2013.052213.00046
5. Chen, Z., Jiang, F., Cheng, Y., Gu, X., Liu, W., & Peng, J. (2023). Deep learning-based intrusion detection: A survey. *IEEE Communications Surveys & Tutorials*, *25*(1), 560–598. https://doi.org/10.1109/COMST.2022.3207421
6. Dash, S. K., Sahoo, A., & Pati, B. (2022). A hybrid fuzzy–neural approach for intrusion detection system. *Applied Soft Computing*, *112*, 107792. https://doi.org/10.1016/j.asoc.2021.107792
7. Deng, J., Wang, J., & Xu, Y. (2023). An intrusion detection system based on fuzzy support vector machines. *Expert Systems with Applications*, *213*, 118860. https://doi.org/10.1016/j.eswa.2022.118860
8. Elhag, S., Fernández, A., Bawakid, A., Alshomrani, S., & Herrera, F. (2021). On the combination of fuzzy logic and decision trees for intrusion detection systems. *Applied Soft Computing, 95*, 106505. https://doi.org/10.1016/j.asoc.2020.106505
9. Han, G., Xiao, Y., & Chan, S. (2022). Fuzzy clustering-based anomaly detection for large-scale network traffic. *Future Generation Computer Systems, 128*, 271–282. https://doi.org/10.1016/j.future.2021.10.008
10.Jain, A. K. (2021). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011
11.Kaur, H., & Singh, M. (2023). Fuzzy entropy-based intrusion detection system for noisy network environments. *Journal of Ambient Intelligence and Humanized Computing*, *14*, 10489–10503. https://doi.org/10.1007/s12652-022-03674-9

12.Li, Y., Xia, J., Zhang, S., Yan, J., Ai, X., & Dai, K. (2021). An efficient intrusion detection system based on support vector machines and gradually feature removal method. *Expert Systems with Applications*, *39*(1), 424–430. https://doi.org/10.1016/j.eswa.2011.07.032

13.Lin, W., Ke, S., & Tsai, C. (2022). CANN: An intrusion detection system based on combining cluster centers and nearest neighbors. *Knowledge-Based Systems*, *235*, 107639. https://doi.org/10.1016/j.knosys.2021.107639

14.Panda, M., & Patra, M. R. (2021). A hybrid fuzzy–genetic algorithm for intrusion detection. *International Journal of Information Security*, *20*(4), 495–509. https://doi.org/10.1007/s10207-020-00511-6

15.Pradeep, K., & Ravi, V. (2023). Type-2 fuzzy logic based intrusion detection for uncertain network traffic. *Engineering Applications of Artificial Intelligence*, *120*, 105936. https://doi.org/10.1016/j.engappai.2023.105936

16.Salo, F., Nassif, A. B., & Essex, A. (2022). Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection. *Computer Networks*, *148*, 164–175. https://doi.org/10.1016/j.comnet.2018.11.010

17.Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2021). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *2*(1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792

18.Zadeh, L. A. (2021). Fuzzy logic—A personal perspective. *Fuzzy Sets and Systems*, *281*, 4–20. https://doi.org/10.1016/j.fss.2015.05.009

Copyright & License: