

# Fuzzy Misuse Detection: Enhancing Intrusion Detection Systems with Data Mining and Machine Learning Technique

Madhuri Kanojiya<sup>1</sup>, Lokesh Chouhan<sup>2</sup>

<sup>1,2</sup>National Institute of Technology,  
Hamirpur, Himachal Pradesh, 177005, India

**Abstract**— Intrusion detection system (IDS) is the main defence which concurrence with firewalls and other security systems are implementing to detect the intrusions, attacks, and unauthorized misbehaviour. Misuse detection is the part of IDS which can be prevent and detect from the security attacks or intrusions. This paper using fuzzy misuse detection approach to detect the attacks or intrusion and describe the various data mining and machine learning techniques.

**Keywords**— Intrusion Detection, Fuzzy Logic, Neuro-Fuzzy, ANFIS, FCM, Feature Selection.

## I. INTRODUCTION

Intrusion detection systems are important and critical components of the security system infrastructures to improve the security in the computer systems. IDSs can be defined into the network-based IDS(NIDS) and host-based IDS(HIDS) [1]. NIDS effort to defend the network from the different types of attacks or intrusions and HIDS intended to distinguish the abnormal patterns of the events in host [1-4].

AN IDS, generally having three methods to detection the intrusion which are indicated as misuse detection, anomaly detection and hybrid method [5-9]. Misuse detection methods distinguish the intrusion based on the pre-specified attack signature and patterns; anomaly detection IDS schema normal behaviour profiles should be defined [10-12]. Any difference in the profiles as intrusion and consider as

## II. RESEARCH BACKGROUND

As Shown in figure- 2, IDSs system using the learning methods in which they illustrate as supervised, semi-supervised and unsupervised learning [15]. IDS system based on supervised learning are more accurate and maybe inaccessible because of it hardly depend on the label dataset, unsupervised learning having high false positive alert rate and do not required label datasets. Semi-

new type of attack. Demerits of IDS can't protect the hosts against the new type of attack.

This paper presents such a survey and taxonomy of the fuzzy signature-based IDS. Fuzzy logic is the fuzzy set theory that was introduced in 1965 by Lotfi Zadeh. Fuzzy set is a characterizati of indefinite and unspecified information in mathematical.

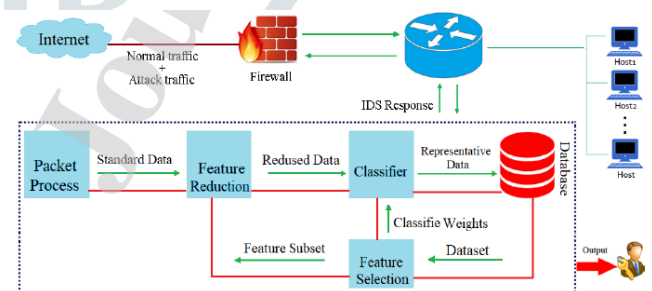


Figure 1: Architecture of Network IDSs

Fuzzy IDS using the fuzzy logic techniques such as fuzzy classification, fuzzy clustering, neuron fuzzy, ANFIS, fuzzy feature extraction and so on. Such techniques having ability to detect the malicious behaviour, intrusions at unsur data [16-18]. ANFIS fuzzy or neuron-fuzzy models that uses for leaning algorithm such as backpropagation, grid partition method, subtractive clustering and so on. Fuzzy clustering improves the IDS accuracy and performance by detect the misuse detection. Fuzzy clustering applies various clustering methods such as A-intuitionistic and K-mean clustering algorithm to produce the alerts with less false alter for new malware.[18-22]

supervised learning using the features of both supervised and unsupervised leaning by using partially labels datasets. IDS monitoring the network, host or both [18]. Host-based ids do not need to any special hardware device and use host's resource to detect the intrusion by monitoring the audit log, network event, and system call. Network IDS system analysis and sniff the traffic packet for detecting the anomalies and intrusion. NIDS need a special hardware and host, it monitoring the packets

crossing the network gateway and can't defend and detect the attack from any other network [19-22].

IDS can be grouped as active and passive approaches. it generates alert message and block the intruders. In other hand passive mode, the IDS logs happening, and network administrator or advise the security. In operation mode, IDS can group into online and offline. Online based-IDS monitoring the packet form network and detect the intrusion, performance depend upon the number of feature use to monitoring the packet. other hand in offline based-IDS using logs files dataset to detect the intrusion [14-18].

In high proportions IDS dataset, using the feature selection or extraction to remove the redundant, irrelevant and noise features. figure -2, feature extraction can be classified as filter, wrapper based and embedded process. Filter based process target the native properties that measured by PCA, information gain, scanning component methods and so on. wrapper based method using the classifier output to measure the value of features and subset of features.

DARPA dataset have been used in the IDS literature and having broad variety of intrusions pattern and signature in a military network environment [17=19]. It classified into TCPDUMP and BSM (Basic security model), in which TCPDUM dataset are collect on an imitated LAN and contain data packets, other hand BSM consists the logs files of execution of system calls. KDDCUP and NSL-KDD datasets are derived from the DARPA dataset by eliminating the unessential records. KDDCUP data is created by manipulating

the TCPDUM dataset, which is the part of DARPA dataset, with 38 numeric features. These features categories as host-based traffic features, basic features, time-based traffic features and content features. In which dataset consists the training dataset and test dataset, training dataset having 4,940K data illustration of normal traffic with 24 attacks. Test dataset having 311029 data illustration of normal traffic with 38 attacks. Often training dataset used 10-20% due to very large dataset. Figure -3 describe the percentage record of attacks in KDDCUP dataset. As shown in figure- 3, Security attacks are divided into 4 categories:

1. DoS (denial of service)
2. R2L (Remote to local)
3. U2R (User to Root)
4. Probe attack

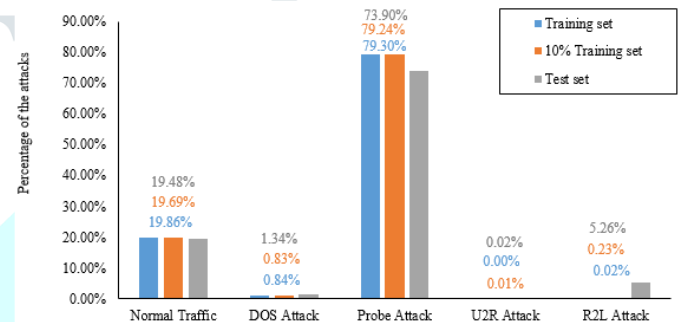


Figure-3: Attacks percentage in the KDDCUP dataset

NSL-KDD dataset is created by reducing the duplicate records in KDDCUP dataset and consists 41 features of KDDCUP dataset.

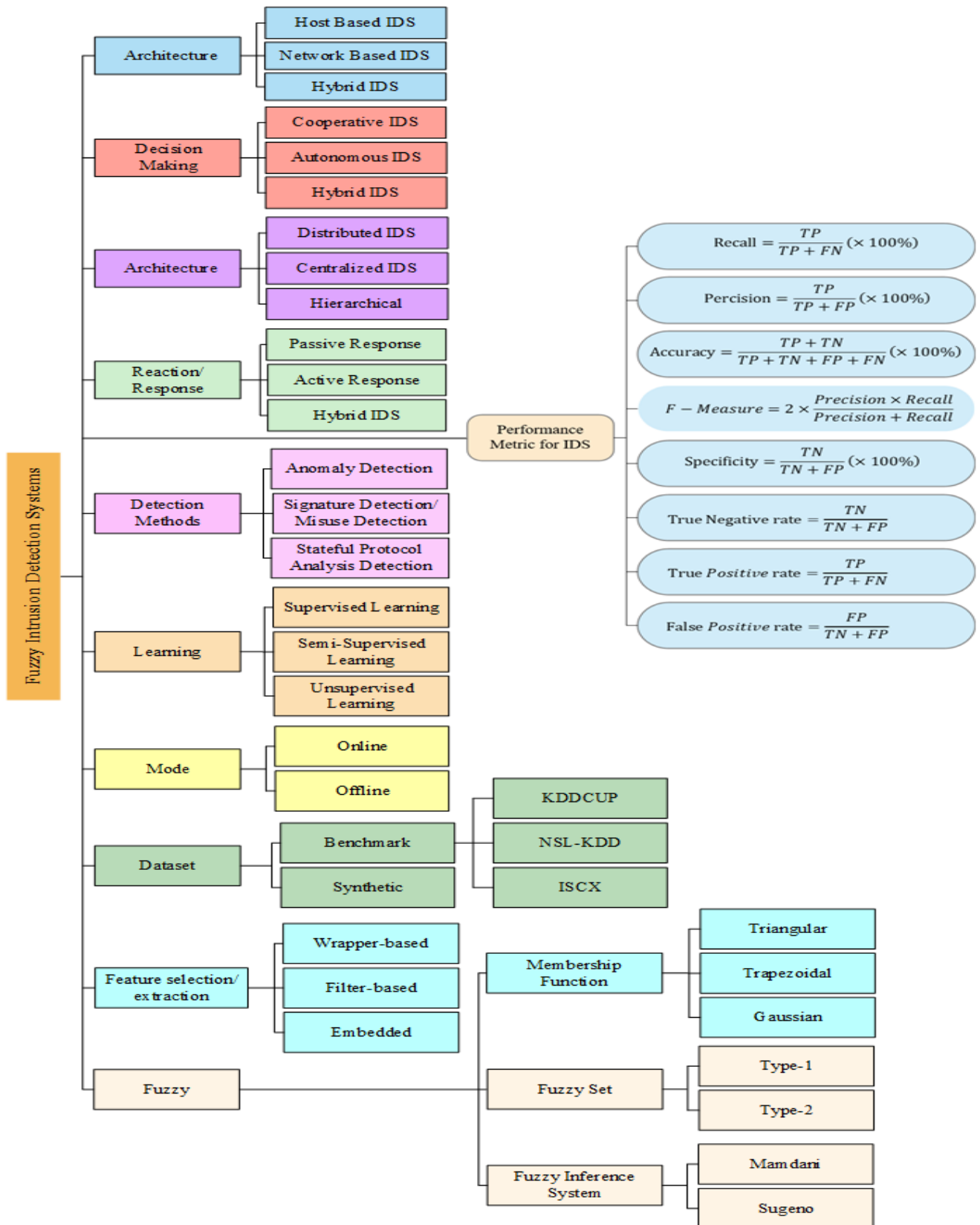


Figure -2: IDSs properties

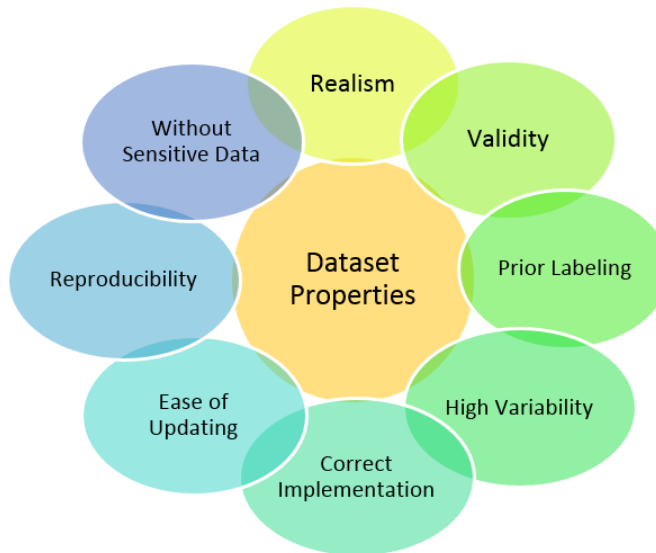


Figure-4: An ideal dataset property

After selecting the dataset and evaluating the IDS solution on it. In testing stage outcome of an IDS can be divided into four classes.

1. **True negative:** Normal traffic analysis as normal traffic.
2. **True positive:** Attack or intrusion traffic analysis as attack or intrusion traffic.
3. **False negative:** normal traffic analysis as attack or intrusion traffic.
4. **False positive:** attack or intrusion traffic analysis as normal traffic.

### III. FUZZY APPROACHES TO IDS

Numerous fuzzy approaches are classified in the IDS literature, such as ANFIZ fuzzy, neuro-fuzzy, fuzzy clustering, fuzzy feature selection and fuzzy classified. These fuzzy techniques and algorithms are used to detect the misuse detection.

#### A. ANFIS OR NEURO-FUZZY APPROACH TO IDS

ANFIS or neuro-fuzzy system intended to learn ANNs features with fuzzy logic. ANFIS using the backpropagation, least square approach to membership function parameters of tuning, subtractive clustering and grid partition. These algorithms using for learning. In grid partition learning approach is best with few membership functions and input for system, in which grid partition method create rules according to membership function. Subtractive clustering is the fast algorithm to find the clustering centroids.

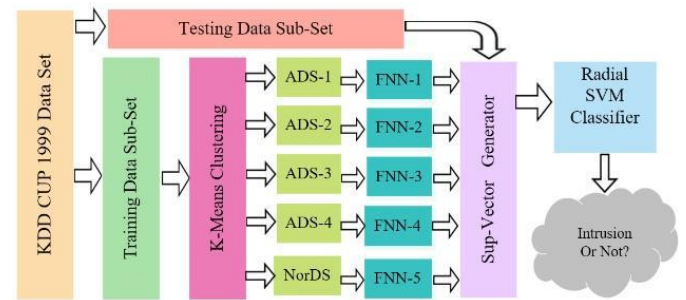


Figure-5: Block diagram of IDS solution

For feature extraction and train Using KDDCUP 1999. ANNIS using the ABD hybrid artificial colony and aABC algorithm for train the dataset. it identified the malware using fuzzy logic and learn the fuzzy rules to detect the new malware.

Improving the cluster convergence to create fuzzy rules to input data and increase the accuracy.

K-mean clustering algorithm generating the tuning parameters subsets. These subsets using to train the various neuro-fuzzy models, SVM classifier identified the intrusion and handling the security attacks with high detection rate. Block diagram as shown in figure -5, provide the solution of misuse detection.

#### B. FUZZY CLUSTERING APPROACH TO IDS

Fuzzy clustering is best approach to detect the misuse detection in IDS, to improve accuracy and performance of IDS. A-intuitionistic FCM method using for clustering to detect the new malware.

FC-ANN method using ANN model and fuzzy clustering to improve the detection rate with less false positive rate. ANN algorithm trained the model to improving the ability of learning. Fuzzy clustering method is use for generating the tuning subset with various parameters. Fuzzy aggregation method aggregating result of ANN models.

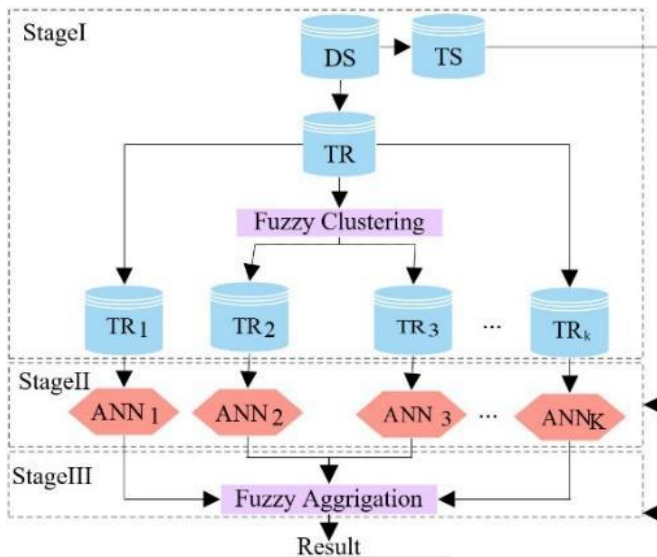


Figure-6: FC-ANN architecture

FCM is a semi-supervised solution based on the pairwise constraints. It is improving the detection rate with ability of learning of system. FCM and KNN algorithm using to reconstruct feature vector and train classifier. It reduces the false positive rate.

Active SVM learning and FCM clustering are hybrid semi-supervised learning approach as shown in figure 7. It is implementing on the NSLKDD IDS dataset. Approach tried to minimize the false positive and false negative rate by using two phase hybrid semi-supervised learning to detect the misuse and anomaly detection. In first phase, implement the FCM clustering algorithm for the abnormal and normal data. In second phase implement the 2 KNN classifier, First KNN classifier for misuse detection and tried to identify the false negative, other KNN classifier for abnormal detection and identified the false positive. finally, both KNN classifier identified the normal and attack traffic.

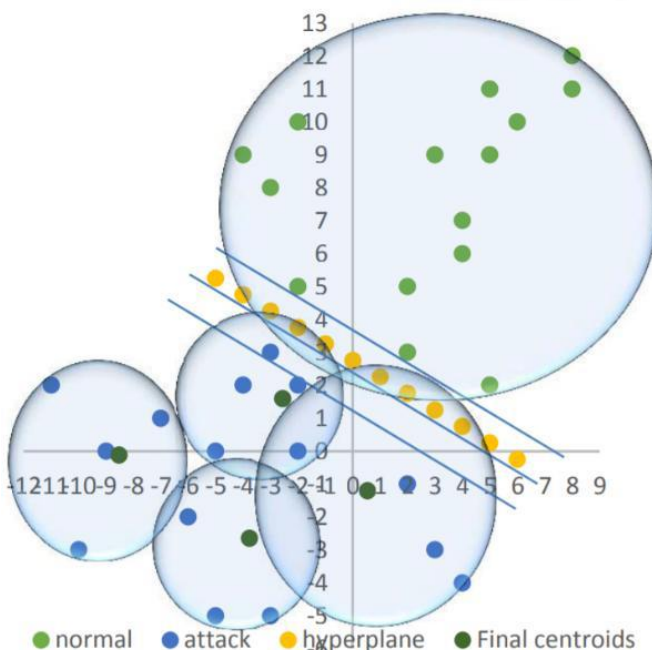


Figure-7: clustering based SVM

### C. FUZZY FEATURE SELECTION APPROACH TO IDS

In Fuzzy feature selection author introduces the heuristic to identify the smallest set of network traffic features, and using the classifier such as random forest, JRIP rules, decision tree and random tree. It identified the account hijacking, probing and DDoS attacks.

### D. FUZZY CLASSIFIER APPROACH TO IDS

Fuzzy classifier approach is based on the semi-supervised hybrid learning. It is having the ability to identify the both external and internal malicious behaviours. To identifying the external malicious behaviour, implement the fuzzy GA (Genetic algorithm), and for internal malicious behaviour detection using the signature match approach.

IDS using GA, clustering and ANN algorithms in order to improving the accurately detect the intrusion. Genetic fuzzy model is the efficient approach for feature selection and minimized the irrelevant features. It is using multi-gene genetic algorithms, Naïve Bayes classifier and NSL-KDD dataset.

## IV. CONCLUSION

misuse and anomaly detection are the part of intrusion detection paradigm and detect the attacks or intrusions by applying the signature matching approach. For improving the accuracy and performance of IDS, integrated the fuzzy approach to detect the misuse or anomaly detection. Fuzzy approach using various algorithm and techniques, they define as fuzzy logic, fuzzy clustering, fuzzy feature selection, fuzzy classifier and so on. Motivation of this paper is to represent the comprehensively survey and taxonomy of fuzzy based misuse and anomaly detection IDS. It maximized the normal packets is correctly identify as normal packets and intrusion packets is correctly identify as intrusion packets or minimized the normal packets is identify as intrusion packets and intrusion packets is identify as normal packets. Utilize the fuzzy approach and summarized the features and focus the contribution of fuzzy techniques to control the intrusion and attacks in IDS.

## REFERENCES

- [1] M. Masdari and H. Khezri, A survey and taxonomy of the fuzzy signature-based Intrusion Detection Systems, *Applied Soft Computing Journal* (2020)
- [2] A. Sung, S. Mukkamala, "Identifying important features for intrusion detection using support vector machines and neural networks" in *Symposium on Applications and the Internet*, 2003, pp. 209–216
- [3] A. Branitskiy and I. Kotenko, "Network attack detection based on combination of neural, immune and neuro-fuzzy classifiers," in *Computational Science and Engineering (CSE), 2015, IEEE 18th International Conference on, 2015, pp. 152-159.*
- [4] A. Chandrasekhar and K. Raghuvver, "Intrusion detection technique by using k-means, fuzzy neural network and SVM classifiers," in *Computer Communication and Informatics (ICCCI), 2013 International Conference on, 2013, pp. 1-7.* [41] C. Azad and V. K. Jha, "Fuzzy min-max neural.
- [5] Z. Liu, W. Wei, H. Wang, Y. Zhang, Q. Zhang, and S. Li, "Intrusion Detection Based on Parallel Intelligent Optimization Feature Extraction and Distributed Fuzzy Clustering in WSNs," *IEEE Access, vol. 6, pp. 72201-72211, 2018.*
- [6] Gao Xiang, Wang Min, Zhao Rongchun, "Applying Fuzzy Data Mining to Network Unsupervised Anomaly Detection", *ISCIT, 2005, IEEE Computer pp. 1249-1253*
- [7] S. Sai Satyanarayana Reddy, P. Chatterjee, and C. Mamatha, "Intrusion Detection in Wireless Network Using Fuzzy Logic Implemented with Genetic Algorithm," *Singapore, 2019, pp. 425-432.*
- [8] Ming-Yang Su , Chun-Yuen Lin , Sheng-Wei Chien and Han-Chung Hsu, "Genetic-Fuzzy Association Rules for Network Intrusion Detection Systems", *IEEE International Conference on Fuzzy Systems, 2011, Taipei, Taiwan*
- [9] R. Shanmugavadivu and Dr. N. Nagarajan, "Network Intrusion Detection System using Fuzzy Logic", *IJCSE, 2011, ISSN: 0976-5166 Vol. 2 No. 1*
- [10] T. Taerat, B. Baler, "Deterministic lossless log message clustering tool," in 2011, pp. 3-4.
- [11] Lokesh Chouhan and H.S. Lalventhangi, "Adaptive Energy Detection Based Solution for Fronthaul Problem in C-RAN," *Wireless Personal Communications (WPC)*, Springer Publication, vol 103, 2018, pp. 2743–2755, ISSN: 0929-6212, 14 Sep, 2018. DOI: [10.1007/s11277-018-5960-6](https://doi.org/10.1007/s11277-018-5960-6).
- [12] Lokesh Chouhan and Aditya Trivedi, "Performance Study of a CSMA based Multiuser MAC Protocol for Cognitive Radio Networks," *Wireless Networks (WINE)*, Springer Publication, vol 22, no 1, 2016, pp. 33-47, ISSN: 1022-0038. DOI: [10.1007/s11276-015-0947-7](https://doi.org/10.1007/s11276-015-0947-7).
- [13] Jayanti Rastogi, Lokesh Chouhan, and Aditya Trivedi, "Multichannel CSMA Based MAC Scheme for Unsaturated Cognitive Radio Networks," *Wireless Personal Communications (WPC)*, Springer Publication, vol 85, no 3, 2015, pp. 1279-1294, ISSN: 0929-6212. DOI: [10.1007/s11277-015-2840-1](https://doi.org/10.1007/s11277-015-2840-1).
- [14] Lokesh Chouhan and Aditya Trivedi, "MAC Layer Protocols for Cognitive Radio Network," *Self Organization and Green Applications in Cognitive Radio Networks*. IGI Global, 2013, pp. 154-189. Web. 5 Jul. 2013. ISBN13: 9781466628120, DOI: [10.4018/978-1-4666-2812-0](https://doi.org/10.4018/978-1-4666-2812-0).
- [15] Rajni Dubey, Sanjeev Sharma, and Lokesh Chouhan, "Security for Cognitive Radio Networks," *Cognitive Radio and Interference Management: Technology and Strategy*. IGI Global, 2013, pp. 238-256. Web. 18 Nov. 2011. ISBN13: 9781466620056, DOI: [10.4018/978-1-4666-2005-6](https://doi.org/10.4018/978-1-4666-2005-6).
- [16] Wali Ullah Farooqui and Lokesh Chouhan, "Coordinated Multi-Robot Navigation Using Sectorization of Environment," *Conference on IT in Business, Industry and Government (CSIBIG) 2014*, IEEE, CSI Indore Chapter, pp. 1-6, 08-09 March 2014. ISBN: 978-1-4799-3064-7. DOI: [10.1109/CSIBIG.2014.7057009](https://doi.org/10.1109/CSIBIG.2014.7057009).
- [17] Purushottam, Aditya Trivedi, and Lokesh Chouhan, "Channel Allocation and Resource Optimization in Cognitive Radio Cloud Network," *Conference on Advances in Mobile Communications, Networking and Computing*, organized by ICEIT, New Delhi, 27 – 28 September, 2013.
- [18] Lokesh Chouhan and Aditya Trivedi, "Analysis of MAC Schemes for Cognitive Radio Network: Perfect and Imperfect Learning Modelling," *Proceedings of 10th IEEE International Conference on Wireless and Optical Communications Networks (WOCN-2013)*, July 26-28, 2013, Bhopal, India, pp. 1-6. ISBN: 151-7703. DOI: [10.1109/WOCN.2013.6616247](https://doi.org/10.1109/WOCN.2013.6616247).
- [19] Lokesh Chouhan and Aditya Trivedi, "Priority based MAC scheme for cognitive radio network: A queuing theory modeling," *Proceedings of 9th IEEE International Conference on Wireless and Optical Communications Networks (WOCN-2012)*, September 20-22, 2012, Indore, India, pp. 1-5. ISBN: 151-7703. DOI: [10.1109/WOCN.2012.6331886](https://doi.org/10.1109/WOCN.2012.6331886).
- [20] Rajni Dubey, Sanjeev Sharma, and Lokesh Chouhan, "Secure and Trusted algorithm for Cognitive Radio Network," *Proceedings of 9th IEEE International Conference on Wireless and Optical Communications Networks (WOCN-2012)*, September 20-22, 2012, Indore, India, pp. 1-7. ISBN: 151-7703. DOI: [10.1109/WOCN.2012.6331887](https://doi.org/10.1109/WOCN.2012.6331887).
- [21] Lokesh Chouhan and Aditya Trivedi, "Cognitive radio networks: Implementation and application issues in India," *Seminar on Next Generation Network - Implementation and Implication*, Telecom Regulatory Authority of India (TRAI), Govt. of India, New Delhi, 25-26th August, 2011. Web. [PDF link](#).
- [22] Lokesh Chouhan and Sanjeev Sharma, "Implementation of RSA shared key algorithms to secure Mobile Ad Hoc Networks," *Proceeding of IEEE International Conference on recent trends in soft computing and information technology*, Bhopal, pp. 416-421, 7-8 Jan 2010