SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
SCHOOL OF COMPUTING
DEPARTMENT OF COMPUTING TECHNOLOGIES
18CSP107L / 18CSP108L - MINOR PROJECT / INTERNSHIP

Identification and Counteraction of DDOS, DOS, and MITM utilizing AI calculations

Batch ID: 3

Guide name: Dr. S. Nikkath Bushra

Designation: Assistant Professor

Department: Computing Technologies

Student Reg. No: RA2111003011799

Student Name: V. Venkat Aditya

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Introduction

- As cyberspace evolves rapidly, increasing cyber threats pose serious risks to the confidentiality, integrity, and availability of sensitive information.
- DDoS, DoS, and Man-in-the-Middle (MitM) attacks are major challenges for cybersecurity professionals.
- This project leverages machine learning algorithms for advanced detection and prevention of DDoS, DoS, and MitM attacks.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Problem Statement

In the rapidly evolving landscape of cybersecurity, Distributed Denial of Service (DDoS) and Denial of Services (DoS) attacks pose significant threats to the availability and reliability of online services. These malicious activities aim to overwhelm target systems or networks, rendering them inaccessible to legitimate users. Traditional methods of detecting and mitigating DDoS and DoS attacks, such as rule-based or signature-based approaches, often fall short in addressing the dynamic and sophisticated nature of these threats. The challenge lies in developing innovative and adaptive solutions that can effectively differentiate between normal and malicious traffic patterns, thereby enhancing network security and minimizing the impact of such attacks on critical services and infrastructure.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Objectives

- Data Collection
- Feature Engineering
- Data Preprocessing
- Model Selection
- Model Tuning
- Performance Comparison
- Scalability and Efficiency

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Literature review

S. No.	Title	Theme Description	Algorithms	Pros
1	Resilient Event-Triggered H∞ Control Under DoS Attacks With an Application to Offshore Structures(2023)	the studied system is integrated into a class of switching time-delay systems, and a series of sufficient conditions are derived to ensure the internally exponential stability of the system and to design the resilient event-triggered H∞ controllers.	Random Forest Algorithm K-means clustering	The proposed algorithm has accuracy value of 81.93% and 69.75% respectively.
2	Distributed Event-Triggered Consensus-Based Control of DC Microgrids in Presence of DoS Cyber Attacks(2023)	this mechanism, transmissions occur while a specified event is triggered to prevent unessential utilization of communication resources. The asymptotic stability of the ET-based controller is shown formally by using Lyapunov stability via linear matrix inequality (LMI) conditions.	Decision tree K-means clustering	The proposed algorithms has accuracy value of 89.36% and 73.02% respectively.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Literature review

S. No.	Title	Theme Description	Algorithms	Pros
3	Observer-Based Secure Control for Vehicular Platooning Under DoS Attacks(2023)	To deal with DoS attacks, we consider an observer-based mechanism to estimate the state of vehicles based on available sensor measurements which significantly improves the resilience and tolerance of platooning system during the attack interval.	K-Means Clustering ERF-KMC Algorithm	The proposed algorithm has accuracy value of 85.33% and 83.75% respectively.
4	Low-Rate DoS Attacks, Detection, Defence, and Challenges: A Survey(2020)	We classify the LDoS attacks and existing defence methods according to time domain and frequency domain in which detection and defence are performed.	Naive Bayes Support Vector Machine	The proposed algorithms has accuracy value of 75.33% and 63.75% respectively.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Literature review

S. No.	Title	Theme Description	Algorithms	Pros
5	Detection and Mitigation of Low-Rate Denial-of-Service Attacks: A Survey(2022)	Evolving type of attack, known as LDoS (Low-rate Denial of Service) attacks, has the potential to produce more damage than its predecessor due to its stealth nature and the lack of suitable detection and defense methods	Support Vector Machine (SVM) Random Forest (RF)	The proposed algorithms has accuracy value of 93.80% and 88.55% respectively.
6	Anomaly Detection IDS for Detecting DoS Attacks in IoT Networks Based on Machine Learning Algorithms (2022)	According to the adaptive threshold technique, a resilient adaptive event-triggered mechanism (AETM) is developed to reduce the transmission frequency and also combating the aperiodic DoS attacks.	Random Forest Support Vector Machines	The proposed algorithms has accuracy value of 99.80% and 98.55% respectively.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Literature review

S. No.	Title	Theme Description	Algorithms	Pros
7	MSR-DoS: Modular Square Root-Based Scheme to Resist Denial of Service (DoS) Attacks in 5G-Enabled Vehicular Networks(2022)	this paper proposes modular square root-based to resist denial of service (DoS) attacks (MSR-DoS) scheme in 5G-enabled vehicular networks.	Modular Square Root (MSR) Algorithms	The proposed MSR-DoS scheme reduces the computation overhead of signing the message and verifying the message by 99.80% and 98.55%, respectively.
8	An Efficient IDS Framework for DDoS Attacks in SDN Environment(2021)	this mechanism, transmissions occur while a specified event is triggered to prevent unessential utilization of communication resources. The asymptotic stability of the ET-based controller is shown formally by using Lyapunov stability via linear matrix inequality (LMI) conditions.	Software Defined Network (SDN) architecture.	The proposed algorithms has accuracy value of 89.36% and 73.02% respectively.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Proposed System/Methodology

A proposed system for the detection and prevention of Distributed Denial of Service (DDoS), Denial of Service (DoS), and Man-in-the-Middle (MITM) attacks leverages advanced machine learning techniques to enhance both detection accuracy and response speed. This system integrates real-time data collection from network traffic, analyzing it using machine learning models such as Support Vector Machines (SVMs) and K-Nearest Neighbor (KNNs), which are capable of identifying complex patterns that might signal an attack.

This proposed system employs a multi-layered approach, combining advanced machine learning techniques with real-time data analysis to detect and prevent DDoS, DoS, and Man-in-the-Middle (MITM) attacks. The methodology involves the various key components:

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Proposed System/Methodology

1. Real-Time Data Collection and Processing
    a. Data Sources - The system collects real-time data from various network traffic sources, including routers, firewall, intrusion detection/prevention systems (IDS/IPS), and network logs.
    b. Preprocessing - Raw data undergoes preprocessing steps, such as data cleaning, normalization, and feature extraction, to ensure high-quality inputs for machine learning models. Tools like Wireshark or Tshark can be used for packet capture, while Python libraries like Pandas or Scikit-learn are employed for preprocessing tasks.
2. Feature Engineering
    a. Key features such as packet size, source/destination IP addresses, frequency of requests, and protocol types are extracted and engineered to enhance the model's ability to identity attack patterns.
    b. Statistical and time-based features are also computed to capture anomalies in network behavior, using tools like NumPy and Scipy.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Proposed System/Methodology

3. Machine Learning Model Selection and Training
    a. Supervised Learning - The system initially employs supervised learning algorithms, such as Extreme Gradient Boosting (XGBoost) and K-Nearest Neighbors (KNN), to train on labeled datasets containing normal and attack traffic patterns.
    b. Unsupervised Learning - To detect unknown or novel attacks, the system integrates unsupervised learning techniques, like K-means clustering or Autoencoders, which can identify deviations from normal traffic behavior.
    c. Training is conducted using Scikit-learn and TensorFlow or Keras for deep learning-based models, with Jupyter Notebooks serving as the development environment.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Proposed System/Methodology

4. Model Evaluation and Tuning
   a. The models are evaluated on metrics such as accuracy, precision, recall, and F-1 score to ensure robust performance in distinguishing between normal and malicious traffic.
   b. Techniques like Cross-Validation and Grid Search are used to fine-tune hyperparameters, optimizing the models for better detection accuracy and response speed.
5. Real-Time Detection and Response
   a. The system continuously monitors incoming network traffic, applying the trained models to classify traffic in real time.
   b. Upon detection of an attack, the system dynamically adjusts network configurations such as routing paths, firewall rules, or access controls to mitigate the threat. SON (Software-Defined Networking) controllers, such as OpenDaylight or ONOS, are utilized for dynamic network adjustments.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Proposed System/Methodology

6. Anomaly and Signature-Based Hybrid Detection
   a. To enhance detection accuracy and reduce false positives, the system employs a hybrid detection approach by integrating anomaly-based detection with signature-based methods.
   b. Snorts or Suricata can be used for signature-based detection, which is complemented by the anomaly detection models to capture previously unseen attack patterns.
7. Continuous Learning and Adaptation
   a. The system incorporates a feedback loop where new attack patterns identified by the unsupervised models are fed back into the supervised learning process to update and retrain models periodically.
   b. This continuous learning mechanism ensures that the system evolves and adapts to emerging threats.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

Proposed System/Methodology

8. Deployment and Monitoring
   a. The entire framework is deployed on a scalable architecture using technologies like Docker and Kubernetes for containerization and orchestration, enabling seamless integration and scaling in different network environments.
   b. Prometheus and Grafana are used for monitoring the system's performance and attack detection metrics, providing real-time insights into the security posture.

This approach leverages a combination of traditional and advanced techniques, integrating various machine learning algorithms and real-time network management tools to provide a robust defense against sophisticated cyber threats.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

System Design

1. Architecture Diagram:

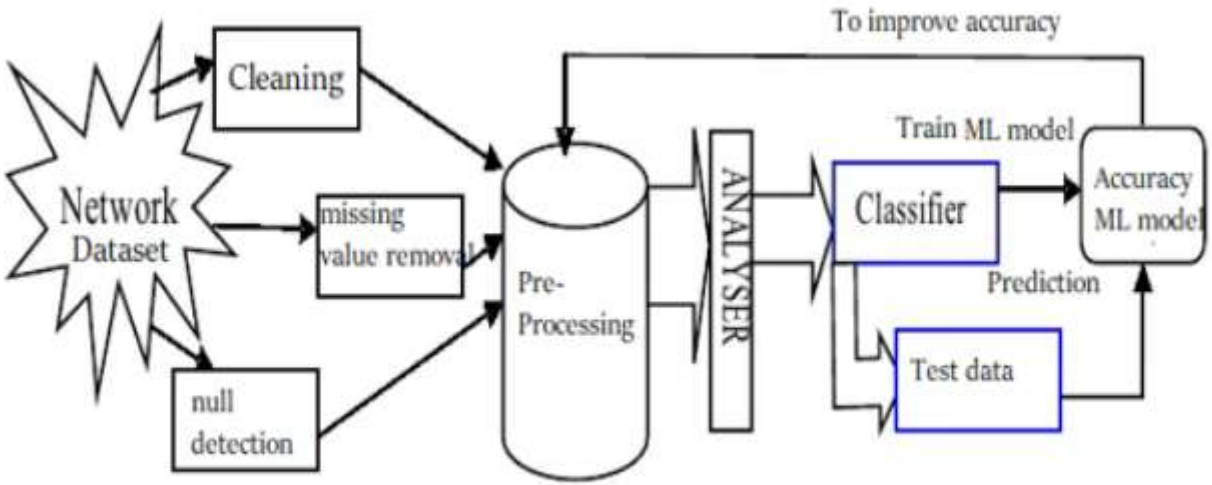

Fig-1

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

System Design

2. Data Flow Diagram:

Data flow diagrams are used to graphically represent the flow of data in a business information system. DFD describes the processes that are involved in a system to transfer data from the input to the file storage and reports generation. Data flow diagrams can be divided into logical and physical. The logical data flow diagram describes flow of data through a system to perform certain functionality of a business. The physical data flow diagram describes the implementation of the logical data flow.

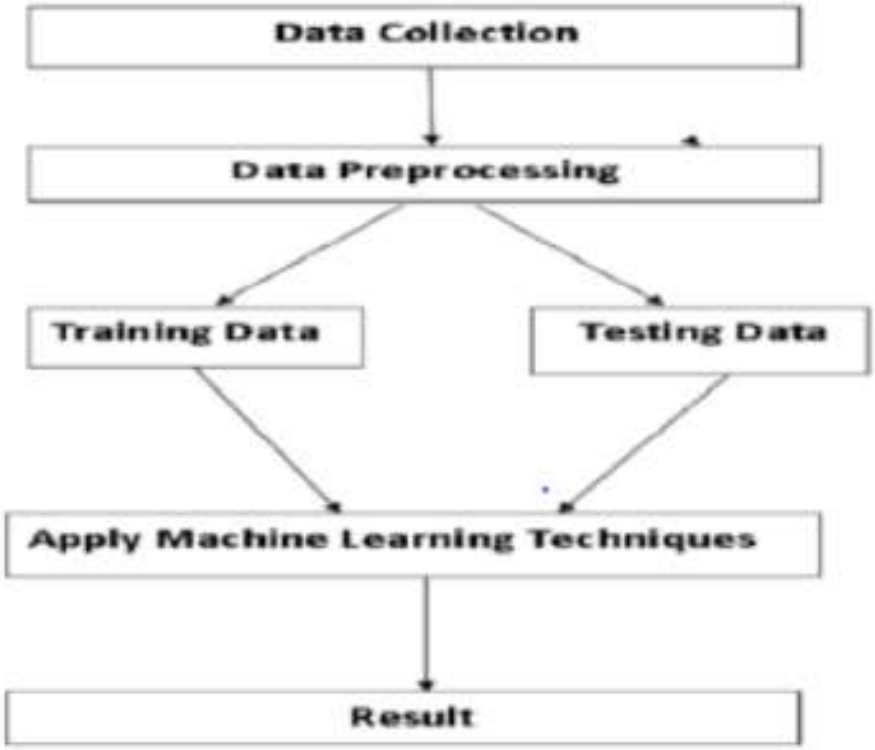

Fig-2

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

System Design

3. Class Diagram:

The class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing and documenting different aspects of a system but also for constructing executable code of the software application.

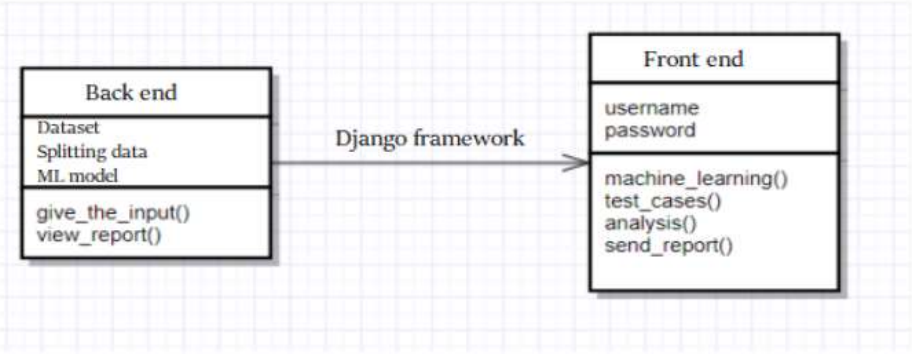

Fig-3

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

System Design

4. Activity Diagram:

The Activity Diagram forms effective while modeling the functionality of the system. Hence this diagram reflects the activities, the types of flows between these activities and finally the response of objects to these activities.

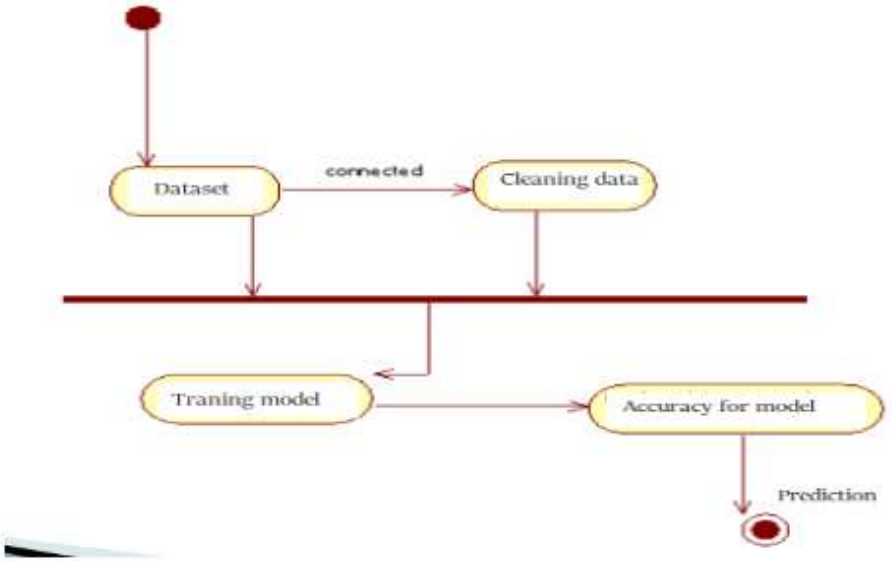

Fig-4

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

CHALLENGES

- Advancement of Attacks Strategies: Attacks such as DDoS, DoS, MiTM attacks are on the rise with several methods being employed by attackers to avoid detection systems. The system needs to learn how to pace itself with the trends in attacking and modify itself to appropriate the attack patterns as they occur.
- Strategies for Management of False Positives: A proper detection apparatus needs to ensure minimal false positive occurrences so as to preserve the normal flow of traffic across the network. This is only achievable by employing artificial intelligence techniques that can discriminate in a meaningful manner between the good and bad traffic.

Scalability: Because the size and complexity of networks has increased over time the detection system needs to be scalable so that it can process large amounts of information with ease. This includes improving the machine learning models' performance while maintaining their precision.

Anomaly Detection: Any flooding attack or MiTM campaign is likely to exhibit some unusual form of network traffic. For that reason, the system ought to be able to identify such anomalous patterns through the management of normal and abnormal behaviours.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

IMPLEMENTATION

1. KNN, or K-Nearest Neighbour

KNN is the most common algorithm in Machine learning, pattern recognition and several other areas. KNN is used to classification related jobs. This method is also called as an instance based algorithm (also termed as lazy learning). All training data samples are kept until new observations need to be classified; a model or classifier is not created immediately. This attribute of a lazy learning algorithm renders it superior than eager learning, which develops a classifier prior to the classification of fresh observations. This attribute of a lazy learning algorithm renders it superior than eager learning, which develops a classifier prior to the classification of fresh observations. This technique is particularly important when dynamic data necessitates rapid changes and updates. KNN was utilised with various distance measures.

Step I: Supply the feature space to the neural network for training.

Step II: Calculate distance utilising the Euclidean distance formula

Step III: Determine the values calculated using the Euclidean distance so that, for each $i = 1, 2, 3,.., k$, $d\,i \leq d\,i + 1$.

Step IV: Depending on the characteristics of the data, technique will be applied or vote cast.

Step V: The proportion and standard of data supplied to KNN determine the value of the constant K, which refers to the nearest neighbors, K. K value is kept low for small sized datasets and kept high for large sized datasets where appropriate.

IMPLEMENTATION

2. XGBoost

XGBoost is another machine learning algorithm that implements the idea of multiple small decision trees by considering their functional approximations. The role of each tree is to improve on the mistakes made by the previous trees, so with every cycle, the combination improves and the performance of the model overall increases. This procedure is referred to as boosting.

XGBoost is used to classify different types of network traffic (Benign, DDoS, DoS, and MiTM). It learns patterns in the data to identify each type correctly, though it sometimes gets confused between similar attack types (like DDoS and DoS).

i. Data Input: focuses on loading the network traffic data (packet details, traffic patterns) and spreadsheets into the XGBoost model and the available data is labeled as Benign, DDoS, DoS, MiTM.

ii. Initialize Trees: XGBoost procedure initially creates a basic decision tree with the aim of categorizing the given data. This tree may not be very precise in its classification at the beginning.

iii. Calculate Errors: Once the first tree completed the prediction process, XGBoost considers the outcome. Specifically, it pays attention to those points which have been classified correctly, as well as those that were erroneous. It concentrates on the errors.

iv. Build New Trees: As new trees are built; each tree seeks to correct the errors introduced by the previous tree. The trees are linearly combined, and the final output prediction is optimized.

v. Repeat the Process: This approach is repeated until enough trees are constructed by the model (according to the maximum number of boosting rounds). With every new tree, the classification becomes more accurate.

vi. Final Prediction: When all the trees have been created XGBoost utilises all available trees in order to formulate a prediction for each instance of network traffic (Benign, DDoS, DoS and MiTM).

vii. Evaluate the Model: This model can also be evaluated based on how well its predictions agree with actual outcomes determined from a confusion matrix, or accuracy rates. It will provide the number of instances that were marked the way and incorrectly so.

IMPLEMENTATION

2. XGBoost

XGBoost is another machine learning algorithm that implements the idea of multiple small decision trees by considering their functional approximations. The role of each tree is to improve on the mistakes made by the previous trees, so with every cycle, the combination improves and the performance of the model overall increases. This procedure is referred to as boosting.

XGBoost is used to classify different types of network traffic (Benign, DDoS, DoS, and MiTM). It learns patterns in the data to identify each type correctly, though it sometimes gets confused between similar attack types (like DDoS and DoS).

i. Data Input: focuses on loading the network traffic data (packet details, traffic patterns) and spreadsheets into the XGBoost model and the available data is labeled as Benign, DDoS, DoS, MiTM.

ii. Initialize Trees: XGBoost procedure initially creates a basic decision tree with the aim of categorizing the given data. This tree may not be very precise in its classification at the beginning.

iii. Calculate Errors: Once the first tree completed the prediction process, XGBoost considers the outcome. Specifically, it pays attention to those points which have been classified correctly, as well as those that were erroneous. It concentrates on the errors.

iv. Build New Trees: As new trees are built; each tree seeks to correct the errors introduced by the previous tree. The trees are linearly combined, and the final output prediction is optimized.

v. Repeat the Process: This approach is repeated until enough trees are constructed by the model (according to the maximum number of boosting rounds). With every new tree, the classification becomes more accurate.

vi. Final Prediction: When all the trees have been created XGBoost utilises all available trees in order to formulate a prediction for each instance of network traffic (Benign, DDoS, DoS and MiTM).

vii. Evaluate the Model: This model can also be evaluated based on how well its predictions agree with actual outcomes determined from a confusion matrix, or accuracy rates. It will provide the number of instances that were marked the way and incorrectly so.

SRM
INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

References

- V. D. M. Rios, P. R. M. Inácio, D. Magoni and M. M. Freire, "Detection and Mitigation of Low-Rate Denial-of-Service Attacks: A Survey," in IEEE Access, vol. 10, pp. 76648-76668, 2022, doi: 10.1109/ACCESS.2022.3191430.
- M. Sathishkumar and Y. -C. Liu, "Resilient Adaptive Event-Triggered Control for Singular Networked Cascade Control Systems Under DoS Attacks," in IEEE Access, vol. 10, pp. 89197-89210, 2022, doi: 10.1109/ACCESS.2022.3199697.
- M. A. Al-Shareeda and S. Manickam, "MSR-DoS: Modular Square Root-Based Scheme to Resist Denial of Service (DoS) Attacks in 5G-Enabled Vehicular Networks," in IEEE Access, vol. 10, pp. 120606-120615, 2022, doi: 10.1109/ACCESS.2022.3222488.
- L. Gao, Y. Li, L. Zhang, F. Lin and M. Ma, "Research on Detection and Defense Mechanisms of DoS Attacks Based on BP Neural Network and Game Theory," in IEEE Access, vol. 7, pp. 43018-43030, 2019, doi: 10.1109/ACCESS.2019.2905812.
- S. Khodadadi, T. K. Tasooji and H. J. Marquez, "Observer-Based Secure Control for Vehicular Platooning Under DoS Attacks," in IEEE Access, vol. 11, pp. 20542-20552, 2023, doi: 10.1109/ACCESS.2023.3250398.