A Comprehensive Survey on AI and ML Approaches for Detecting DDoS Attacks

Dr R Anurekha
Assistant Professor
Department of Information Technology
Institute of Road and Transport Technology, Erode – 638316, Tamil Nadu, India

Abstract: Distributed Denial of Service (DDoS) attacks have become one of the most persistent and evolving threats in cybersecurity, affecting organizations, governments, cloud computing environments and IoT networks. These attacks overwhelm target systems with massive amounts of malicious traffic, leading to service disruptions, financial losses and security breaches. Traditional detection mechanisms, such as signature-based and heuristic methods, struggle to cope with increasingly sophisticated and adaptive attack strategies. Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful tools in combating DDoS attacks, offering adaptive learning capabilities, anomaly detection and real-time response mechanisms. This paper presents a comprehensive survey of AI and ML methodologies used for DDoS detection, including supervised learning, unsupervised learning and deep learning approaches. Various algorithms such as decision trees, support vector machines, convolutional neural networks, recurrent neural networks and autoencoders, publicly available datasets, benchmarking strategies and real-world applications of AI-driven DDoS detection systems are explored. The survey also discusses major challenges, such as scalability, adversarial evasion techniques and interpretability concerns, while suggesting future research directions to enhance AI-based security solutions.

Keywords: DDoS detection, artificial intelligence, machine learning, anomaly detection, cybersecurity, deep learning, network security, supervised learning, unsupervised learning, real-time detection

I. INTRODUCTION

In the ever-expanding digital landscape, Distributed Denial of Service (DDoS) attacks have emerged as one of the most formidable cybersecurity threats. These attacks are designed to flood targeted networks, applications, or services with an overwhelming volume of malicious traffic, rendering them inaccessible to legitimate users. DDoS attacks are often orchestrated using botnets – large networks of compromised devices – allowing attackers to exploit vulnerabilities across various infrastructures, including cloud computing, IoT networks and enterprise systems. The increasing sophistication of these attacks necessitates the development of advanced detection and mitigation strategies.

DDoS attacks can be broadly categorized into three main types: volumetric attacks, protocol attacks and application-layer attacks. Volumetric attacks, such as UDP floods and ICMP floods, overwhelm network bandwidth, making services unavailable. Protocol attacks, including SYN floods and Ping of Death, exploit weaknesses in network protocols to exhaust server resources. Application-layer attacks, like HTTP floods and Slowloris, specifically target web applications, mimicking legitimate requests to evade detection. As attackers continue to develop multi-vector attack strategies that combine different techniques, conventional security mechanisms struggle to provide adequate protection.

Traditional DDoS detection mechanisms, such as signature-based, rate-limiting and heuristic approaches, struggle to keep pace with the evolving tactics of attackers. These conventional methods rely heavily on predefined rules and thresholds, making them less effective in identifying novel and adaptive attack patterns. Additionally, rule-based systems require constant updates, which may not be feasible in real-time attack scenarios. As cyber

threats become more complex and dynamic, there is a growing need for intelligent, scalable and real-time detection techniques that can adapt to new attack methodologies.

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as promising solutions to address these challenges. AI/ML-based models can analyse vast amounts of network traffic, identify anomalous behaviours and adapt to evolving attack patterns with minimal human intervention. These techniques leverage supervised learning, unsupervised learning and deep learning methodologies to classify, predict and mitigate DDoS attacks effectively. Supervised learning algorithms, such as decision trees and support vector machines, rely on labelled datasets to distinguish between normal and malicious traffic. Unsupervised learning methods, including clustering and autoencoders, excel in anomaly detection by recognizing deviations from established behavioural patterns. Deep learning approaches, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), further enhance detection capabilities by capturing complex traffic patterns and temporal dependencies.

AI-driven DDoS detection systems offer several advantages over traditional methods. They provide improved accuracy in identifying attacks, reduce false positives and enable automated responses to mitigate threats in real time. Additionally, these systems can detect zero-day attacks by learning from historical traffic patterns and identifying anomalies that deviate from expected behaviours. However, AI-based approaches also face challenges, including the need for large and diverse datasets, computational complexity and susceptibility to adversarial attacks where attackers manipulate data to evade detection.

This survey aims to provide an in-depth analysis of AI and ML techniques applied to DDoS attack detection. We explore various methodologies, benchmark datasets, evaluation metrics and real-world applications, shedding light on the advantages and limitations of each approach. Additionally, key challenges – including scalability, adversarial attacks, and interpretability – are analysed, along with future research directions aimed at improving AI-driven security frameworks. As the landscape of cybersecurity continues to evolve, integrating AI/ML into security infrastructures will be crucial in strengthening defences against increasingly sophisticated DDoS threats.

II. BACKGROUND AND DDoS ATTACK LANDSCAPE

As the internet continues to evolve, cyber threats have become increasingly sophisticated, with Distributed Denial of Service (DDoS) attacks emerging as one of the most disruptive security challenges. These attacks target organizations, cloud infrastructures, IoT networks and critical services, overwhelming them with malicious traffic to cause downtime, financial losses and reputational damage. DDoS attacks have evolved from simple volumetric floods to complex multi-vector strategies that exploit vulnerabilities in network protocols and application layers. Understanding the different types of DDoS attacks, their impact on various sectors and traditional detection mechanisms is crucial for developing advanced defence strategies. This section provides a detailed overview of the DDoS attack landscape, highlighting the classification of attacks, their consequences on cybersecurity and the limitations of conventional detection techniques.

2.1 Definition and Types of DDoS Attacks

A Distributed Denial of Service (DDoS) attack is a malicious attempt to disrupt the normal functioning of a targeted server, service, or network by overwhelming it with a flood of internet traffic. Attackers use multiple compromised devices, often forming botnets, to generate excessive requests and exhaust system resources.
DDoS attacks can be classified into several categories:

- **Volumetric Attacks**: These attacks, such as UDP floods, ICMP floods and DNS amplification attacks, aim to saturate the target's bandwidth by sending massive amounts of data packets.

- **Protocol Attacks**: These attacks, including SYN floods, Smurf attacks and Ping of Death, exploit vulnerabilities in network protocols to exhaust server resources.
- **Application-Layer Attacks**: Targeting web applications, these attacks, such as HTTP floods and Slowloris, mimic legitimate user behavior, making them harder to detect.
- **Multi-Vector Attacks**: A combination of different attack types, making detection and mitigation more challenging.

2.2 Impact on Cybersecurity

DDoS attacks have significant consequences, affecting various sectors:

- **Businesses and Enterprises**: Downtime caused by DDoS attacks leads to revenue loss, customer dissatisfaction and reputational damage.
- **Cloud Service Providers**: Disruptions in cloud environments can impact multiple clients, leading to severe operational disruptions.
- **IoT Networks**: As IoT devices often lack robust security, they are prime targets for botnet-based DDoS attacks, leading to widespread network failures.
- **Critical Infrastructure**: Government services, financial institutions and healthcare systems are at high risk, with attacks causing national security concerns.

2.3 Traditional Detection Methods

Traditional DDoS detection mechanisms include:

- **Signature-Based Detection**: Relies on known attack patterns and signatures. While effective for recognizing previously identified threats, it fails to detect novel or evolving attack strategies.
- **Anomaly-Based Detection**: Monitors network traffic for deviations from established baselines. While more adaptive, it can generate false positives due to legitimate traffic fluctuations.
- **Heuristic Techniques**: Utilize predefined rules to classify traffic. Though effective in some cases, these methods require constant updates and manual tuning to remain effective against sophisticated attacks.

As DDoS threats continue to evolve, traditional methods alone are insufficient, necessitating the integration of AI and ML for more robust, adaptive and real-time detection strategies.

III. AI/ML TECHNIQUES FOR DDoS DETECTION

The increasing frequency and sophistication of DDoS attacks demand advanced detection mechanisms capable of adapting to evolving attack patterns. Traditional rule-based approaches struggle to identify novel and stealthy attacks, necessitating the adoption of AI and ML techniques. AI/ML-based detection systems leverage vast amounts of network traffic data to identify anomalous behaviour indicative of an ongoing DDoS attack. These models can be trained using labelled data (supervised learning), detect patterns in unlabelled data (unsupervised learning), or extract complex features from raw traffic data (deep learning). The various AI/ML methodologies for DDoS detection, their implementation, strengths and limitations are as below.

3.1 Supervised Learning Approaches

Supervised learning techniques require labelled datasets where each instance is classified as either normal or malicious. These approaches are widely used in DDoS detection due to their ability to generalize well when sufficient training data is available.

- **Decision Trees & Random Forests**: Decision trees partition data based on feature thresholds to classify traffic as normal or malicious. Random Forests improve upon

this by aggregating multiple decision trees, reducing overfitting and enhancing robustness. These models provide interpretability and efficient classification, though they may struggle with high-dimensional data. Feature selection techniques play a crucial role in improving accuracy by identifying the most relevant network parameters, such as packet rate, source IP frequency and protocol types.

- **Support Vector Machines (SVM)**: SVMs are effective for both binary and multi-class classification. They construct hyperplanes in high-dimensional spaces to separate attack traffic from legitimate traffic. Kernel-based SVMs, such as radial basis function (RBF) and polynomial kernels, enhance classification in non-linear data distributions. However, SVMs require careful parameter tuning and suffer from scalability issues when applied to large-scale network datasets.
- **Neural Networks**: Traditional feedforward neural networks and Multi-Layer Perceptrons (MLPs) can classify traffic patterns using learned feature representations. These models excel in capturing complex relationships within traffic features but require substantial training data and computational power. Hybrid approaches combining neural networks with feature engineering techniques, such as Principal Component Analysis (PCA) and genetic algorithms, can enhance detection efficiency.

3.2 Unsupervised Learning Approaches

Unsupervised learning methods do not require labelled data, making them valuable for anomaly detection, where attack patterns may not be previously defined.

- **Clustering Methods (K-Means, DBSCAN):** Clustering techniques identify anomalies by grouping similar traffic patterns. K-Means partitions data points into clusters based on similarity, while DBSCAN detects dense regions of normal traffic, isolating outliers that could indicate potential DDoS attacks. These methods are useful for detecting novel attack patterns but may require careful parameter tuning. Distance-based clustering techniques, such as hierarchical clustering, have also been applied to improve anomaly detection accuracy.
- **Autoencoders**: These deep-learning-based models learn compressed representations of normal network traffic. When presented with anomalous traffic, autoencoders exhibit higher reconstruction errors, signalling potential DDoS attacks. Variational autoencoders (VAEs) introduce probabilistic encoding to improve generalization, making them effective in reducing false positives but requiring extensive training on clean traffic data. Self-taught learning and transfer learning techniques can be applied to enhance autoencoder performance on limited datasets.

3.3 Deep Learning Techniques

Deep learning has revolutionized DDoS detection by capturing intricate traffic patterns and temporal dependencies. These methods leverage large datasets and advanced architectures to enhance detection accuracy.

- **Convolutional Neural Networks (CNNs):** CNNs extract spatial patterns from network traffic features, making them effective in identifying attack signatures. They excel in feature extraction and pattern recognition but require substantial labeled training data. CNN architectures such as ResNet and MobileNet have been explored to enhance real-time detection capabilities. Feature augmentation techniques, including synthetic data generation and transfer learning, have also been applied to address data scarcity challenges.
- **Recurrent Neural Networks (RNNs) & Long Short-Term Memory (LSTM) Networks:** RNNs and LSTMs capture sequential dependencies in network traffic, making them well-suited for detecting DDoS attacks with evolving patterns over time. LSTMs address the vanishing gradient problem, improving long-term sequence

learning and anomaly detection. Bidirectional LSTMs (BiLSTMs) and attention-based mechanisms further enhance pattern recognition by considering past and future traffic states simultaneously. These models are particularly effective in identifying low-rate and slow-moving DDoS attacks.

- **Generative Adversarial Networks (GANs):** GANs consist of a generator and discriminator network, where the generator creates synthetic attack data and the discriminator differentiates between real and fake traffic. This technique enhances model robustness by generating diverse attack scenarios for training, improving generalization to real-world attack patterns. Conditional GANs (cGANs) have been explored for targeted attack simulations, while semi-supervised GANs integrate labeled and unlabeled data to refine DDoS detection performance. Adversarial training strategies further strengthen detection models against evolving attack techniques.

3.4 Hybrid AI/ML Approaches

To leverage the strengths of multiple AI/ML techniques, hybrid models combine different approaches to enhance detection accuracy and resilience against adversarial attacks.

- **Ensemble Learning**: Combining multiple models, such as Random Forests with CNNs or SVMs with autoencoders, helps mitigate individual model weaknesses and improves overall detection rates.
- **Hybrid Deep Learning Models**: Stacking CNNs with LSTMs enables both spatial and temporal feature extraction, making detection more robust against sophisticated attacks.
- **Feature Engineering & Dimensionality Reduction**: Applying techniques like PCA, t-SNE and mutual information maximization optimizes input data representation, improving detection performance with reduced computational overhead.

AI and ML techniques provide promising advancements in DDoS detection, each with unique strengths and challenges. While supervised learning offers high accuracy with labelled data, unsupervised learning excels in anomaly detection and deep learning methods enhance feature extraction and temporal analysis. These approaches can be optimized for real-time, scalable and adversarial-resistant DDoS mitigation systems. Additionally, integrating explainable AI (XAI) techniques can improve model interpretability and trustworthiness in critical cybersecurity applications.

IV. DATASETS AND BENCHMARKING

The effectiveness of AI/ML-based DDoS detection models largely depends on the quality and diversity of datasets used for training and evaluation. Datasets provide labelled or unlabelled network traffic data, including both benign and malicious activities, enabling researchers to develop, benchmark and refine detection techniques. Given the evolving nature of DDoS attacks, datasets must encompass various attack vectors, network environments and traffic patterns to ensure robust model generalization. Benchmarking these models requires standardized performance metrics that assess accuracy, detection capability and real-time efficiency. Some of the key publicly available datasets – both historical and recent – that have contributed to the advancement of DDoS detection and the essential metrics used to evaluate AI/ML models are discussed below.

4.1 Overview of Publicly Available Datasets

Publicly available datasets play a crucial role in training and evaluating AI/ML models for DDoS detection. These datasets contain network traffic information, including both normal and attack patterns, enabling researchers to develop robust detection models. Several datasets have been used in the literature that continue to serve as foundational benchmarks.

- **DARPA 1998 and 1999 Intrusion Detection Evaluation Dataset**: Created by the MIT Lincoln Laboratory, these datasets include simulated network attacks, including DDoS traffic. While somewhat outdated, they provide structured attack scenarios that aid in understanding traditional DDoS patterns.
- **CAIDA (Center for Applied Internet Data Analysis) Dataset**: CAIDA has provided numerous datasets since the early 2000s, capturing real-world DDoS attack traffic. The 2007 CAIDA dataset, for example, includes traces from actual network attacks and is widely used for anomaly detection research.
- **KDD Cup 1999 Dataset**: Originally designed for intrusion detection, this dataset contains various network attack types, including DDoS. Despite its limitations, such as redundant records, it remains a common benchmark for early-stage AI models.
- **ISCX 2012 Dataset**: Developed by the Canadian Institute for Cybersecurity, this dataset includes diverse attack scenarios, including different DDoS variants. It serves as a precursor to more advanced datasets like CIC-DDoS 2017.

These datasets provide detailed and updated attack patterns, offering improved benchmarking capabilities for AI/ML models. The availability of both older and modern datasets allows for comparative analysis of how attack patterns and detection techniques have evolved over time.

4.2 Performance Metrics for DDoS Detection

Evaluating AI/ML models for DDoS detection requires robust performance metrics to ensure accuracy, reliability and real-time responsiveness. The following key metrics are commonly used:

- **Accuracy**: Measures the proportion of correctly classified instances (both normal and attack traffic). While widely used, accuracy alone can be misleading when datasets are imbalanced.
- **Precision**: Defines the proportion of correctly identified attack instances among all instances classified as attacks. A high precision score indicates fewer false positives, reducing unnecessary mitigation efforts.
- **Recall (Detection Rate):** Measures the proportion of actual attack instances that are correctly identified. High recall ensures minimal false negatives, crucial for security applications.
- **F1-Score**: The harmonic mean of precision and recall, providing a balanced metric that accounts for both false positives and false negatives.
- **False Positive Rate (FPR):** Represents the percentage of normal traffic misclassified as an attack. A low FPR is essential for minimizing disruptions in legitimate network activities.
- **False Negative Rate (FNR):** Indicates the percentage of attacks misclassified as normal traffic. A lower FNR is crucial to avoid undetected threats.
- **Latency**: The time required for a model to detect an attack in real-time scenarios. Low latency is essential for rapid response mechanisms to mitigate ongoing threats.
- **Throughput**: Measures the number of traffic instances processed per unit time, impacting the scalability of the AI/ML model.
- **Area Under the Curve (AUC-ROC):** Provides a graphical representation of a model's performance across various threshold levels, highlighting its ability to distinguish between normal and attack traffic.

A combination of these metrics is typically used to ensure comprehensive evaluation, balancing detection effectiveness with practical deployment considerations. Models must

achieve high precision and recall while maintaining low false positives and low latency for optimal real-world implementation.

V. CHALLENGES AND FUTURE DIRECTIONS

Despite the promising advancements in AI and ML driven DDoS detection, several challenges hinder their widespread adoption and effectiveness in real-world scenarios.

- **Scalability Issues**: Handling large-scale, high-speed network traffic efficiently without compromising detection accuracy.
- **Evasion Techniques**: Attackers using adversarial AI, encryption and obfuscation techniques to bypass detection mechanisms.
- **Integration with Security Frameworks:** Seamless deployment in cloud, edge and hybrid environments while maintaining efficiency.
- **High False Positive Rates**: AI/ML models often misclassify benign traffic as malicious, leading to unnecessary mitigation actions.
- **Evolving Attack Patterns**: Newer DDoS attack variants, such as AI-generated and adaptive botnets, require continuous model updates.
- **Data Imbalance**: Many datasets contain more benign traffic than attack data, leading to biased models that struggle to detect rare or sophisticated attacks.
- **Computational Overhead**: Deep learning models require significant computational power, limiting their deployment in resource-constrained environments.
- **Real-Time Detection Constraints**: The need for ultra-low latency in DDoS detection to mitigate attacks before significant damage occurs.
- **Privacy and Data Security Concerns**: Training AI models requires large amounts of network data, raising issues related to user privacy and compliance with data protection regulations.
- **Cross-Network Generalization**: Many AI models perform well on specific datasets but struggle to generalize across different network environments.
- **Lack of Standardized Evaluation Frameworks**: The absence of universally accepted benchmarking standards makes it difficult to compare different detection methods.

Looking ahead, future research in AI/ML-based DDoS detection should focus on enhancing model interpretability to build trust among cybersecurity professionals. The development of adaptive and self-learning models that can detect emerging threats with minimal retraining will be crucial in staying ahead of evolving attack patterns. Additionally, improving dataset diversity and creating standardized evaluation frameworks will allow for better benchmarking and cross-network generalization. Advancements in edge AI and federated learning could also play a significant role in enabling efficient, privacy-preserving detection solutions in distributed environments. By addressing these challenges and leveraging cutting-edge AI innovations, the cybersecurity community can build more resilient and intelligent DDoS mitigation strategies for the future.

VI. Conclusion

This survey highlights the critical role of AI and ML in advancing DDoS attack detection, offering superior adaptability and efficiency compared to traditional rule-based approaches. AI driven solutions leverage advanced learning techniques to identify complex attack patterns, enabling real-time mitigation strategies to protect network infrastructure. The comparative analysis of supervised, unsupervised and deep learning models demonstrates their strengths in handling diverse attack scenarios while addressing evolving threats. Additionally, publicly available datasets and benchmarking methodologies provide valuable insights into model performance and effectiveness.

Despite these advancements, significant challenges remain in ensuring the robustness, scalability and real-world applicability of AI-based DDoS detection systems. High false positive rates, adversarial attacks and data privacy concerns present obstacles that must be addressed to enhance the reliability of these solutions. Furthermore, integrating AI models into existing cybersecurity frameworks while maintaining computational efficiency and low latency remains a key area for future research.

Hybrid models that combine multiple AI techniques, such as deep learning with heuristic-based methods, may offer improved accuracy and adaptability. The integration of federated learning and edge AI can also facilitate distributed and privacy-preserving DDoS detection solutions. By continuously evolving AI/ML methodologies and addressing existing limitations, researchers and cybersecurity professionals can develop more resilient, scalable and intelligent DDoS mitigation strategies, strengthening network security in an increasingly interconnected world.

References

[1]. S. Gunduz, B. Arslan and M. Demirci, "A Review of Machine Learning Solutions to Denial-of-Services Attacks in Wireless Sensor Networks," 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 2015, pp. 150-155, doi: 10.1109/ICMLA.2015.202.

[2]. O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 2017, pp. 2186-2193, doi: 10.1109/BigData.2017.8258167.

[3]. K. Kalkan, G. Gür and F. Alagöz, "SDNScore: A statistical defense mechanism against DDoS attacks in SDN environment," 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, 2017, pp. 669-675, doi: 10.1109/ISCC.2017.8024605.

[4]. Gulshan Kumar and Krishan Kumar. 2011. AI based supervised classifiers: an analysis for intrusion detection. In Proceedings of the International Conference on Advances in Computing and Artificial Intelligence (ACAI '11). Association for Computing Machinery, New York, NY, USA, 170–174.

[5]. Dr R Anurekha, "Understanding the Mirai Botnet: Architecture, Attack Mechanism and Defense Strategies", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.4, Issue 6, page no.839-843, June-2017.

[6]. P. Amaral, J. Dinis, P. Pinto, L. Bernardo, J. Tavares and H. S. Mamede, "Machine Learning in Software Defined Networks: Data collection and traffic classification," 2016 IEEE 24th International Conference on Network Protocols (ICNP), Singapore, 2016, pp. 1-5, doi: 10.1109/ICNP.2016.7785327.

[7]. P. Arun Raj Kumar, S. Selvakumar, “Distributed denial of service attack detection using an ensemble of neural classifier”, Computer Communications, Volume 34, Issue 11, 2011, Pages 1328-1341, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2011.01.012.

[8]. Dr R Anurekha. "A Conceptual Framework for Distributed Denial-of-Service (DDoS) Attacks: Taxonomy, Analytical Models and Defence Strategies", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Vol.4, Issue 2, pp.723-734, June 2016.

[9]. Hakem Beitollahi, Geert Deconinck, “Analyzing well-known countermeasures against distributed denial of service attacks”, Computer Communications, Volume 35, Issue 11, 2012, Pages 1312-1332, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2012.04.008.

[10]. Peng Xiao, Wenyu Qu, Heng Qi, Zhiyang Li, “Detecting DDoS attacks against data center with correlation analysis”, Computer Communications, Volume 67, 2015, Pages 66-74, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2015.06.012.

[11]. R. Das and T. H. Morris, "Machine Learning and Cyber Security," 2017 International Conference on Computer, Electrical & Communication Engineering (ICCECE), Kolkata, India, 2017, pp. 1-7, doi: 10.1109/ICCECE.2017.8526232.