IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

The Realm of IoT Security Infested with Botnets: A Comprehensive Survey to Research Proposed

[1]Pravin U. Chokakkar, [2] Rohit K. A. Suryawanshi, [3] Prof. Dr. Sunil B. Mane, [4]Prof. S. K. Gaikwad

[1]M.Tech Student, [2] M.Tech Student, [3]Associate Professor, [4]Assistant Professor

[1]Department of Computer Engineering and IT

[1]COEP Technological University, Pune, India

Abstract: This survey paper provides a comprehensive analysis of the information security practices within the realm of IoT security. It explores the challenges, strategies, and best practices for implementing security principles to enhance the security posture of IoT systems. By synthesizing existing literature and real-world case studies, this paper offers insights into the evolving landscape of in the context of IoT security.

Index Terms - IoT Security, Botnet, Machine Learning, Deep Learning, Information Security.

1. INTRODUCTION

In today's rapidly evolving digital environment, web applications have become an important part of the world, essential tools for communication, business, and information storage. As dependence on web applications increases, the importance of security cannot be ignored. Cyber threats continue to evolve, becoming more complex and diverse, making it important for organizations to adopt an effective security assessment.

1.1 Introduction to IoT security challenges

Based on threat intelligence gathered from 2.6 million smart homes worldwide shielded by NETGEAR Armor powered by Bitdefender, this report was produced. We looked at about 120 million IoT devices that produced an astounding 3.6 billion security events globally in order to find weaknesses and potential attack vectors and make everyone's smart home secure. [1] Change will be required due to privacy issues. Big data is essential to IoT devices. Less than 10,000 houses are estimated to be able to "generate 150 million discrete data points a day," or around one data point every six seconds for each household, according to a 2015 FTC study. Things are much worse now. [2] According to Deloitte's 2022 Connectivity and Mobile Trends Survey, one in two Internet of Things consumers voiced worries about security.

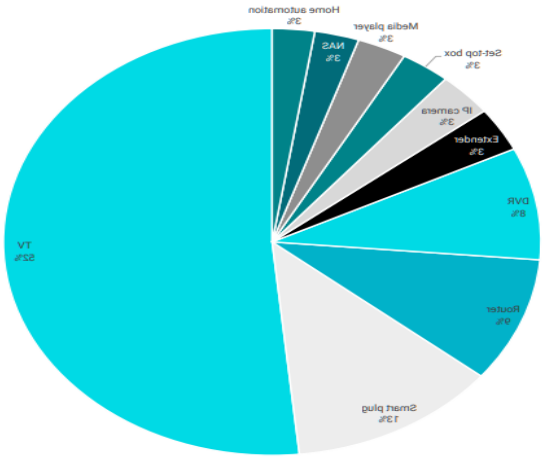

Figure 1. Survey Statistics

2. BACKGROUND

To understand the concepts of IoT Security and brief knowledge and ideas in the survey paper.

2.1 IoT Security

2.1.1 Large Scale Attacks

Almost every aspect of life has benefited greatly from IoT visibility, control, and opportunity. These days, the benefits of IoT may be observed in self-driving cars, issues with traffic congestion, patient monitoring, high-quality medical care, smart home products, etc. People's lives have improved, and productivity has grown because to these solutions. But because of their processing power and widespread connectivity, these gadgets are open to cyberattacks. Cybercriminals can take control of these gadgets and turn them into bots or zombies due to a lack of security safeguards. These devices join a botnet when hundreds of millions of them are infected. The command and control (C&C) server use these botnets to perform a variety of large-scale malicious assaults. There are several kinds of large-scale assaults. Some of these include maintaining an open HTTP connection on web servers and retransmitting TCP timeouts[3].

2.1.2 Intrusion Detection Taxonomy

Based on many factors, including detection strategy, detection location, and detection methodology, intrusion detection taxonomy classifies many kinds of intrusion detection methods and approaches [4]. The common categories in the intrusion detection taxonomy are broken down as follows:

(A) Signature-based Detection: These systems use signatures, or patterns of known attacks, to compare observed events. An alert is set off if a match is discovered. While signature-based detection works well against known assaults, it may have trouble identifying new or unidentified threats.

(B) Anomaly-based Detection: When observed behavior considerably departs from a baseline of typical activity, anomaly detection systems create an alarm. Although this method can identify assaults that have not been identified before, it may result in false positives and needs to be adjusted constantly to adjust to changing conditions.

(C) Hybrid Detection: By combining anomaly- and signature-based techniques, hybrid detection systems seek to maximize the benefits of each technique while lowering false positives and increasing detection accuracy.

Based on Detection Location:

(A) Network-based Intrusion Detection System (NIDS): NIDS analyzes packets and payloads to identify unusual activities including malware transmission, denial-of-service attacks, and port scans. It monitors network traffic in real-time.

(B) Host-based Intrusion Detection System (HIDS): HIDS is a system that monitors system logs, file integrity, registry changes, and other host-specific activities for indications of malicious activity or unauthorized access. It may be installed on individual hosts or endpoints.

(C) Wireless Intrusion Detection Systems (WIDS): WIDS are made expressly to keep an eye on wireless networks and identify potential security risks such as rogue devices and illegal access points.

2.1.3 IoT Threat Taxonomy

2.1.3.1 Botnet Attacks

Botnet assaults come in a variety of sizes and shapes, and each has its own goals and strategies[5]. The following are some prominent kinds of botnet attacks:

(A) Distributed Denial of Service (DDoS) Attacks: Botnet attacks which trigger distributed denial of service (DDoS) are among the most popular uses for them. A denial-of-service (DDoS) attack occurs when a botmaster directs the compromised devices to overload a target server or network with traffic, rendering it unusable for authorized users.

(B) Phishing and spamming: Phishing attempts and spam emails are commonly sent out in big quantities via botnets. Attackers can disseminate malicious emails more effectively and avoid being caught by spam filters by utilizing the processing power of several compromised machines.

(C) Credential Stuffing: Automated credential stuffing attacks, which include systematically testing stolen login and password combinations against a variety of websites and online services, can be carried out via botnets. Unauthorized access to user accounts may result from successful login attempts.

(D) Click Fraud: Botnets have been utilized in click fraud schemes to create phony clicks on web ads, inflating click-through rates and resulting in financial losses for advertisers. This may be used to influence internet advertising marketplaces, damage rival businesses, or make money through ad networks.

(E) Data Theft and Espionage: Sensitive information, including financial information, login passwords, intellectual property, and personal data, can be taken from compromised machines via botnets. It is possible to use this stolen data for espionage or gain financially.

(F) Crypto jacking: Certain botnets are used to mine cryptocurrencies without the owners' permission. The computing power of the compromised devices is used to mine cryptocurrency for the botmaster, such as Bitcoin or Monero.

2.1.3.2 Packet Flooding Attacks

A packet flooding assault, often referred to as a packet storm or flood, is a kind of Denial of Service (DoS) attack in which many packets are sent in rapid succession to a target network or server, overloading its resources and rendering it unavailable to authorized users [5]. When conducting a packet flooding assault, the attacker usually uses a variety of methods, including:

(A) UDP Flood: The attacker bombards the victim system with a huge quantity of User Datagram Protocol (UDP) packets sent to arbitrary ports. It is simpler to spoof the source IP address of the packets sent over UDP as it is connectionless and does not require a handshake, unlike TCP, which makes it more difficult to identify the attacker.

(B) ICMP Flood: The attacker bombards the target system with many Internets Control Message Protocol (ICMP) echo request (ping) packets. This kind of attack has the potential to overload the target's processing power and use up a large amount of network traffic.

(C) SYN Flood: To initiate a connection but prevent the target from completing the handshake, an attacker sends several TCP SYN packets to the target's TCP port. By packing the target's half-open connection table, this depletes its resources and stops authorized users from connecting.

(D) HTTP Flood: This attack targets the application layer of the OSI model by delivering a huge number of HTTP requests to a web server. It is sometimes referred to as a Layer 7 or application-layer flood. This might result in a denial-of-service attack by overloading the web server's processing power.

2.1.3.3 TCP SYN Flooding Attacks

A type of denial-of-service (DoS) attack known as TCP SYN flooding uses the TCP three-way handshake procedure to overload the resources of a target server and prevent authorized users from accessing it [6]. This is how it operates:

(A) Three-way handshake: When establishing a standard TCP connection, a client sends a server a SYN (synchronize) packet to start a conversation. A SYN-ACK (synchronize-acknowledgment) packet is returned by the server in response, signaling that it is prepared to establish a connection. The connection is finally formed when the client sends an ACK (acknowledgment) packet.

(B) SYN flooding: The attacker pretends to establish a connection while sending a high number of SYN packets to the target server. But the attacker either doesn't reply to the server's SYN-ACK packets or spoofs the originating IP addresses. Consequently, every time a SYN packet arrives, the server spends resources to create a half-open connection, but it never gets the last ACK needed to finish the three-way handshake.

(C) Resource exhaustion: The server uses up all its RAM and the number of half-open connections it can support while waiting for ACK packets that never show up. The server's resources become overloaded with half-open connections from the assault, eventually making it impossible for legitimate users to connect to the service.

2.1.3.4 Ping of Death

A Denial of Service (DoS) attack known as the "Ping of Death" occurs when an attacker takes advantage of flaws in network protocols to send large or improperly formatted ICMP packets to a target system. The target system may crash or become unresponsive due to resource depletion when it tries to handle these packets [7]. The target's services are not available to authorized users because of this attack. Patching vulnerabilities, limiting packet sizes on systems, and utilizing firewalls to stop excessive packets are examples of mitigation techniques.

2.1.3.5 Slow Loris Attacks

A particular kind of Denial of Service (DoS) assault known as the Slow Loris attack targets web servers by flooding them with connections and depleting their resources to stop them from providing services to authorized users. It sends incredibly sluggish HTTP queries to the server, sends incomplete requests all the time, and maintains open connections without finishing them. This causes a denial of service for genuine users attempting to contact the site by progressively using up the CPU, RAM, and connection slots that are available on the server.

3.REVIEW METHODOLOGY

3.1. Objectives and research questions

The purpose of this research project is to examine current and suggested defenses against widespread attacks on Internet of Things systems. To accomplish this goal, it is first necessary to have a thorough grasp of IoT systems, the many attacks that have recently occurred to cause widespread disruptions to IoT systems, and the varied defense strategies that researchers have come up with or implemented. Second, we will investigate various deep learning and machine learning methods that researchers employ to analyse network traffic and distinguish between harmful and benign traffic patterns.

Research questions	Question Objective	Objective
RQ1	What are the main security flaws that make IoT devices open to malicious attacks?	To understand various security needs, obstacles, and the necessity of defending internet-enabled devices from malicious, targeted, and widespread attacks.
RQ2	Which techniques are employed to safeguard Internet of Things systems?	To comprehend the many technical defenses against IoT system threats that researchers have offered at various network tiers.
RQ3	Which large-scale assaults affect Internet of Things devices?	To list well-researched and well-known widespread attacks that target Internet of Things devices.
RQ4	What kinds of machine learning and deep learning methods do researchers use?	To find out which deep learning and machine learning techniques are widely used and advised by researchers to defend IoT devices against advanced threats.
RQ5	What steps are being made to stop widespread attacks?	To identify the best defenses against large-scale attacks that the researchers proposed for IoT devices.
RQ6	How frequently is deep learning suggested as a defense against widespread IoT attacks?	To find out whether deep learning significantly exceeds all other machine learning techniques by analysing and contrasting various machine learning and deep learning solutions.

4. IDENTIFICATION OF RESEARCH

4.1 Search Strategy

In a research review, finding the most pertinent papers to the topic of interest is crucial. We looked up our subject on Google Scholar and IEEE Explore.

A. Google Scholar

B. ACM Digital Library

C. IEEE/IET Electronic Library (IEL) were used for the main search.

The databases advised for searching academic journal articles and conference papers include the IEEE/IET and ACM databases. The terms that search engines employ to find pertinent studies about large-scale assaults and IoT security in libraries.

Google Scholar has gained popularity as a tool for doing extensive searches for academic publications. However, a thorough examination was conducted to ensure that only journal-published publications and their complete texts were included in this study evaluation.

4.2 Research Selection

We searched all the derived references, and all known approached for a definite path of research selection. We got to learn the top-down approach followed by many authors to reach the smaller topic selection. They devised a way to narrow down our focus through each layer and reach a proper subset to the

field. The approach was from the flow of selection, IoT as the main subject with the security approach and decide to select network security as sub-topic. The domain has vast fields to explore, and attacks devised to it were present. The trend recently popularized had the Denial of Service (DOS) attacks in major numbers affecting the IoT devices in wide numbers. The DOS attacks had a major proportion of attacks made by the botnets with a defined hijacking techniques with the large-scale threat attacks made to disrupt the system.

5. Data Analysis

Data analysis for research involves several steps aimed at extracting insights, identifying patterns, and drawing conclusions from collected data.

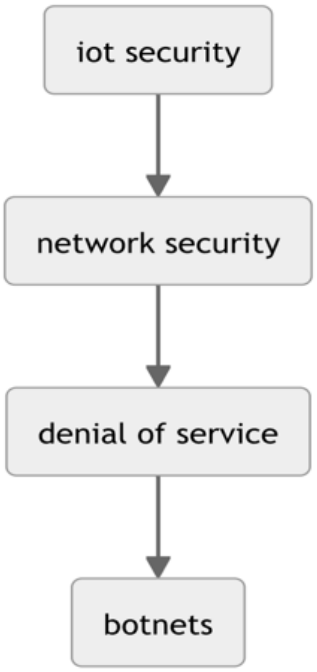

Figure 2. Research selection

5.1 Dataset Collection:

To simulate actual settings, a representative and diversified collection of IoT network traffic was assembled. The dynamic nature of botnet-driven assaults on IoT devices is reflected in this dataset, which contains both benign and dangerous actions. Because of privacy and security considerations, it might be difficult to get precise statistics for IoT botnets. For research reasons, you may find certain publicly accessible datasets helpful, nevertheless. Here are some locations for you to check out:

1. Internet of Things Dataset (IoT-23)-comprising 23 devices: This collection of network traffic statistics from 23 IoT devices is called IoT-23. Both favorable and detrimental traffic conditions are included. Although it does not focus directly on botnets, it can be useful in identifying both benign and potentially malicious activity in IoT traffic[8].

2.Bot-IoT- The dataset is a network traffic dataset for IoT botnet detection and cybersecurity research. Published in 2018 by the Stratosphere Lab research group at the Czech University of Technology in Prague, this dataset provides valuable insights into IoT botnet activity and network behavior. By using features extracted from network traffic packets and metadata about IoT devices, researchers can develop and evaluate intrusion detection systems suitable for the IoT environment [9].

3. CICIDS 2017 - Intrusion Detection Set from the Canadian Institute for Cybersecurity (2017):
This collection includes several IoT-related scenarios together with network traffic statistics. Although it does not only target Internet of Things botnets, it also offers a wide range of network traffic that may contain threats connected to IoT [10].

4. N-BaIoT – The dataset is a network traffic dataset created for research in the field of Internet of Things (IoT) security. Developed by the Australian Centre for Cyber Security (ACCS) at the University of New South Wales (UNSW), the dataset comprises traffic traces collected from various IoT devices such as surveillance cameras, smart TVs, and smart bulbs. With features extracted from network traffic and

metadata about device types, the N-BaIoT dataset enables researchers to study IoT device communication patterns, detect anomalies, and develop intrusion detection systems tailored to IoT environments[11].

5.2 Dataset Analysis

Dataset	Year Published	IoT Specific	Attacks Captured	Total features	Total benign records	Total Malicious Records
IoT-23	2020	Y	11	23	30,858,735	294,449,255
Bot-IoT	2018	Y	6	45	9,543	73,360,900
CICIDS	2017	Y	14	80	2,273,097	557,646
N-BaIoT	2018	Y	8	115	17,936	831,298

5.3. Machine Learning-Deep learning models in IoT security

Machine learning and deep learning along with enabling the development and training of models by learning from traffic patterns and offering an efficient response to IoT large-scale attacks, machine learning and deep learning may be extremely beneficial for developing an efficient network intrusion detection system (NIDS)[12]. Understanding machine learning and deep learning capabilities and approaches for safeguarding IoT systems against various large-scale threats is one of the paper's goals.. To identify large-scale attacks like DDoS, numerous research papers have been undertaken using a variety of machine learning and deep learning models, individually and as an ensemble of classifiers. Analysis of big data sets using machine learning (ML) may discover trends that are informative. The process of identifying intricate and challenging patterns in data by applying statistical methods and algorithms is known as machine learning (ML). Deep learning (DL) is one branch of machine learning that exhibits promise in detecting intrusions in Internet of Things networks. Deep learning models are developed to study how the human brain processes information. In order to generate an artificial neural network (ANN), multiple hidden layers are utilized.. Processional volume processing (ML) is not equally capable of processing massive amounts of data as deep learning (DL). The performance of ML models stabilizes when a certain threshold is met, whereas DL models continue to perform better as the quantity of data increases. Because of the enormous amount of data that Internet of Things devices generate, DL algorithms are consequently an ideal fit for these kinds of intrusion detection systems. There are various DL models; for example, Recurrent Neural Networks (RNN) are widely used for speech recognition and sequential data, whereas Convolutional Neural Networks (CNN) are specialized for image processing. Researchers have used these and a few more powerful models separately and in combination to determine the most effective defenses against widespread IoT threats. ML and DL techniques are mainly divided into the following groups:

1. Supervised learning:
Prior to model training, this kind of learning requires labelling the input and output data. This helps algorithms learn from the patterns in the incoming data to provide predictions or judgements. This method requires human data labelling according to correct classifications before models can be trained. An intrusion detection system can classify network traffic as malicious or benign. It might also classify it as a particular type of attack, such DDoS, Slow Loris, or TCP SYN packet.

2. Unsupervised learning:
Data that has not been labelled is provided as part of this learning process. Human involvement in the dataset creation process is usually negligible or nonexistent. By using this technique, models group data into classes according to the hidden structures in the dataset that they share. While supervised learning produces robust cybersecurity results, unsupervised learning is the better choice since network data is less constrained.

3.Reinforcement learning:
It's a method of learning through mistakes and applying it to the next decision-making in situations that are uncertain. It's a sort of game-like system where users are either awarded or punished depending on the algorithm's actions when the goals are not accomplished.

5.4. Evaluation metrics

Numerous measures are available in machine learning to assess the performance of the classifier. Selecting the appropriate performance indicators based on the real-world needs of a particular application is crucial. Certain classifiers could have strong performance in one metric but weak performance in another. Getting the most out of performance measurements is one aim of the model assessment process. We are restricting our attention to describing those performance measures taken from the list of accessible ML performance metrics and included in the research papers we examined.

5.4.1. Accuracy (ACC)

Accuracy is one of the most often used performance indicators for classifiers. It evaluates a classifier's ability to identify intrusions or assaults from an intrusion detection system's point of view. Put differently, it offers the proportion of accurately identified intrusion attempts to all inputs .

$$\text{Accuracy} = \frac{Correctly\ classified\ intrusion}{total\ number\ of\ inputs} \times 100\%$$

5.4.2. Precision (PR)

While accuracy is a useful indicator of a model's training efficiency, it is not the sole metric used in decision-making. An uneven dataset has an associated uncertainty factor. When the input varies, a model that returns a high accuracy score on the input dataset performs badly because it is unable to properly materialize the data [13]. Then, alternative performance metrics are used, such as precision (PR). It shows the percentage of positive cases projected to be positive. Stated differently, it refers to the percentage of malicious packets that are accurately notified. A greater precision score indicates that the attack data is being classified by the model with accuracy.It is calculated in this way:

$$\text{Precision} = \frac{Truepositives}{Truepositives + Falsepositives}$$

5.4.3. Recall (R)

The recall relates to sensitivity and our level of confidence that every favorable case was anticipated to be positive. Put otherwise, it offers a percentage of fraudulent packets that are accurately recognized. It is an additional crucial measure since fraudulent traffic may go undetected if a model is unable to identify widespread attacks. The security of the systems will be severely impacted by this. It is calculated in this way:

$$\text{Recall} = \frac{True\ positives}{True\ positives + False\ negative}$$

5..4.3 F-measure (F1)

It provides an overall accuracy score and an assessment of the model's performance by combining precision and recall. When a model has a high F-measure score, it has effectively detected attack traffic while reducing false positives and false negatives. The accuracy and recall harmonic mean is called the F-measure. It is calculated in this way:

$$\text{F} - \text{measure} = 2 \times \frac{precision \times Recall}{precision + Recall}$$

6. Results of Review

The goal of this research project is to identify the most reliable and efficient novel methods for protecting IoT systems against advanced attacks, as well as to understand the importance and need for safeguarding IoT systems from vulnerabilities. After reviewing each study and focusing on the research questions that were earlier laid out in the paper, the results are presented in the part that follows:

RQ1: What are the main security flaws that make IoT devices open to malicious attacks?

The goal of this inquiry was to better understand the various security requirements of the Internet of Things and the importance of protecting Internet-connected devices. Devices connected remotely can become targets of widespread, malicious attacks. IoT devices are open to numerous assaults. If you don't take strong security precautions, hackers might be able to infect your device and expand their botnet networks.

RQ2: Which techniques are employed to safeguard Internet of Things systems?

The goal of this research is to illustrate different technical strategies developed by researchers at different network tiers for safeguarding Internet of Things systems. Typically, a three-layered strategy is needed to safeguard a network. First, network traffic passes through filters by firewall rules to prevent harmful traffic from entering the network, creating a proactive layer. Second, intrusion detection systems (IDS) are used to implement the detective layer in order to detect potential network intrusions that have already evaded firewalls. As mentioned earlier, reactive security is the ability to strengthen defences before an attacker can exploit more vulnerabilities or to respond to an intrusion detection system (IDS) alarm that signals that the network has been compromised. Finally, the reactive or response layer is implemented to either quickly recover systems after an adversary was able to circumvent the other security layers.[14]

RQ3: Which large-scale assaults affect Internet of Things devices?

Identifying known and thoroughly researched large-scale attacks that affect IoT devices was the aim of this topic. With the development of technology, hackers are now aggressively searching for Internet of Things vulnerabilities including open ports, default passwords, and unencrypted network data using advanced and automated techniques. Malware is installed to take control of the device (bot) after a target has been found. [15] claimed that low-security setups make it simple to compromise a single IoT device. Millions of devices, however, may be compromised, giving rise to a powerful weapon that could harm the service. A group of compromised computers may be deployed as a botnet or an army of bots to carry out extensive attacks, depending on the attacker's objectives. A botnet attack can have disastrous effects. According to [24], there are two main architectures for botnets: P2P and Client-Server models. In a client-server architecture, all of the bots receive orders from a command and control (C&C) server.

RQ4: What kinds of machine learning and deep learning methods do researchers use?

The goal of this topic is to list the widely used and suggested machine learning and deep learning techniques by researchers for IoT system protection. A wide range of already developed and custom models have been used in numerous research studies to determine the optimal response to large-scale attacks originating from Internet of Things devices. The models we found from our data analysis that researchers use frequently and produce decent results for identifying malicious network traffic are listed in the shortlist below.

1.Support Vector Machine (SVM)

One of the most widely used and efficient classification methods that offers good accuracy while requiring less processing power is support vector machine (SVM). The SVM method divides the amount of features into a hyperplane or decision boundaries, and then it uses classification techniques to distinguish between distinct classes. Compared to neural networks, SVM performs faster and with more efficiency when the dataset size is smaller.

2. Random Forest (RF)

Based on ensemble approaches, bootstrapping, and bagging techniques, Random Forest (RF) is another well-known machine learning algorithm. Several individuals and distinct decision trees are trained in parallel using this technique, and the best results are found by aggregating (or "bagging") each individual's training outcomes. By dividing the features into small samples at random, random forest avoids feature association, which is one advantage over regular decision trees. Random Forest has been employed by researchers in a variety of domains, such as identifying anomalies and malicious network traffic in Internet of Things networks. [17] suggested a machine learning method that uses a variety of algorithms to detect DDoS attacks on IoT devices.

Deep Learning (DL)

Unstructured information is now prevalent everywhere in the big data and digital age of today, including social media, e-commerce, search engines, etc. A subfield of machine learning called deep learning simulates how the human brain processes and evaluates large datasets in order to make decisions. As dataset sizes increase, deep learning methods begin to outperform conventional machine-learning models in terms of performance. Deep learning's capacity to handle massive amounts of data and its ability to use Graphical Processing Units (GPUs) for parallel computing are the two key reasons it has grown in strength and shown encouraging results. These two characteristics both aid in enhancing training time.. [18] By using GPUs, which can execute iterative matrix multiplication and make use of thousands of processing units, deep neural networks with numerous layers increase the accuracy of deep learning models. We discussed several studies in the data analysis part that achieved above 99% accuracy by combining the non-linear

sigmoid and ReLU activation functions with artificial neural networks (ANN). ANN outperformed ML models in every study when large-scale assaults were identified using both ANN and conventional ML models.

Convolution Neural Network (CNN)

Convolution Neural networks classify input data, primarily images, into many groups using one or more convolutional and subsampling layers. The three stacked layers that comprise CNN's basic architecture are the convolutional layer, pooling layer, and fully connected layer. CNNs are well known for their superior accuracy in solving challenging tasks. CNN claims that great precision requires a lot of computing power, while the Internet of Things operates in a resource-constrained context. [19] In their study on effective, accurate CNNs for Internet of Things devices, it is critical to find a balanced CNN model that works effectively and gives the best accuracy with the least amount of processing overhead. This means that either the CNN model's size must be decreased to fit an IoT device, which requires skill in addition to trial and error.

Recurrent Neural Network (RNN)

The recurrent neural network (RNN) is another powerful deep neural network that is widely used to process and identify patterns in time series, natural language, and sequential data. RNNs are an excellent choice for assessing IoT network traffic for anomaly detection since sequential data includes information like stock price, timestamps, and network traffic, among other things. [20] notes that because RNNs can analyze and learn from network traffic to identify unusual traffic packets, they can be a powerful tool against bot and fraud detection. In a similar vein, RNNs can also successfully identify hostile actors by detecting user activity. Long Short-Term Memory is the name of one kind of specialized RNN (LSTM).

Autoencoders (AE)

Using autoencoders (AE) is one of the best unsupervised neural network techniques. An autoencoder consists of the input layer, the hidden encoding layer (sometimes referred to as the bottleneck), and the decoded output layer. The input and output layers have the same size. An input's size and dimensions are reduced once it is transmitted to an AE, encoded, and compressed. The compressed data is fed onto the output layer, which then restores it to its original dimensions.

RQ5: What steps are being made to stop widespread attacks?

Listing the various defenses that researchers have developed to protect IoT systems from mass attacks was the aim of this article. The problem of protecting IoT devices from widespread attacks is made more difficult by their limited computing and energy resources. If one IoT device is compromised, it is not a big deal. Yet, adversaries might use millions of these infiltrated IoT devices as bots to launch a denial-of-service attack (DDoS) on critical infrastructure, potentially causing major damage to household, commercial, medical, and transportation equipment. It is difficult for manufacturers to put security first because of the fierce rivalry for new, inexpensive IoT devices and their rapid yearly growth. [21]

RQ6: How frequently is deep learning suggested as a defense against widespread IoT attacks?

The aim of the investigation was to compare different machine learning and deep learning strategies in order to ascertain whether deep learning is superior to all other methods. Traditional machine learning methods struggle to understand the dynamic nature of DDoS attacks. This is a result of the heavy reliance on human network traffic monitoring in complicated feature extraction techniques employed by machine learning algorithms [22].

7. Challenges and Future Directions

7.1 IoT Security

7.1.1 Selection of the right IDS approach

Even after their initial detection in 2016, DDoS assaults such as Mirai continue to pose a significant risk. Mirai-like versions have been found as of July 2019 by expanding the attack surface and utilising various payloads. In order to protect susceptible devices, it is essential to detect botnets and use a multi-layered defence strategy.

7.1.2 Scalable IDS solutions for IoT

Using deep learning algorithms to scan network packets provides a more scalable method of detecting zero-day attacks in the Internet of Things. Cyberattacks are becoming more complex, advanced, and frequent these days. By examining the available data and analysing user behaviour, deep learning models are able to self-learn from previous attacks and identify hidden patterns in the data that can be used to identify malicious attempts [23] .

7.1.3 Selection of proper dataset

The distinctive characteristics of the Internet of Things ecosystem necessitate meticulous model construction and training for intrusion detection. Comparably, different solutions are needed for IoT device network traffic and resource constraints (reduced memory, processing power, energy, etc.) than for traditional computer systems. It's crucial to train machine learning and deep learning models on an IoT-specific dataset that includes a variety of assaults.

7.1.4. Continuous training and labeling of dataset

Training a model using a benchmark dataset is necessary to create the best possible ML and DL model. When datasets are big, scientists use a smaller dataset to train their machine learning models to maximize performance. If the model in the training phase of an IDS does not monitor all patterns of network traffic, this can occasionally result in the creation of a biased model. Although such models would initially yield better performance measures, their application in a real-world setting would be unsuccessful due to a lack of generalization of previously unseen patterns.

8. FUTURE WORK

Proposing an enhanced model with higher rate of accuracy and precisions as to detect the comparative study.

9. CONCLUSION

The scope of the research has been widely looked upon and narrowed on the prospects of machine learning topics and the future work on the dataset analysis using Artificial Intelligence has been set.

10. REFERENCES

1.2023-IoT-Security-Landscape-Report.pdf (bitdefender.com)
2.https://www2.deloitte.com/content/dam/insights/articles/us175371_tmt_connectivity-and-mobile-trends-interactive-landing-page/DI_Connectivity-mobile-trends-2022.pdf
3. Osterweil, Eric, Angelos Stavrou, and Lixia Zhang. "20 years of DDoS: A call to action." *arXiv preprint arXiv:1904.02739*(2019).
4. Anthi, Eirini, et al. "A supervised intrusion detection system for smart home IoT devices." *IEEE Internet of Things Journal* 6.5 (2019): 9042-9053.
5. Kelly, Christopher, et al. "Testing and hardening IoT devices against the Mirai botnet." *2020 International conference on cyber security and protection of digital services (cyber security)*. IEEE, 2020.
6. Haris, S. H. C., et al. "TCP SYN flood detection based on payload analysis." *2010 IEEE Student Conference on Research and Development (SCOReD)*. IEEE, 2010.
7. Yusof, Mohd Azahari Mohd, Fakariah Hani Mohd Ali, and Mohamad Yusof Darus. "Detection and defense algorithms of different types of DDoS attacks." *International Journal of Engineering and Technology* 9.5 (2017): 410.
8. IoT-23 Dataset: A labeled dataset of Malware and benign IoT traffic. (n.d.). Stratosphere IPS. Retrieved October 22, 2020, from https://www.stratosphereips.org/datasets-iot23
9. Yusof, Mohd Azahari Mohd, Fakariah Hani Mohd Ali, and Mohamad Yusof Darus. "Detection and defense algorithms of different types of DDoS attacks." *International Journal of Engineering and Technology* 9.5 (2017): 410.
10. Chaabouni, Nadia, et al. "Network intrusion detection for IoT security based on learning techniques." *IEEE Communications Surveys & Tutorials* 21.3 (2019): 2671-2701.
11. Alsamiri, Jadel, and Khalid Alsubhi. "Internet of things cyber attacks detection using machine learning." *International Journal of Advanced Computer Science and Applications*10.12 (2019)..
12. Chaabouni, Nadia, et al. "Network intrusion detection for IoT security based on learning techniques." *IEEE Communications Surveys & Tutorials* 21.3 (2019): 2671-2701..
13. Rawat, S. "Is accuracy EVERYTHING?." *Medium* (2019).

14. Graves, J. "Reactive vs. proactive cybersecurity: 5 reasons why traditional security no longer works." (2019)..
15. http://M. Pratt, Learn the IoT botnets basics every IT expert should know, IoT Agenda (2020) https://internetofthingsagenda.techtarget.com/feature/ Learn- the- IoT- botnets- basics- every- IT- expert-should- know .
16. Goyal, Mohit, Ipsit Sahoo, and G. Geethakumari. "HTTP botnet detection in IOT devices using network traffic analysis." *2019 International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC)*. IEEE, 2019.
17. Chaudhary, Pooja, and Brij B. Gupta. "Ddos detection framework in resource constrained internet of things domain." *2019 IEEE 8th global conference on consumer electronics (GCCE)*. IEEE, 2019..
18. Yeung, Gingfung, et al. "Towards {GPU} utilization prediction for cloud deep learning." *12th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 20)*. 2020.
19. Lawrence, Tom, and Li Zhang. "IoTNet: An efficient and accurate convolutional neural network for IoT devices." *Sensors* 19.24 (2019): 5541..
20. http://] Volodymyr, B. (2020). Recurrent neural networks appications guide [8 Real-Life RNN Applications]. https://theappsolutions.com/blog/development/ recurrent- neural- networks/.
21. http://N. McKinley, Challenges in Software Security for IoT Devices (and How to Tackle Them) March 2, Heimdal Security Blog, 2020 https://heimdalsecurity. com/blog/challenges-security-for-iot/.
22. DeBeck, C., J. Chung, and D. McMillen. "I can't believe mirais: tracking the infa-mous IoT malware." (2019).
23. Muncaster, Phil. "Cyber-attacks up 37% over past month as# COVID19 bites." *Infosecurity Magazine. Retrieved* 25 (2020).