Contents lists available at ScienceDirect

International Journal of Cognitive Computing in Engineering

journal homepage: www.keaipublishing.com/en/journals/international-journal-of-cognitive-computing-in-engineering/

Adversarial learning for Mirai botnet detection based on long short-term memory and XGBoost

Vajratiya Vajrobol [a,b], Brij B. Gupta [c,f,g,h,*], Akshat Gaurav [b,d], Huan-Ming Chuang [e]

[a] *Institute of Informatics and Communication, University of Delhi, India*
[b] *International Center for AI and Cyber Security Research and Innovations. Asia University, Taiwan*
[c] *Department of Computer Science and Information Engineering, Asia University, Taichung 413, Taiwan*
[d] *Ronin Institute, Montclair, NJ, USA*
[e] *Department of Information Management, National Yunlin University of Science and Technology, Yunlin, Taiwan*
[f] *Symbiosis Centre for Information Technology (SCIT), Symbiosis International University, Pune, India*
[g] *Department of Electrical and Computer Engineering, Lebanese American University, Beirut 1102, Lebanon*
[h] *Center for Interdisciplinary Research, University of Petroleum and Energy Studies (UPES), Dehradun, India*

ARTICLE INFO

Keywords:
Mirai botnet cyber attack
Long short term Memorys XGboost
Adversarial training

ABSTRACT

In today's world, where digital threats are on the rise, one particularly concerning threat is the Mirai botnet. This malware is designed to infect and command a collection of Internet of Things (IoT) devices. The use of Mirai attacks has intensified in recent times, thus threatening the smooth operation of numerous devices that are connected to a network. Such attacks carry adverse consequences that include interference with services or the leakage of confidential information. To fight this growing threat, smart and flexible detection techniques are required to counter the new methods cyber attackers use. The aim of this research is to develop a resilient defense against Mirai botnet attacks. The Long Short Term Memory term (LSTM) and XGBoost combined have the best performance of 97.7% accuracy score. With this combination, the aim is to strengthen our cyber defenses against sophisticated and dynamically operating Mirai botnets to further enhance the security of our digital world.

1. Introduction

IoT's efficiency and convenience have revolutionized technology connections and interactions. But growing secu- rity threats (Omolara et al., 2022; Singh and Gupta, 2022; Liu et al., 2022), particularly botnets, jeopardize the integrity of IoT ecosystems. The environment also faces dangerous opponents such as the Mirai botnet, which is infamous for facilitating the coordination of large distributed denial-of- service (DDoS) attacks that can disrupt critical infrastruc- ture and services (Hallman et al., 2017; Wang et al., 2022).

Our study seeks to offer an elaborate and advanced insight into the activities of the Mirai botnet and other associated botnets in response to this rising menace. In this paper, five advanced detection models are presented to address the challenges involved in detecting Mirai bot- nets in an Internet of Things (IoT) context. The models include Deep Neural Network (DNN), Convolutional Neu- ral Network (CNN), Long Short-Term Memory (LSTM), RandomForest-Combined LSTM, and XGBoost-Integrated LSTM. These models act as a toolkit that enhances detection accuracy by exploiting different architectural strengths.

In contrast with more traditional machine learning and deep learning techniques (Almomani et al., 2022; Hu et al., 2022; Tembhurne et al., 2022; Yousaf et al., 2021), our research takes into consideration the dynamic nature of cyber threats. We understand that hostile attacks on the security systems of the IoT are becoming more complex. In this work, we focus on combining adversarial training approaches to improve our models' resilience to such adver- sarial challenges. We use adversarial perturbations for our model training to make models that are more robust to any malicious attempts.

The implications of our findings are not confined to the Mirai botnet detection domain and are relevant for adver- sarial learning in IoT security as a whole. As adversarial attacks become sophisticated and pervasive, our work lays the foundation for developing stronger and more versatile security approaches within the IoT landscape.

Our contributions are presented as follows;

[*] Corresponding author.
E-mail addresses: bbgupta@asia.edu.tw (B.B. Gupta), chuanghm@yuntech.edu.tw (H.-M. Chuang).

https://doi.org/10.1016/j.ijcce.2024.02.004
Received 21 December 2023; Received in revised form 29 February 2024; Accepted 29 February 2024
Available online 3 March 2024

- Integrating hybrid deep neural networks based on Mi- rai botnets such as RandomForest-Combined LSTM and XGBoost-Integrated LSTM with adversarial train- ing practice.
- A comparative study of five algorithms for detecting Mirai botnets such as Deep Neural Network (DNN), Convolutional Neural Network (CNN), Long Short- Term Memory (LSTM), RandomForest-Combined LSTM, and XGBoost-Integrated LSTM.
- Enhanced the protection by improving the accuracy of Mirai botnet detection.

The research structure starts with an introduction, fol- lowed by the presentation of the dataset and methods. In the fourth section, the results and discussion are elaborated, and finally, the conclusion is drawn.

2. Literature surveys

Prior studies on the Mirai botnet have constituted a cru- cial foundation for understanding its behaviour and devising effective detection strategies. Mirai Botnet, which was ini- tially discovered in 2016, became popular by penetrating IoT devices and using them to cause mass destruction through distributed denial-of-service (DDoS) attacks. Much has been written on Mirai with regard to its propagation techniques, attack methods, and possible remedies (Kambourakis et al., 2017).

McDermott et al. (2018) discusses the spike in dis- tributed denial of service (DDoS) attacks that are IoT-related and suggests a deep learning solution, more specifically a bidirectional long short-term memory-based recurrent neural network (BLSTM-RNN). The model uses word em- bedding for text recognition, converting attack packets into tokenized integer format. In this study, a traditional LSTM- RNN was considered, and the results were compared to BLSTM-RNN in order to detect four attack vectors that can be used by Mirai Botnet, such as accuracy and loss. Though dynamic programming has the highest processing time, the research shows that it is the most efficient model in practice (McDermott et al., 2018).

Another architecture was established by Ahmed et al. (2019), which uses blockchain for better IoT security against Mirai Botnet attacks. The approach is based on divid- ing the network into autonomous systems (AS) and using blockchains for the storage and distribution of lists with internet protocol (IP) addresses of connected hosts, marking as malicious (Ahmed et al., 2019).

Mezher et al. (2022) proposed a model for adversarial attacks in images. Deng and Karam (2022) suggest tech- niques for the detection of adversarial attacks on texture recognition. Mishra et al. (2021) proposed DDoS attack in SDN-cloud network. Kiran et al. (2022) suggest an encryption scheme for cloud-assisted IoT. Kumar et al. (2022) propsoed key management schem for cloud-based vehicular IoT networks.

Wahab et al. (2017) proposed DDoS attack detection in a cloud environment. Shaikh et al. (2022) proposed bonnet detection in VANETs. Abbas et al. (2021a) proposed feature selection technique for attack detection in SDN.Al-Qerem et al. (2020b) proposed technique for IoT transaction process in fog environment. Gupta and Quamara (2020); Khanam et al. (2022); Sadatacharapandi and Padmavathi (2022); Sharma and Sharma (2022) presents a survey on IoT security challenges. Memos et al. (2018) proposed a surveillance system for IoT in smart cities. Mishra et al. (2022) proposed DDoS attack detection using a majority vote system.

Al-Qerem et al. (2020a) used the combination of ma- chine learning and feature selection algorithms with Multi- phase Genetic Algorithm and Particle Swarm Optimization (PSO) to determine the best set of features. The Random Forest algorithm is used through this Feature Selection Algorithm, achieving a highly impressive one 100%

Furthermore. Recent research from Nakip and Gelenbe (2021) introduced an attack detection scheme based on Auto-Associative Dense Random Neural Network (AADR) with high accuracy in the detection of attacks and fewer false alarms .

GÜVEN et al. (2023) presented a methodology for analyzing con- strained IoT devices and focuses on simulat- ing Mirai attacks. The research explores the relationship between various sample sizes from the Kitsune dataset, aiming to optimize accuracy in Mirai attack detection. To achieve this, the study evaluates different approaches with smaller sample sizes to reduce training time on resource- constrained devices. Cross-validation techniques are em- ployed to prevent overfitting during the learning process, and Bootstrapping generates samples of varying sizes (1000, 10000, and 100000) to assess performance metrics. The findings indicate that a sample size of 10000 achieves an impressive accuracy rate of 99.56% for Mirai attack detection in IoT devices .

Sharma et al. (2023) focused on subtypes of the Mirai attack, including ACK, SYN, Plain UDP, UDP flood, and Scan, to assess their impact on IoT systems. The research employs a dataset and evaluates its performance using metrics like accuracy, precision, recall, and F1-score. The results indicate that the CNN model outperforms LSTM and GRU in terms of these metrics when applied to the NBaIoT dataset.

Affinito et al. (2023) conducted a comprehensive inves- tigation into the evolution of the Mirai botnet spanning a six-year period, from 2016 to 2022. The analysis primarily focuses on TCP SYN packets that exhibit the Mirai signa- ture, characterized by a TCP sequence number equal to the destination IP address.

The study’s noteworthy findings indicate that the Mirai signature continues to be utilized by malicious actors today, contrary to previous assessments. Furthermore, the research reveals an upward trend in the number of compromised devices and the volume of TCP SYN packets involved in scanning activities over time. Notably, the study highlights that cybercriminals predominantly target Telnet port 23, with fewer requests made on Telnet port 2323. Conversely, there is a decrease in the number of probes on SSH ports, followed by a resurgence in 2022. Finally, the research identifies several ports that had not been targeted until 2018 but have since experienced a substantial influx of TCP SYN packets exhibiting the Mirai signature. These previously uncontacted ports are linked to the emergence of new variants of the Mirai botnet, underscoring the evolving nature of this cybersecurity threat (Affinito et al., 2023).

Focusing on applying machines and deep learning to de- tect the pattern of Mirai botnets. The inclusion of adversarial learning strengthens the efficacy of cyber attack defence. In the adversarial phase, the model is trained by exposing it to intentionally adversarial examples designed to emulate variances sophisticated attackers may introduce. This assists the model generalised well and become resistance in volatile to cyber threats, such as those constituted by Mirai botnets (Mustapha et al., 2023; Novaes et al., 2021; Huang et al., 2020)

In a nutshell, our investigation highlights the need to combine adversarial learning with existing detection meth- ods to strengthen the resistance against Mirai botnets. The model pays particular attention to this fusion due to the fact that it helps in bridging a major shortcoming existing in other approaches, thereby improving the overall robustness of our study’s proposed detection mechanisms.

3. Methods

In this study, we leveraged the Mirai botnet dataset ex- tracted from the extensive CICIoT2023 dataset as illustrated in Fig. 1, purposefully crafted for the high-throughput detection of infiltrations and direct DDoS attacks in IoT- based ecosystems. This dataset encompasses a spectrum of attack labels, encompassing GREIP, GREETH, UDP Plain, alongside benign traffic. To ensure the dataset’s readiness for analysis, we meticulously implemented preprocessing techniques, including label encoding and min-max scaling. Subsequently, we segregated the data into distinct training (80%) and testing (20%) subsets. Our research harnessed the power of five distinct and versatile machine learning algorithms: Deep Neural Network (DNN), Convolutional Neu- ral Network (CNN), Long Short-Term Memory (LSTM), LSTM combined synergistically with Random Forest, and LSTM strategically integrated with

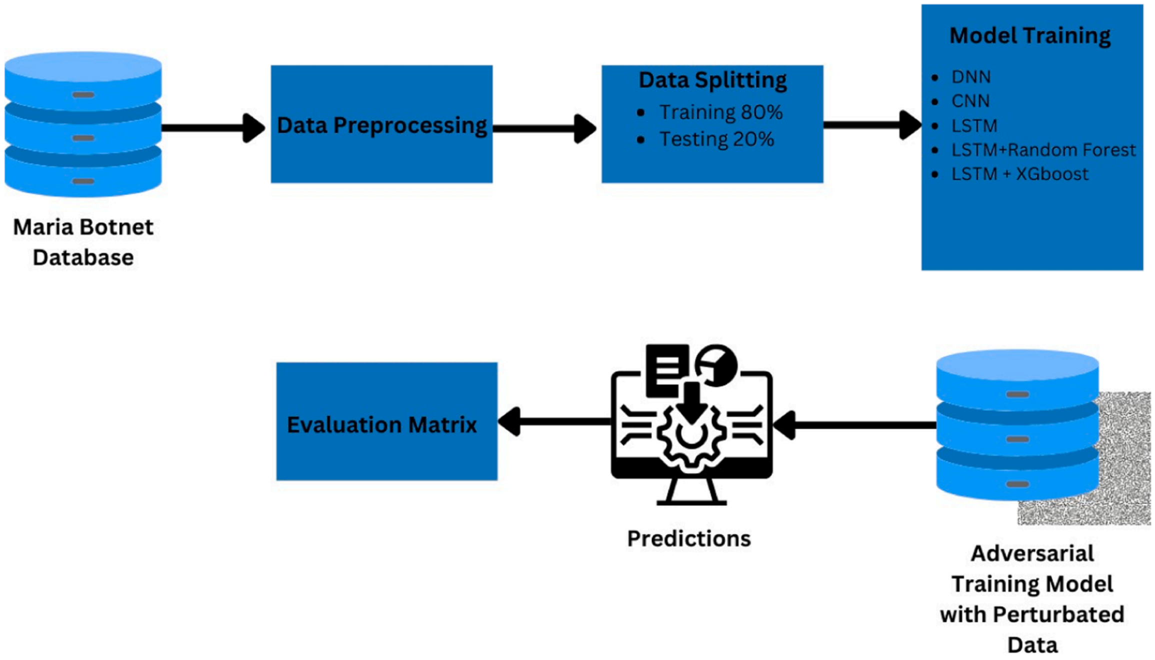

Fig. 1. Mirai botnet detection Framework using adversarial training method.

XGBoost for training and predictive purposes. To thoroughly assess the models' efficacy, we employed a comprehensive array of evaluation metrics, encompassing accuracy, precision, recall, F1-score, and the intricate analysis provided by the confusion matrix. In addition to this structured approach, we conducted adver- sarial training, fortifying our models against sophisticated attacks. This comprehensive methodology facilitates an in- depth exploration of the Mirai botnet dataset within IoT en- vironments, advancing our understanding of security threats and illuminating potential countermeasures.

4. Dataset and data pre-processing

The dataset was retrieved from CICIoT2023, which is a real-time dataset for big data-based high-throughput detec- tion of infiltrations and direct DDoS attacks in the IoT-based environment (Neto et al., 2023).

In this dataset, there are multiple forms of Mirai attacks, notably massive Distributed Denial of Service (DDoS) at- tacks against IoT gadgets. The data collection included five different Raspberry Pis, representing five different variants of the Mirai attack. A Netgear unmanaged switch is used to provide connectivity between the attackers and general IoT devices. Attacks are implemented by a set of tools, and customized Mirai configurations are used for experiments.

The role of an online IoT supervisor is that of a coordina- tor, managing the operation of multiple IoT devices in the topology, including sensors, cameras, and smart speakers.

Specifically examined attack methodologies include GREIP and GREETH. In the GREIP attack, encapsulated packets with random IPs and ports for internal data and real IPs in the external layer are flooded into the target system. Like the GREETH attack, while morphing a GRE IP tunnel, it enters the comparable process but underscores a unique packet encapsulation method according to the ethernet header. It also pays attention to UDP Plain, which points out that the victim systems to be attacked should be UDP packets. This type of attack uses a duplicated packet segment, and more specifically, every packet has a different payload. The variety of attack strategies added to a robust understanding of poten- tial threats to IoT devices under real-world circumstances.

Additionally, adding benign labeling to make a distinc- tion between them and Mirai attacks. This dataset has 22,115 Mirai-great floods, 24,476 benign traffic, 16,952 Mirai-grip floods, and 20,166 records of Mirai-up plain, as shown in Fig. 2. The data preprocessing steps, including label encoding and min-max scaling, have been demonstrated to ensure uniformity in the data range before proceeding to the 80:20-train-test split. The training set is made up of 66,967 records, and the testing set has 16,742 records, making the dataset size 83709 traffic records.

5. Botnet

In this complex process, a botnet is built using a multi- staged strategy that includes a scanner server, a loader server, and a C2 (Common and control) server that cybercriminals often employ. Fig. 3 shows the scanner server's initial process, which entails actively seeking out vulnerable de- vices on the internet by probing open ports, discovering unpatched software, and exploiting known vulnerabilities. The loader server is informed strictly by the scanner server regarding the presence of devices containing vulnerabilities. The purpose of the loader server is to inject the target devices with a malicious payload (such as

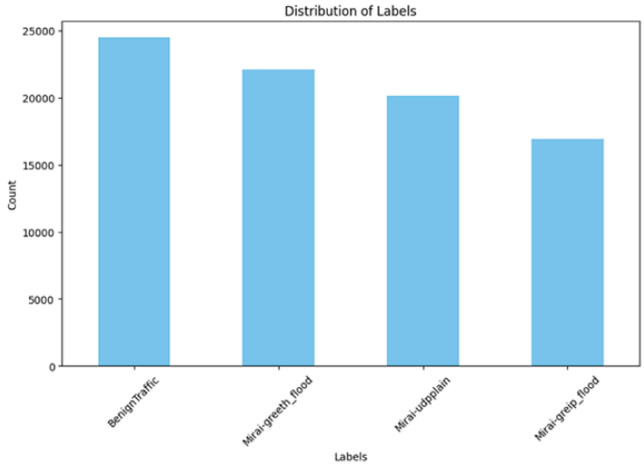

Fig. 2. Distribution of dataset

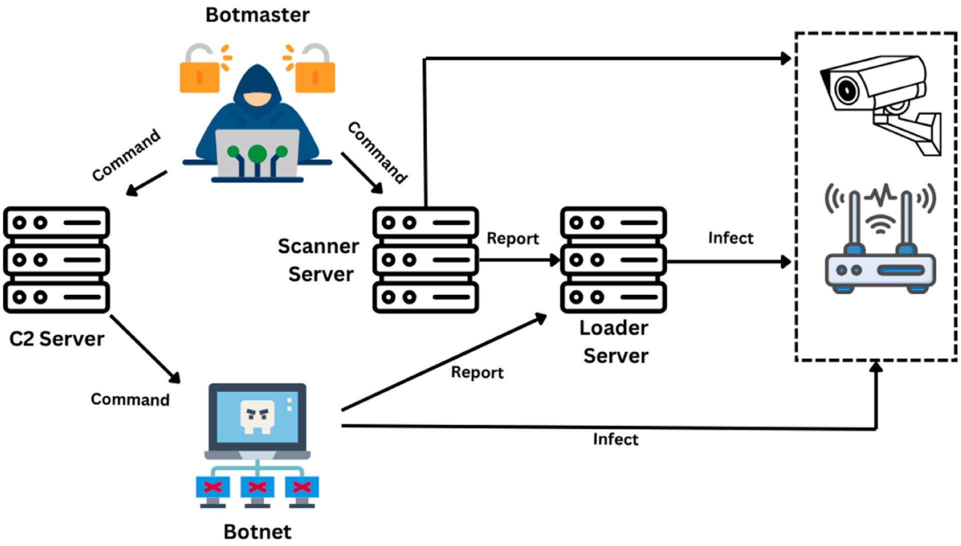

Fig. 3. Botnet Working

malware or a trojan), which makes them vulnerable at those points and gives unauthorized access. The compromised devices, in turn, try to connect to the C2 server that acts as its central command center. The botnet operator, or the botmaster, issues commands to the infected devices through the C2 server and orders them to undertake various evil operations. The loading, commanding, and scanning processes in this sequence enable the creation and control of a botnet in an organized manner (Abbas et al., 2021b).

Next section, five algorithms for Mirai botnet detection frameworks will be illustrated.

6. Deep neural network (DNN)

Fig. 4. shows that DNN layers in the architecture are closely connected. The first layer is the dense layer with 128 units and rectified linear unit (ReLU) activation function. The size of the features in the scaled training data is such that this layer acts as the first feature extractor, transforming the input data. The one-dimensional convolutional layer with 64 filters and ReLU activation is then followed by the dense layer with 64 units and ReLU activation. The final layer is also a dense layer; it uses a softmax activation function for class probabilities. This informs the use of the sparse cate- gorical cross-entropy loss function, which is advantageous when dealing with multi-class classification, and as such, it is constructed with an Adam optimizer. In particular, during training, the model changes its parameters iteratively so as to reduce the provided loss. In tasks where feature hierarchies and relationships between input features and target classes need to be captured, this DNN architecture can be utilized (Kudugunta and Ferrara, 2018).

7. Convolutional neural network (CNN)

The design includes a Conv1D layer with 64 filters of kernel size three and the ReLU activation function in the CNN. This implies that a MaxPooling1D layer with a pool size of 2 follows the convolutional layer in Fig. 5. and reduces the spatial dimensions of the output feature maps in an effort to cut down on redundancy and computational expense. To make it a one-dimensional vector, the multi- dimensional output of this process is run through a flattening layer. The last architecture is dense (Alkahtani and Aldhyani, 2021).

8. Long short term memory (LSTM)

In the LSTM model, a 50-unit Long Short-Term Memory (LSTM) layer is integrated into the architecture for capturing temporal dependencies within sequential data. The shape of the input is defined by the dimensions of the reshaped training data. Next, a dense layer with an activation function of softmax is introduced after the LSTM layer that provides a probability distribution over the classes by having as many units as all the classes in the classification task. The model was created using the sparse categorical cross-entropy loss function and the Adam optimizer for multi-class classifica- tion problems. In training, the model modifies its parameters to minimise the given loss and evaluates measures like accuracy (Alkahtani and Aldhyani, 2021).

9. LSTM+Random forest

A fusion of LSTM (long short-term memory) networks together with conventional machine learning models such as RandomForest will create a hybrid architecture to take advantage of both. Given their

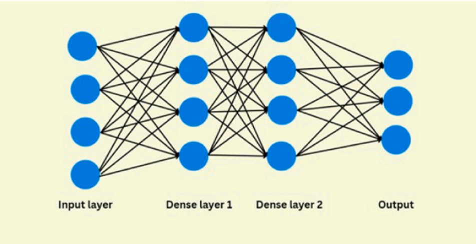

Fig. 4. DNN Architecture

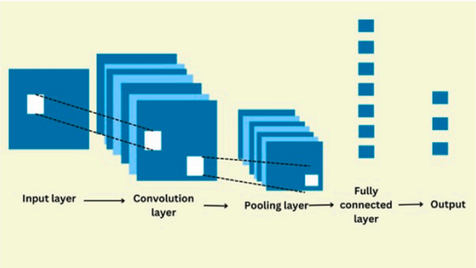

Fig. 5. The architecture of CNN

ability to capture sequential dependencies and patterns in sequential or time-series data, LSTMs have been useful for tasks like time-series forecasting and natural language processing. The LSTM model works by establishing complex temporal relationships from the data using a hybrid approach. Then train a RandomForest classifier with the LSTM features that were extracted. This classifier does well with high-dimensional feature spaces and makes accurate predictions based on ensemble learning (Begum et al., 2022). LSTM+RandomForest takes advan- tage of the robustness and interpretability of Random Forests but retains the ability of LSTMs to learn complex tem- poral dependencies. This combination is especially useful when dealing with datasets that consist of sequential and non-sequential features. When combining the extra non- sequential features with the relevant temporal patterns that were extracted by LSTM, you get a full picture of the data using the RandomForest model.

10. LSTM +XGBoost

The LSTM layer is configured with 50 units, making it ideal for learning complex temporal patterns in the sequen- tial data. The input shape is adapted to the reshaped train- ing data and ensures compatibility. Combining this LSTM model with XGBoost offers synergistic benefits: The pri- mary advantage of the 50-unit LSTM is that it can learn com- plex features and long-term dependencies. These learned features are then extracted and utilised as built-in features in the XGBoost classifier to improve generalisation. XGBoost is a powerful gradient boosting technique that cuts across LSTM features with the help of an ensemble of decision trees and can help with more complex data relationships. This method exploits the particular virtues of LSTM and XGBoost, resulting in a highly efficient paradigm for ap- plications associated with sequential data and intricate pat- terns and, hence, enhancing accuracy and generalisation. It is, therefore, recommended that extensive parameter fine- tuning for LSTM and XGBoost be conducted on the basis of special traits of the data to attain better performance(Luo et al., 2021).

11. Adversarial training

Adversarial examples are inputs that humans have care- fully created to trick machine learning models into making incorrect predictions or classifications. Adversarial training is an important machine-learning method for fortifying models against the influence of adversarial examples. The main idea of this approach is to make a model more robust against adversarial attacks by training it using real data supplemented with attached examples. Fig. 6 is an illustration of adversarial examples, which are inputs that are made in a lab to trick machine learning models and look like real data. The adversarial training attacks have methods like the Fast Gradient Method that create perturbed inputs for the model.

This method generates adversarial examples that inform the model's decision-making through minor alterations to input data. When these adversarial examples are incorporated into its training, the ability of the model to adapt to performance is achieved. An example of an adversarial attack is the Fast Gradient Method (FGM), which tactfully alters input data by perturbing it along the direction of its largest loss in value. Training in this technique, therefore, makes it easy for the model to be robust with useful input even after minor adversarial manipulations. This exposes the model to many sophisticated inputs that help it not overfit the data and generalize well. Adversarial training proves to be very handy when it comes to machine learning models that may undergo hostile cases or manipulation, as witnessed in cybersecurity operations. When the model is trained, it learns more about how to deal with adversarial inputs. This makes it more reliable when dealing with adversarial examples in the real world, which is how it can be exposed to cases made by the Fast Gradient Method. (Andriushchenko and Flammarion, 2020).

12. Evaluation metrics

The evaluation metrics employed, such as accuracy, pre- cision, recall, and F1-score, are conventional for assessing the performance of classification models in Eq. (1-4), which are defined as;

- **Accuracy**- Accuracy measures the overall correctness of the model's predictions.

$$\text{Accuracy} = \frac{Number\ of\ Correct\ \text{Predictions}}{\text{TotalNumberofPredictions}} \tag{1}$$

- **Precision**- Precision focuses on the correctness of positive predictions, indicating how many predicted positive instances are actually positive.

$$\text{Pre}csision = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives} \tag{2}$$

- **Recall**- Recall measures the ability of the model to capture all the positive instances, indicating how many actual positives were correctly predicted.

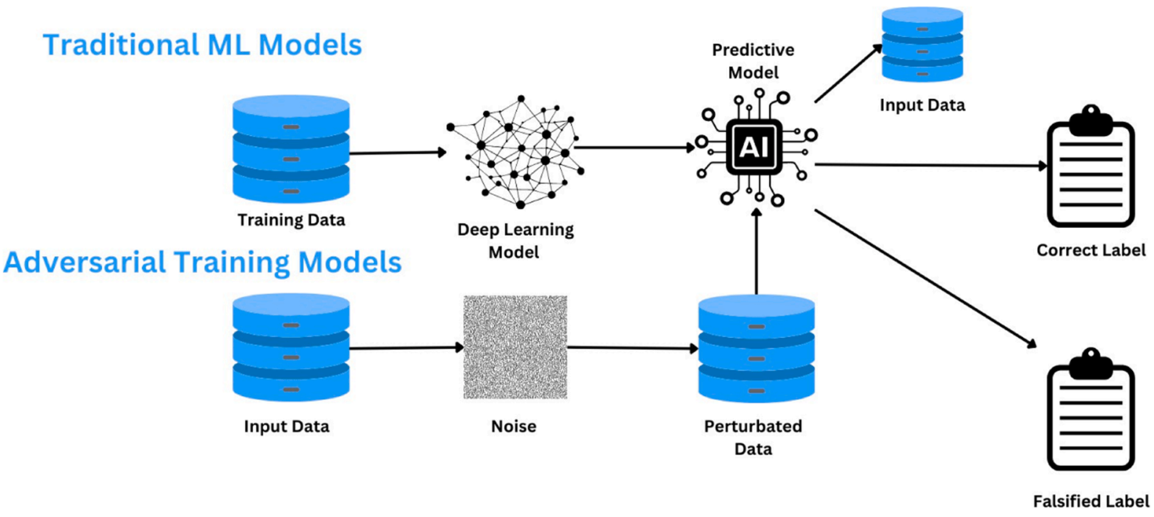

Fig. 6. The difference between Traditional Machine Learning Models and Applying Adversarial Training

$$\text{Recall} = \frac{True\,Positives}{True\,Positives + False\,Negatives} \tag{3}$$

- **F1-score**- The F1-score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance.

$$F1 - score = 2*\frac{\text{Precision*Recall}}{\text{Precision + Recall}} \tag{4}$$

13. Results and discussion

13.1. Machine learning and deep learning results

Through evaluating diverse algorithms in a classification exercise such as Deep Neural Network (DNN), Convo- lutional Neural Network (CNN), and Long Short Term Memory (LSTM) and complemented with some ensemble methods, which are LSTM + Random Forest and LSTM + XGBoost, a distinctive scale of performance orders arises. As shown in Fig. 7., the ensemble models LSTM + Random Forest and LSTM + XGBoost have better results in terms of accuracy metrics above 97%. Most importantly, these models exploit the power of LSTM for sequence-based challenges in combination with the stability of ensemble methods. In Table 1, the DNN model is also a good per- former in terms of detection accuracy and precision, with over 95%, which verifies that deep learning is useful in Mirai botnet detection. Although the CNN model falls slightly behind, performing a little worse for all metrics, recall and F1-score differ the most. These findings underpin the necessity of combining LSTM with ensemble techniques to produce superior classification results.

The performance of combining Long Short-Term Mem- ory (LSTM) models with ensembles including XGBoost and Random Forest is superior over single models espe- cially LSTM with XGBoost presents 97.7% because both approaches have their strengths in predicting splendidly and they are mixed towards achieving quality results. This makes it relevant in problems that involve temporal dynamics, as LSTM captures context dependency in time-series data. LSTM precisely enables learning and maintaining informa- tion about sequence-based patterns in the input data, thus making it outstanding at differentiating difficult rela- tion- ships in the problem of classification.

Table 1
The performance of models

Algorithms	Accuracy	Precision	Recall	F1- score
DNN	0.958	0.953	0.956	0.954
CNN	0.868	0.877	0.838	0.836
LSTM	0.833	0.832	0.796	0.783
LSTM +Random Forest	0.971	0.967	0.968	0.968
LSTM + XGboost	0.977	0.974	0.974	0.974

Notably, XGBoost has been recognized for its ability to work well with imbalanced datasets, capture intricate patterns, and offer remedies for overfitting. XGBoost has some advantages over Random Forest with its decision tree ensemble approach capable of handling high-dimensional data and few outliers. These take advantage of the technique of converting multiple weak learners into an accurate model. Additionally, LSTM can be used with it to improve the gen- eralisation of the model and prevent overfitting. Probabilis- tically, ensemble methods bring diversity, thus improving upon the possible weaknesses of the individual models for better overall enhancement.

The confusion matrices for the LSTM + XGBoost and LSTM models in Fig. 8. and 9. for LSTM were examined to evaluate their performance in classifying traffic into four categories: BenignTraffic, Mirai-greeth_flood, Mirai- greip_flood, and Mirai-udpplain. The LSTM + XGBoost model had significant true positives in all classes at 4,962, 4,229, 3,097, and 4,067 for each of the classes. Yet, the model suffered some fallout on both sides. However, the standalone LSTM model had high true positives, for exam- ple, 4,962, 4,090, 1,554, and 4,061, but a proportion of false

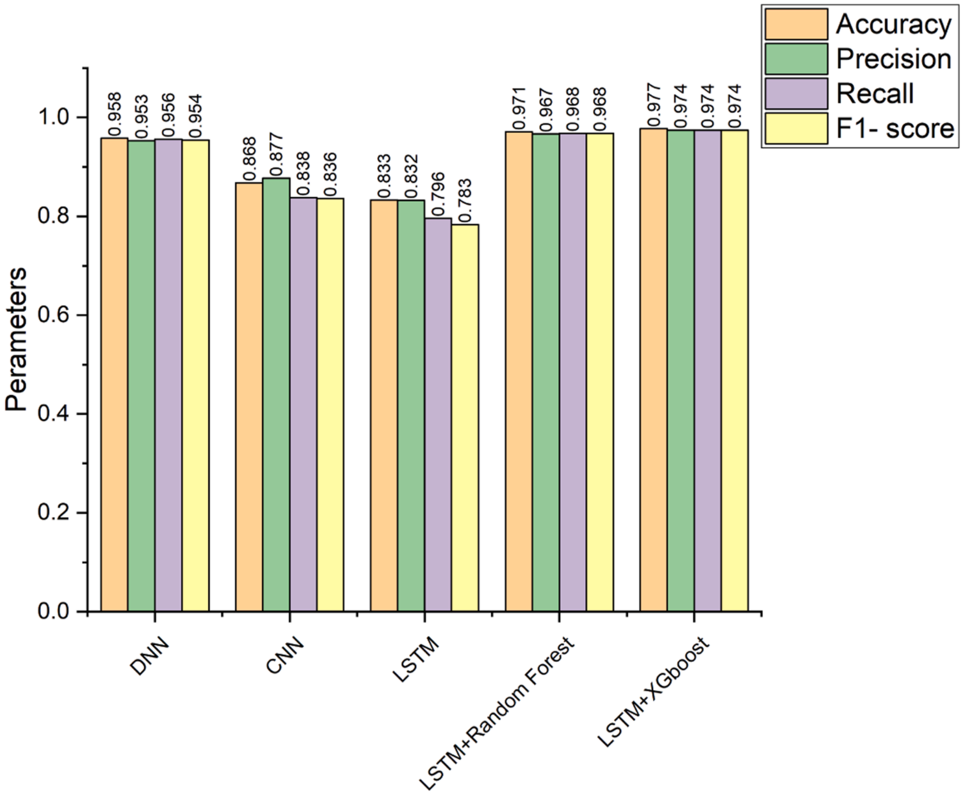

Fig. 7. The performance of model

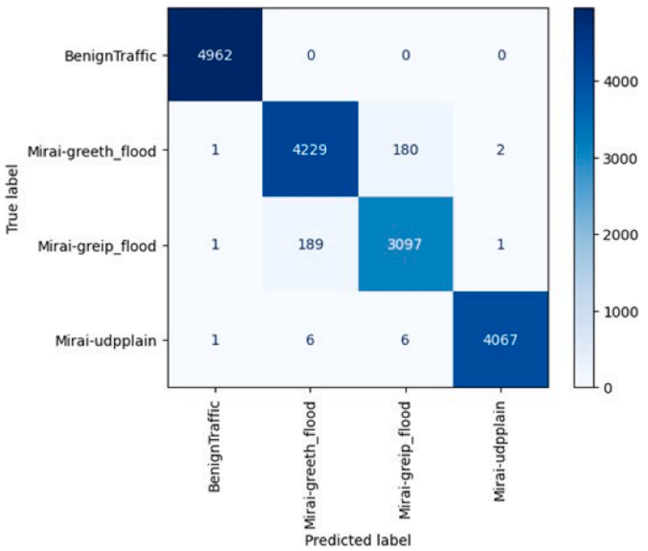

Fig. 8. The confusion matrix of LSTM+XGBoost

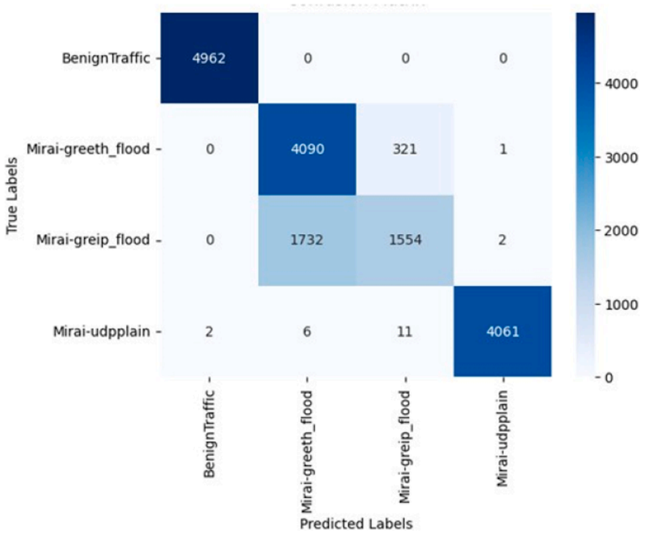

Fig. 9. The confusion matrix of LSTM

positives and false negatives. From the confusion matrix, it is clear that the LSTM model shows more incorrect predictions as compared to its hybrid form, called LSTM+XGBoost.

14. Adversarial training results

Tables 2 and 3 for both normal and adversarial training with XGBoost reveal incredibly high precision, recall, and F1-score values across all classes,numbers that are highly indicative of excellent performance.

The model has a high precision, recall, and F1-score for the ‘BenignTraffic’ class during normal training, which means successful detection of benign network traffic. The model also achieves high precision and recall values in the identification of diverse Mirai botnet attacks such as ‘Mirai- greeth_flood,’ ‘Mirai-greip_flood’ and ‘Mirai-udpplain’.

The adversarial training gives a similar performance as was noted in the normal training. The model has very high precision, recall, and F1-scores for all the classes, which shows that it is quite resilient to adversarial examples. The similar behaviour of these metrics between normal training and adversarial training indicates that the model has learned to effectively handle adversarial perturbation without compromising the overall classification accuracy.

Table 2
The performance of LSTM-XGBoost with normal training across different classes

classes	precision	recall	f1-score
BenignTraffic	0.999	1.000	1.000
Mirai-greeth_flood	0.956	0.959	0.957
Mirai-greip_flood	0.943	0.942	0.943
Mirai-udpplain	0.999	0.997	0.998

Table 3
The performance of LSTM-XGBoost with adversarial training across different classes

classes	precision	recall	f1-score
BenignTraffic	1.000	1.000	1.000
Mirai-greeth_flood	0.955	0.959	0.957
Mirai-greip_flood	0.945	0.941	0.943
Mirai-udpplain	0.998	0.997	0.997

15. Conclusions

To conclude, our mixed approach of using Long Short- Term Memory (LSTM) neural networks and XGBoost with adversarial learning seems to be a good way to find the Mirai botnet. The robustness of Mirai botnet attack defense is realized through the combined powers of LSTM in temporal dependencies, XGBoost’s ensemble learning, and the re- silience that adversarial training brings forth. Our integrated model exhibits a practical effectiveness of 97.7% in counter- ing the dynamic challenges of Mirai botnet activities across various datasets with consistently impressive accuracy. The motivations of this research go well beyond just aiding Mirai botnet detection and provide a trustworthy basis for resilient cybersecurity frameworks that highlight the necessity for proactive and intelligent defense to prevail over an ever-changing cyber threat landscape. However, it is important to acknowledge the limitation of our current model’s po- tential reduced effectiveness against new or evolving Mirai botnet variants not represented in the training dataset. This highlights a critical area for future work, where we aim to enhance the model’s adaptability through the integration of online learning algorithms. Such advancements will allow our detection system to update its knowledge base in real- time as new threat data emerges, ensuring sustained effec- tiveness against the dynamic nature of cyber threats. This future direction not only seeks to mitigate the identified limitation but also to advance the field of cybersecurity defense mechanisms, ensuring that our digital world remains secure against novel and sophisticated cyber attacks.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research work is supported by National Science and Technology Council (NSTC), Taiwan Grant No. NSTC112-2221-E-468-008-MY3.

References

Abbas, N., Nasser, Y., Shehab, M., & Sharafeddine, S. (2021a). Attack-specific feature selection for anomaly detection in software-defined networks. In *2021 3rd IEEE Middle East and north Africa Communications Conference (menacomm)* (pp. 142–146). IEEE.

Abbas, S. G., Hashmat, F., Shah, G. A., & Zafar, K. (2021b). Generic signature development for iot botnet families. *Forensic Science International: Digital Investigation, 38*, Article 301224.

Affinito, A., Zinno, S., Stanco, G., Botta, A., & Ventre, G. (2023). The evolution of mirai botnet scans over a six-year period. *Journal of Information Security and Applications, 79*, Article 103629.

Ahmed, Z., Danish, S. M., Qureshi, H. K., & Lestas, M. (2019). Protecting iots from mirai botnet attacks using blockchains. In *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)* (pp. 1–6). IEEE.

Al-Qerem, A., Abutahoun, B. M., Nashwan, S. I., Shakhatreh, S., Alauth- man, M., & Almomani, A. (2020a). Network-based detection of mirai botnet using machine learning and feature selection methods. In *Handbook of Research on Multimedia Cyber Security* (pp. 308–318). IGI Global.

Al-Qerem, A., Alauthman, M., Almomani, A., & Gupta, B. B. (2020b). Iot transaction processing through cooperative concurrency control on fog– cloud computing environment. *Soft Computing, 24*, 5695–5711.

Alkahtani, H., & Aldhyani, T. H. (2021). Botnet attack detection by using cnn- lstm model for internet of things applications. *Security and Communication- Cation Networks, 2021*, 1–23.

Almomani, A., Alauthman, M., Shatnawi, M. T., Alweshah, M., Alrosan, A., Alomoush, W., & Gupta, B. B. (2022). Phishing website detection with semantic features based on machine learning classifiers: A comparative study. *International Journal on Semantic Web and Information Systems (IJSWIS), 18*, 1–24.

Andriushchenko, M., & Flammarion, N. (2020). Understanding and improving fast adversarial training. *Advances in Neural Information Processing Systems, 33*, 16048–16059.

Begum, A., Kumar, V. D., Asghar, J., Hemalatha, D., & Arulkumaran, G. (2022). A combined deep cnn: Lstm with a random forest approach for breast cancer diagnosis. *Complexity*, 2022.

Deng, Y., & Karam, L. J. (2022). Frequency-tuned universal adversarial attacks on texture recognition. *IEEE Transactions on Image Processing, 31*, 5856–5868.

Gupta, B. B., & Quamara, M. (2020). An overview of internet of things (IoT): Architectural aspects, challenges, and protocols. *Concurrency and Computation: Practice and Experience, 32*, e4946.

GÜVEN, E. Y., et al. (2023). Mirai botnet attack detection in low-scale network traffic. *Intelligent Automation & Soft Computing, 37*.

Hallman, R., Bryan, J., Palavicini, G., Divita, J., & Romero-Mariona, J. (2017). Ioddos-the internet of distributed denial of sevice attacks. In *2nd international conference on internet of things, big data and security* (pp. 47–58). SCITEPRESS.

Hu, B., Gaurav, A., Choi, C., & Almomani, A. (2022). Evaluation and comparative analysis of semantic web-based strategies for enhancing educational system development. *International Journal on Semantic Web and Information Systems (IJSWIS), 18*, 1–14.

Huang, W., Peng, X., Shi, Z., & Ma, Y. (2020). Adversarial attack against lstm- based ddos intrusion detection system. In *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 686–693). IEEE.

Kambourakis, G., Kolias, C., & Stavrou, A. (2017). The Mirai botnet and the iot zombie armies. In *MILCOM 2017-2017 IEEE Military Communica- tions Conference (MILCOM)* (pp. 267–272). IEEE. https://doi.org/10.1109/MILCOM. 2017.8170867.

Khanam, S., Tanweer, S., & Khalid, S. S. (2022). Future of internet of things: Enhancing cloud-based iot using artificial intelligence. *International Journal of Cloud Applications and Computing (IJCAC), 12*, 1–23.

Kiran, M. A., Pasupuleti, S. K., & Eswari, R. (2022). Efficient pairing-free identity-based signcryption scheme for cloud-assisted iot. *International Journal of Cloud Applications and Computing (IJCAC), 12*, 1–15.

Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. *Information Sciences, 467*, 312–322.

Kumar, R., Singh, S. K., Lobiyal, D., Chui, K. T., Santaniello, D., & Rafsanjani, M. K. (2022). A novel decentralized group key management scheme for cloud-based vehicular iot networks. *International Journal of Cloud Applications and Computing (IJCAC), 12*, 1–34.

Liu, R. W., Guo, Y., Lu, Y., Chui, K. T., & Gupta, B. B. (2022). Deep network- enabled haze visibility enhancement for visual iot-driven intelligent transportation systems. *IEEE Transactions on Industrial Informatics, 19*, 1581–1591.

Luo, J., Zhang, Z., Fu, Y., & Rao, F. (2021). Time series prediction of covid-19 transmission in america using lstm and xgboost algorithms. *Results in Physics, 27*, Article 104462.

McDermott, C. D., Majdani, F., & Petrovski, A. V. (2018). Botnet detection in the internet of things using deep learning approaches. In *2018 international joint conference on neural networks (IJCNN)* (pp. 1–8). IEEE.

Memos, V. A., Psannis, K. E., Ishibashi, Y., Kim, B. G., & Gupta, B. B. (2018). An efficient algorithm for media-based surveillance system (eamsus) in iot smart city framework. *Future Generation Computer Systems, 83*, 619–628.

Mezher, A. H., Deng, Y., & Karam, L. J. (2022). Visual quality assessment of adversarially attacked images. In *2022 10th European Workshop on Visual Information Processing (EUVIP)* (pp. 1–5). IEEE.

Mishra, A., Gupta, N., & Gupta, B. (2021). Defense mechanisms against ddos attack based on entropy in sdn-cloud using pox controller. *Telecommunication Systems, 77*, 47–62.

Mishra, A., Joshi, B. K., Arya, V., Gupta, A. K., & Chui, K. T. (2022). Detection of distributed denial of service (ddos) attacks using computational intelligence and majority vote-based ensemble approach. *International Journal of Software Science and Computational Intelligence (IJSSCI), 14*, 1–10.

Mustapha, A., Khatoun, R., Zeadally, S., Chbib, F., Fadlallah, A., Fahs, W., & El Attar, A. (2023). Detecting ddos attacks using adversarial neural network. *Computers & Security, 127*, Article 103117.

Nakip, M., & Gelenbe, E. (2021). Mirai botnet attack detection with auto- associative dense random neural network. In *2021 IEEE Global Com- munications Conference (GLOBECOM)* (pp. 01–06). IEEE.

Neto, E.C.P., Dadkhah, S., Ferreira, R., Zohourian, A., Lu, R., Ghorbani, A.A., 2023. Ciciot2023: A real-time dataset and benchmark for large- scale attacks in iot environment .

Novaes, M. P., Carvalho, L. F., Lloret, J., & Proença, M. L., Jr (2021). Adversarial deep learning approach detection and defense against ddos attacks in sdn environments. *Future Generation Computer Systems, 125*, 156–167.

Omolara, A. E., Alabdulatif, A., Abiodun, O. I., Alawida, M., Alabdulatif, A., Arshad, H., et al. (2022). The internet of things security: A survey encompassing unexplored areas and new insights. *Computers & Secu- rity, 112*, Article 102494.

Sadatacharapandi, T. P., & Padmavathi, S. (2022). Survey on service placement, provisioning, and composition for fog-based iot systems. *International Journal of Cloud Applications and Computing (IJCAC), 12*, 1–14.

Shaikh, S., Rupa, C., Srivastava, G., & Gadekallu, T. R. (2022). Botnet attack intrusion detection in iot enabled automated guided vehicles. In *2022 IEEE International Conference on Big Data (Big Data)* (pp. 6332–6336). IEEE.

Sharma, A., Mansotra, V., & Singh, K. (2023). Detection of mirai botnet attacks on iot devices using deep learning. *Journal of Scientific Research and Technology*, 174–187.

Sharma, R., & Sharma, N. (2022). Attacks on resource-constrained iot devices and security solutions. *International Journal of Software Science and Computational Intelligence (IJSSCI), 14*, 1–21.

Singh, A., & Gupta, B. B. (2022). Distributed denial-of-service (ddos) attacks and defense mechanisms in various web-enabled computing platforms: Issues, challenges, and future research directions. *International Journal on Semantic Web and Information Systems (IJSWIS), 18*, 1–43.

Tembhurne, J. V., Almin, M. M., & Diwan, T. (2022). Mc-dnn: Fake news detection using multi-channel deep neural networks. *International Journal on Semantic Web and Information Systems (IJSWIS), 18*, 1–20.

Wahab, O. A., Bentahar, J., Otrok, H., & Mourad, A. (2017). Optimal load distribution for the detection of vm-based ddos attacks in the cloud. *IEEE Transactions on Services Computing, 13*, 114–129.

Wang, T., Pan, Z., Hu, G., Duan, Y., & Pan, Y. (2022). Understanding universal adversarial attack and defense on graph. *International Journal on Semantic Web and Information Systems (IJSWIS), 18*, 1–21.

Yousaf, I., Ali, S., Bouri, E., & Dutta, A. (2021). Herding on funda-mental/ nonfundamental information during the covid-19 outbreak and cyber-attacks: Evidence from the cryptocurrency market. *Sage Open, 11*, Article 21582440211029911.