Comparative Research Based on Internet Worms

Mundlamuri Venkata Rao*[1], Divya Midhunchakkaravarthy[2], Sujatha Dandu[3]

[1]Research Scholar, [2]Associate Professor, [1,2]Dept. of Comp. Sci. and Multimedia,
Lincoln University College, Malaysia
[3]Professor, Dept. of Comp. Sci. and Engg., Malla Reddy College of Engineering and Technology, India
Corresponding E-mail: vrmundlamuri@gmail.com

Abstract: Strong detection and categorization frameworks are required since the threat on internet worms is still a top issue in the field of cyber security. In this comparative study, we compare and contrast the Deep Learning CNN Framework & the Joint Detection and Classification of Signature and NetFlow inspired Internet Worms using MBGWO-based Hybrid LSTM techniques for detecting and classifying internet worms. Convolutional Neural Networks (CNNs) are used by the Deep Learning CNN Framework to extract and learn complex information from network traffic data related to worms. This framework seeks to accomplish precise worm identification and classification by utilising deep learning. To evaluate the success of the Deep Learning CNN Framework, we examine its design, training procedure, and performance measures. On the other hand, the Joint Detection and Classification strategy combines MBGWO (Modified Grey Wolf Optimization)-based Hybrid LSTM (Long Short-Term Memory) model with Signature and NetFlow-based techniques. The advantages of flow-based analysis, signature-based detection, and the LSTM model's capacity to detect temporal dependencies are all combined in this hybrid technique. We look into this approach's essential elements, optimisation plan, and performance results. We assess the benefits, drawbacks, and overall effectiveness between the two frameworks with regard to of detection precision, recall, accuracy, and F1-score through a comparison analysis. We also evaluate their viability, effectiveness, and possibility of real-world application. The results of this study aid in and understanding the various methods of internet worm detection. We offer insights into the developments, difficulties, and potential future directions in this important field of cyber security by looking at the Deep Learning CNN Framework and the Joint Detection and Classification technique. The creation of stronger and more effective frameworks for thwarting internet worms and protecting network infrastructures is guided by the comparative analysis.

Keywords: Internet worms, Intrusion detection system (IDS), Hybrid long short-term memory, Deep Learning

1 Introduction

Electronic devices are now required to meet our everyday demands, and their users are increasingly involved in a variety of activities, such as information sharing while linking their gadgets over the web [1]. Due to the widespread use of the internet, numerous cybercriminals have easier access to their targets. Attacks by zero-day Internet worms have a profound and wide-ranging effect on every country in the globe. Three billion web-based worm infections occurred worldwide in 2020, according to the Symantec Internet Security Threat (SIST) research [2, 3], and the number of these attacks has been steadily rising each year. The findings of this inquiry indicate that attackers stole and then lost almost nine billion records of data related to one million people.

Malicious software known as computer worms poses serious risks to computer networks, systems, and the general state of cyber security. These self-replicating programmes are made to

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_93

propagate automatically among connected devices, taking advantage of weaknesses to cause harm or disruption. Worms are more harmful than viruses because they can spread quickly and does not need a host programme to do so.

The development of computer worms has completely changed how cyber attacks are conducted. Worms have the ability to take advantage of different security flaws, such as unpatched software, insecure passwords, or unsecured network connections, which gives them an alarmingly quick and effective way to penetrate and compromise computers. Upon entering a network, worms can quickly propagate, infecting several devices and potentially inflicting extensive harm, data loss, or other problems.Computer worms can be developed and spread for a variety of reasons. Some worms are created by online criminals looking to profit from ransom ware assaults, identity theft, or the theft of confidential data. Others might have been created with strategic, political, or ideological, objectives by state-sponsored actors or hacktivist organisations. Computer worms pose serious hazards to people, organisations, and essential infrastructure regardless of their motivations.Computer worm detection and mitigation require a thorough understanding of their traits, transmission techniques, and defense mechanisms. To recognize and mitigate the risks posed by worms, cyber security experts use a variety of approaches such as intrusion detection systems, monitoring of networks, and vulnerability management. Additionally, it is essential to conduct continuing research and build sophisticated detection algorithms and defense systems.In this paper, we use Deep Learning and MBGWO-based hybrid LSTM algorithms to give a comparative analysis on the combined detection of both signature-based and NetFlow-based Internet worms with respect to diverse attacks. We will examine several techniques, frameworks, and methodologies used to discover and categorise computer worms as well as the techniques, frameworks, and approaches used to attack them. We hope to deepen our knowledge of computer worms and advance the creation of stronger, more potent defences against them by contrasting and comparing existing works in this area.

2 Existing Methods

Over the years, a variety of techniques have been considered for identifying and classifying internet worms. The use of PCAP (traffic capture) files for signature-based detection is covered in [4]. This method analyses an online traffic signature and compares it to established rules to determine whether the data being transferred contains normal or attack signatures. Before assessing if the request has normal or attack signatures, NetFlow-based intrusion detection analyses UDP and TCP signatures. As seen in [5], it frequently takes the form of a DoS/DDoS attack or a worm-based virus infection. A denial-of-service (DoS) attack overloads a server, causing it to crash and disable a website or other resource. In a distributed denial-of-service, or DDoS, assault, several computers or workstations overload the target resource. Using the IDS dataset and a range of machine learning methods, including Bayesian networks, decision trees, and random forests, researchers deployed network-based detection algorithms in [6]. The forecasting of whether incoming requests will be regarded as normal or an attack was then done using the trained model. The Bayesian network (BN), a probabilistic model, is used to convey knowledge acquired from uncertain domains. Each node within the domain's structural representation mimics a random variable throughout this time, and each edge shows the conditional likelihood associated with the random variable. Raw NetFlow data can be used to detect Internet intrusions, which has been discussed in [7]. Monitoring at the NetFlow information allowed for the early detection of malicious payloads, preventing the planned catastrophic loss. Generally speaking, viral dispersion via NetFlow attacks is both connection- and packet-oriented [8]. During packet-oriented worm attacks, harmful payloads are created in the network packages [9]. However, there are numerous ways to identify these attack patterns, and it is very rapid and easy to analyse each bytes of ASCII data found in network packets.

In connection-oriented attacks, the bad conduct makes a remarkable number of conversations with the intended users [10] by using a decision tree classifier. For example, the Red Code [11] malware uses port scanning to roam between hosts and broadcasts the most packets. If the fraudulent data transmission behaviour is demonstrated over a number of time periods, the details regarding the spam connection users will be made public. A network like that is seen as being attacked. Here, the number and length of unprotected connections were employed to gauge the aberrant NetFlow. The random forest approach was then utilized to detect worm infestations utilizing conditional possibilities using hypothesis testing. It is challenging to determine the characteristics of the traceable data that the NetFlow attacks leave behind. Recent research has shown that deep neural networks, such as artificial neural networks and back propagated neural networks (BPNN) [12-16], are an efficient way to extract the features, attributes, and latent relationships between the features of various internet worms. Back propagation is a supervised learning method for gradient descent artificial neural networks. The method determines the gradient of an error function in relation to the neural network weights given an artificial neural network and an error function.

3 Proposed Method

The two methods used for comparing are:

Method1:“Deep Learning CNN (DLCNN) Framework for Detection and Classification of Internet Worms”:

Methodology: To identify and categorise internet worms, this research suggests a deep learning system built around convolutional neural networks (CNNs).

CNN Architecture: The article makes use of a CNN architecture with convolutional layers, pooling layers, and fully linked layers that was created expressly for worm detection shown in figure 1.

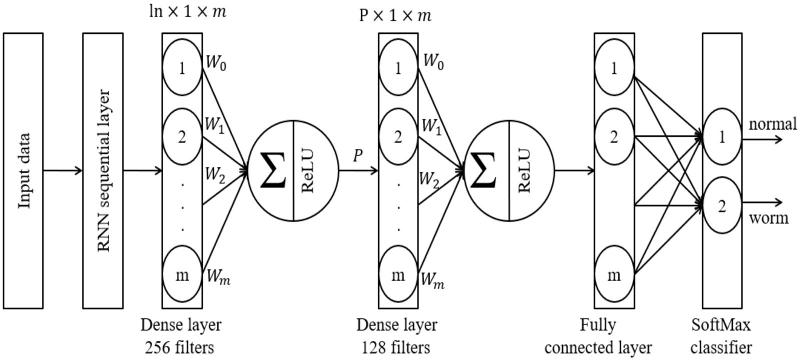

Fig 1. CNN Framework for Detection and Classification of Internet Worms

Data Representation: Images are used to represent the input data, with each image standing for a flow of network traffic.

Training and Classification: A labeled dataset of network flows that includes both regular and worm traffic is used to train the CNN model. The incoming network flows are then classified as either regular or worm traffic using the trained model.

Performance Evaluation: Using a variety of criteria, including accuracy, precision, recall, and F1-score, the study assesses the suggested framework's performance based on the following algorithms.

Table 1: Comparison of different methods with DLCNN method

METHOD	Accuracy (in %)	Precision	Recall	F1-score
Naïve Bayes [23]	44.7238	27.846711	27.0322	23.6494
Decision tree[24]	99.161	75.0390	75.4429	75.21277
Random forest[26]	99.4082	82.6867	81.5245	82.0788
BPNN [29]	96.9428	70.1460	66.7654	65.7605
DLCNN	99.7352	100	100	100

Method 2: Joint Detection and Classification of Signature and NetFlow based Internet Worms using MBGWO-based Hybrid LSTM

Methodology: For the purpose of jointly detecting and classifying both signature-based and NetFlow-based internet worms, this work offers a hybrid strategy combining Multiple Binary Grey Wolf Optimization (MBGWO) and Long Short-Term Memory (LSTM).

Signature-based and NetFlow-based Detection: The research takes into account both signature-based detection and NetFlow-based detection, which examines network flow data for aberrant behaviour. Signature-based detection depends on recognizing particular patterns or signatures connected to known worms.

Hybrid Approach: A hybrid approach is used in this paper to model the temporal relationships in network flows and to optimize the feature selection process.

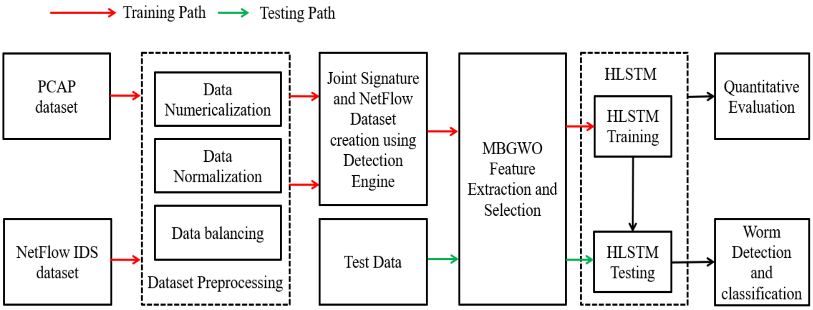

Fig.2. Architecture of Hybrid approach- MBGWO-based Hybrid LSTM

Training and Classification: A dataset with labeled examples of signature-based and NetFlow-based worm traffic is used to train the suggested model. The incoming network flows are then categorized using the learned model.

The features of the two research papers can be described as follows:

Convolutional Neural Network (CNN) Architecture: The paper utilizes a CNN architecture specifically designed for worm detection and classification. This architecture includes convolutional layers, pooling layers, and fully connected layers.Image-based Representation: The input data, which consists of network traffic flows, is represented as images. Each image represents a network flow, allowing the CNN to learn visual patterns associated with worms.Training and Classification: The CNN model is trained on a labeled dataset of network flows that contains both normal and worm traffic. The model learns to classify incoming network flows as normal or worm traffic based on the learned features.Deep Learning Techniques: The paper leverages the power of deep learning techniques, specifically CNNs, to automatically extract relevant features from the network flow data and make accurate predictions.Evaluation Metrics: The performance of the framework is evaluated using metrics such as accuracy, precision, recall, and F1-score to assess its effectiveness in detecting and classifying internet worms.

Joint Detection and Classification of Signature and NetFlow based Internet Worms using MBGWO-based Hybrid LSTM:

Multiple Binary Grey Wolf Optimization (MBGWO): The paper utilizes MBGWO as an optimization technique for feature selection. MBGWO helps identify the most informative features from the signature-based and NetFlow-based worm detection methods.Long Short-Term Memory (LSTM) Architecture: The hybrid model combines the MBGWO-selected features with LSTM, which is recurrent neural network (RNN) architecture capable of modeling temporal dependencies in sequential data.Signature-based and NetFlow-based Detection: The model considers both signature-based detection, which relies on identifying specific patterns or signatures associated with known worms, and NetFlow-based detection, which analyzes network flow data for anomalous behavior.Training and Classification: The proposed model is trained on a dataset that includes labeled instances of signature-based and NetFlow-based worm traffic. The hybrid LSTM model learns to classify incoming network flows as worm or non-worm traffic based on the combined features and temporal dependencies. Performance Evaluation: The performance of the hybrid model is evaluated using metrics such as accuracy, detection rate, false positive rate, and receiver operating characteristic (ROC) curve to measure its effectiveness in jointly detecting and classifying internet worms.These features highlight the specific methodologies and techniques employed in each research paper to address the challenge of internet worm detection and classification.

4 Comparison Of Two Methods

Method1: "Deep Learning CNN Framework for Detection and Classification of Internet Worms", scenarios in which the paper can be used are:

Real Time Worm Detection: The paper proposes a deep learning convolutional neural network (CNN) framework specifically designed for the detection and classification of internet worms. It can be used by network security professionals to identify and classify worms in real-time.**Large-Scale Network Monitoring:** The CNN framework presented in Paper A can be applied to monitor and analyze large-scale networks, making it useful for internet service providers (ISPs) or organizations with extensive network infrastructure.**Machine Learning Researchers:** Researchers interested in applying deep learning techniques to the field of network security can benefit from the insights and methodology presented.

Method 2: "Joint Detection and Classification of Signature and NetFlow based Internet Worms using MBGWO-based Hybrid LSTM", this work can be used in the following scenarios

Integration of Signature and NetFlow Data: The method proposes a hybrid LSTM (Long Short-Term Memory) model that combines both signature-based and NetFlow-based features for worm detection and classification. It can be utilized by network administrators who have access to both types of data to improve the accuracy of their detection systems. **Enhanced Worm Detection Performance**- The hybrid LSTM model presented in Paper B offers improved detection performance by leveraging the strengths of both signature-based and NetFlow-based approaches. This makes it valuable for organizations seeking highly accurate worm detection solutions. **Network Traffic Analysis-** Researchers interested in the analysis of network traffic and the development of hybrid machine learning models can benefit from the methodology and insights presented.

Performance Evaluation: This study assesses the performance of the hybrid model shown in table 2, using metrics such as accuracy, detection rate, false positive rate, and receiver operating characteristic (ROC) curve.

Table 2: Comparison of MBGWO-based HLSTM model with other methods

METHOD	Accuracy	Precision	Recall	F1-score
Naïve Bayes [23]	43.73	32.22	38.59	28.98
Decision Tree [24]	99.161	75.0390	75.4429	75.21277
Random forest [26]	99.4082	82.6867	81.5245	82.0788
BPNN [29]	96.9428	70.1460	66.7654	65.7605
DLCNN [41]	99.7352	99.98	99.52	99.75
MBGWO-based HLSTM model	99.84	100	100	100

The limitations of method 1 are Data Limitations: The effectiveness of the proposed CNN framework heavily relies on the availability of high-quality and diverse training data. Obtaining such data can be challenging, especially for emerging or rare types of worms.Model Complexity: Implementing and training a deep learning CNN model requires substantial computational resources and expertise, which might limit its accessibility for organizations with limited resources or technical capabilities.Generalization to New Worm Types: While the CNN framework might perform well on known worm types, its ability to generalize and detect new or unknown worm variants might be limited.

The limitations of method 2 are Data Availability- Similar to Paper A, the availability of labeled training data that encompasses both signature-based and NetFlow-based features might be limited, making it challenging to build and evaluate the hybrid LSTM model. Complexity and Resource Requirements-Implementing and training a hybrid LSTM model with both signature-based and NetFlow-based features requires significant computational resources and expertise in deep learning. This might restrict its applicability to organizations with sufficient resources and technical capabilities. Scalability-The scalability of the proposed model to large-scale networks or real-time scenarios might not be explicitly addressed in the paper, potentially limiting its effectiveness in those contexts.

In summary, the procedures and techniques used for worm identification and classification are the primary distinction between the two research publications. The second paper presents a hybrid strategy combining MBGWO and LSTM for the joint detection of signature-based and NetFlow-based worms, whereas the first paper focuses on a CNN-based deep learning system that analyses network flows as images. While both articles seek to increase the precision and efficacy of worm detection, their methods for doing so differ.

5. Conclusion And Future Work

In conclusion, computer worms represent a significant and persistent threat to computer systems and networks. These self-replicating malicious programs can quickly propagate and cause widespread damage, making them a top concern in the field of cyber security. Throughout this research paper, we have explored different methodologies, frameworks, and approaches used for the detection and classification of computer worms. By comparing and contrasting various research works, we have gained insights into the evolving nature of worms and the strategies employed to combat them. Efforts to detect and mitigate computer worms involve a combination of proactive measures, such as vulnerability management, network monitoring, and intrusion detection systems. Additionally, the development of advanced detection algorithms and defensive mechanisms is crucial to stay one step ahead of evolving worm-based threats. However, the battle against computer worms remains a constant challenge. Worm authors continue to refine their techniques, exploit new vulnerabilities, and find ways to evade detection. Therefore, ongoing research and collaboration among academia, industry, and cyber security professionals are essential to develop innovative approaches and defenses against emerging worm threats.By advancing our understanding of computer worms and continuously improving detection and classification techniques, we can enhance the security posture of computer systems, networks, and the overall digital landscape. Future research should focus on improving the accuracy and efficiency of detection mechanisms, exploring behavioral analysis and machine learning techniques, and staying updated with emerging worm trends and attack vectors. Ultimately, through collaborative efforts and a comprehensive understanding of computer worms, we can better safeguard our digital ecosystems, protect sensitive data, and mitigate the potential damage caused by these malicious entities.

The detection and classification of computer worms continue to be a dynamic field with several potential avenues for future research and development. Some key areas of future scope in this domain include: Advanced Machine Learning Techniques, Behavior-based Detection, Real-time Monitoring and Response, Hybrid Approaches, IoT and Mobile Worm Detection, Collaborative Defense Mechanisms, Worm Mitigation Strategies, Evaluation Frameworks. By exploring these future research directions and investing in innovative approaches and technologies, we can enhance the resilience and effectiveness of worm detection and classification systems, ultimately contributing to a safer and more secure digital environment.

References

[1]. Koganti, V. S., Galla, L. K., & Nuthalapati, N. (2016, December). Internet worms and its detection. In 2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT) (pp. 64-73). IEEE.

[2]. Rasheed, Mohammad M., Alaa K. Faieq, and Ahmed A. Hashim. "Android Botnet Detection Using Machine Learning." Ingénierie des Systèmesd'Information 25.1 (2020).

[3]. Y. Li, W. Dai, Z. Ming, and M. Qiu, "Privacy protection for preventing data over-collection in smart city," IEEE Transactions on Computers, vol. 65, no. 5, pp. 1339–1350, 2016.

[4]. Rasheed, Mohammad M., et al. "Detection algorithm for internet worms scanning that used user datagram protocol." International Journal of Information and Computer Security 11.1 (2019): 17-32.

[5]. Eskandari, R., Shajari, M., & Ghahfarokhi, M. M. (2019). ERES: an extended regular expression signature for polymorphic worm detection. Journal of Computer Virology and Hacking Techniques, 15(3), 177-194.

[6]. Wu, B., Li, Q., Xu, K., Li, R., & Liu, Z. (2018, October). Smartretro: Blockchain-based incentives for distributed iot retrospective detection. In 2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS) (pp. 308-316). IEEE.

[7]. Ochieng, Nelson, Waweru Mwangi, and Ismail Ateya. "Optimizing computer worm detection using ensembles." Security and Communication Networks 2019 (2019).

[8]. Yadav, Ajit, Rahul Yadav, and Mangesh Tiwari. "Website Security for Detection and Prevention of Attacks." i-Manager's Journal on Software Engineering 14.3 (2020): 37.

[9]. Nath, Sayan, Debdutta Pal, and Avijit Mondal. "Destination source correlation algorithm to monitor local networks: A worm detection technique." Materials Today: Proceedings (2021).

[10]. Kaur, Sanmeet, and Maninder Singh. "Hybrid intrusion detection and signature generation using deep recurrent neural networks." Neural Computing and Applications (2019): 1-19.

[11]. Chen, Liguo, et al. "Detection of DNS DDOS attacks with random forest algorithm on spark." Procedia computer science 134 (2018): 310-315.

[12]. Reddy Madhavi, K., Madhavi, G., Rupa Devi, B., Kora, P., (2020), "Detection of Pneumonia Using Deep Transfer Learning architectures", International Journal of Advanced Trends in Computer Science and Engineering, 9(5), ISSN 2278-3091, Pp:8934- 8937,2021.

[13]. Rajani, Akula & Kora, Padmavathi & Madhavi, Reddy & Jangaraj, Avanija. (2021). Quality Improvement of Retinal Optical Coherence Tomography. 1-5. 10.1109/INCET51464.2021.9456151.

[14]. Srinivas, K., Gagana Sri, R., Pravallika, K. et al. COVID-19 prediction based on hybrid Inception V3 with VGG16 using chest X-ray images. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023-15903-y.

[15]. Tamrakar, S., Choubey, S.B., & Choubey, A. (Eds.). (2023). Computational Intelligence in Medical Decision Making and Diagnosis: Techniques and Applications (1st ed.). CRC Press. https://doi.org/10.1201/9781003309451

[16]. Dejene, Dawit, Basant Tiwari, and Vivek Tiwari. "TD2SecIoT: temporal, data-driven and dynamic network layer based security architecture for industrial IoT." International Journal Of Interactive Multimedia And Artificial Intelligence, 6, no. 4 (2020): 146-156.