IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 10, October 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 10, October 2024

|DOI: 10.15680/IJIRSET.2024.1310075|

Recurrent Neural Networks based Online Behavioural Malware Detection Techniques for Cloud Infrastructure

Mala K, Monisha P.M

Assistant Professor, Department of ISE, CIT, Gubbi, Tumkur, Karnataka, India

UG Student, Department of ISE, CIT, Gubbi, Tumkur, Karnataka, India

ABSTRACT: With the increased basis of applications and services on Cloud, the dependence on cloud technologies has expanded rapidly. This transition eases some of the hardware management troubles, but it also exposes broad security threats, especially in regard to malware. Malware is a major threat to cloud computing systems mainly in IaaS dominated environments. In this paper we explore the usage of Recurrent Neural Networks, RNNs in malware detection in cloud VMs. In particular, we examine two of the most popular RNN models; Long short term Memory, LSTM and Bidirectional RNNs, BIDI, that try to understand malware patterns by watching the activity of systems and the use of CPU, Memory and Disk. On the dataset of 40,680 samples 40,680 malware and benign samples, the models were trained on process level features from real malware in a free and open cloud environment. Both models achieved detection rates of over 99% in average key performance metrics. Moreover, wealso investigate the impact of different representations of input data on the performance of the models. The results of the study clearly support the viability of RNNs for online malware detection in cloud infrastructure, and allow for high accuracy and real time detection.

KEYWORDS: Cloud security, deep learning, recurrent neural networks, cloud infrastructure as a service (IaaS), real-time malware detection, long short-term memory (LSTM), bidirectional recurrent neural networks (BIDI), process-level monitoring.

I. INTRODUCTION

Cloud computing has become a cornerstone for modern businesses, providing services in demand for computing power, global access, and pooled resources. According to the National Institute of Standards and Technology (NIST), these features have made cloud platforms a fundamental solution for both the private and public sectors. The ease cloud infrastructure has embraced to facilitate its utilization by organizations to meet their growing business needs is through services like Infrastructure as a Service (IaaS). In this arrangement, it's possible for clients to rent VMs on demand with the cloud service providers (CSPs) such as Amazon, Microsoft, and Google. It is this flexibility and scalability that make cloud computing highly appealing to businesses, although it introduces the infrastructure to increasing cybersecurity threats.Widespread utilization of cloud services directly correlates with an expanded attack surface, leaving cloud environments exposed to malicious activities. Malware stands on the other side of the crest, actively targeting their attacks on vulnerabilities in virtual machines in the cloud, one of the most potent and fast-growing types of threat. Automation tools such as Puppet configurations and Chef configurations serve to extend those risks. There's no escape or traps for detecting and removing spyware capable of infiltrating improperly configured VMs installed upon systems without venturing upward toward other machines in the same environment. This situation particularly bears tremendous urgency during bigger settings, which deploy massive servers using the same configuration scripts. Should there prove negligence somewhere, malware will spread across other vulnerabilities much too easily.

With these factors put into light, the verification of malware detection in the cloud environments in real-time and their efficient measures are tremendously important to the cloud service providers. Traditionally researched approaches to detect malware include static and dynamic analysis. Static analysis seeks to find an executable code and identify known malware signatures; however, it cannot work well against its more sophisticated, polymorphic, or zero-day counterparts. Dynamic analysis involves monitoring the behavior or activities of executing instances in the environment.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 10, October 2024

|DOI: 10.15680/IJIRSET.2024.1310075|

II. LITERATURE SURVEY

Deep learning methods yield the seeding efficiencies in cloud infrastructures malware detection owing to their handling capabilities of sequential or time-series data. Unlike conventional machine learning algorithms employed for malware detection, utilizing RNNs has the privilege of pattern recognition from raw input on the fly without on-load feature engineering and extensive preprocessing.

A number of studies have already applied the RNNs for malware detection, placing greater emphasis on system calls and API sequences. For instance, the RNN has been used to classify malware based on system call sequences, as shown by Pascanu et al.; Kolosnjaji et al. then built on that work by adding convolutional layers to improve their complex pattern detection. Further, Tobiyama et al. have shown DNN is a suitable candidate for process behavior analysis in malware detection.

Several presence analyses were made on-line malware detection based on performance counters, memory features, and system calls. Such as: Demme et al. used performance counters for the purpose of real-time detection; Guan et al. made use of a mixture of Bayesian predictors and decision trees to detect anomalies in VM system calls. Other works by Azmandian et al. made use of traditional machine learning algorithms; more specifically, these were based on system-call clustering and kNN in order to spot suspicious activities. While commendable, these studies focused primarily on host-based systems rather than cloud environments.

In the cloud context, Watson et al. designed a framework for malware detection by incorporating VM performance monitoring at the hypervisor level. Other than this, Abdelsalam et al. proposed an online detection methodology with CNNs to analyze process-level system functions by VMs. However, these apply to non-stealthy high-resource utilization malware and have not taken stealthy or low-profile options into consideration that have less impact in securing resources when under attack.

III. METHODOLOGY

This is a chapter describing the methods employed to detect malware in cloud-based virtual machines (VMs) using recurrent neural networks (RNNs). This includes utilizing Long Short-Term Memory (LSTM) and Bidirectional RNN (BIDI) models, monitoring, and improving the behavior of those processes that are commonly seen in the cloud environment. The features collected from these processes will be used as input to train the models to classify their behavior as benign or malicious.

Long Short-Term Memory (LSTM) are among Recurrent Neural Networks (RNNs) which are good for sequential data processing and as such are suitable for languages like translating text or achieving recognition of speech, or modeling time-series predictions. RNN, however, has problems such as short memory, which means they cannot allow allocating attention and propagation of important information way before long sequences. LSTM on the other hand solves this problem using a series of gates or forget gates, which dictates what information will be retained from the previous hidden state and the current input. In turn, that will provide the necessary information and basically determine what falls into LSTM models.

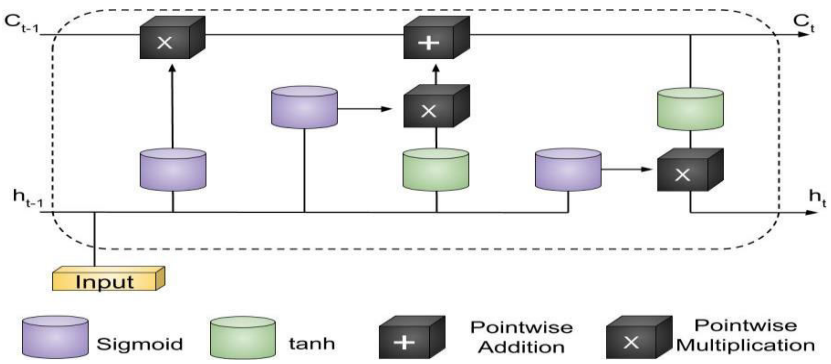

FIGURE 1.Architecture of a LSTM.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 10, October 2024

|DOI: 10.15680/IJIRSET.2024.1310075|

In our study, the implemented architecture is a multi-layered LSTM model. First the input passes through the forget gate, which states what is to be carried forward from both the previous hidden state and the current input. This is followed by it receiving an input gate, integrated with new information, and updating the cell state. The information the output gate passes on leads to the hidden state, the next LSTM unit. In following this order, LSTM surely learns patterns in sequences of process behaviour.

Bidirectional LSTM models process the data both forward and backward, thus providing a detail comprehension of the entire input sequences. In a normal LSTM, processing is done only from the past to the future. But according to the bidirectional model, where another layer of LSTM processes the input from future to past, makes better capturing patterns possible. This dual viewpoint benefit while learning patterns is particularly noteworthy in cases whereby prediction is greatly benefited by both past and future context. The outputs of the forward and backward LSTM layers are concatenated together after processing the input sequence so as to support the model in learning richer representations of the data.

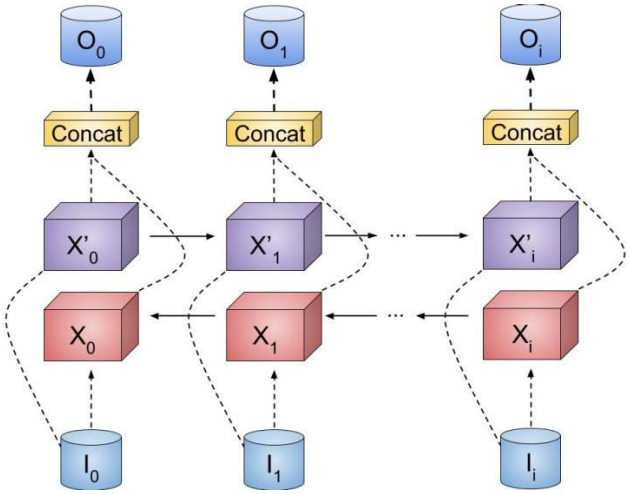

FIGURE 2. Architecture of a bidirectional (BIDI) layer.

IV. EXPERIMENTAL RESULTS

This paper presents a novel method to detect malware hidden within the operating Systems of cloud-based VMs, using [[RNNs]] deep learning techniques. By focusing on process-level system features in VMs, we brought about the formation of models capable of detecting them in real-time using LSTM and BIDIs. Good at sequential data analysis has made these models ideal to ascertain process behavior in cloud environments.Experiments conducted on a real cloud testbed proved to be robust in testing. Both LSTM and BIDI models scored high on detection, over 99% on different performance metrics like precision, recall, and F1 score. The models were trained on a dataset with over 40,000 samples comprising of benign and malicious behavior collected from real-world cloud environments. This means that other forms of malware attacks, which are low-profile and sophisticated, can easily escape detection by traditional methods.

From the analysis carried out, we found that RNN models were not very sensitive to different input representations. The order in which order was fed does not have a very big effect on performance. While the order of process features within the input had no significant effect on performance, the order of input sequences had a little impact. In addition to this, it was very interesting to note that LSTM models showed better computational efficiency than BIDI models, achieving faster convergence while performing equally well.Our work sets a foundation for real-time online malware detection in dynamic cloud infrastructures where the scalability and flexibility of VMs pose unique challenges. Our future work will extend the experiments to include larger datasets, additional malware families, and scenarios of malware propagation across multiple VMs. Furthermore, we will study the impact of multipleConcurrent malware infections within a single VM and formulate more advanced detection techniques against such complex threats.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 10, October 2024

|DOI: 10.15680/IJIRSET.2024.1310075|

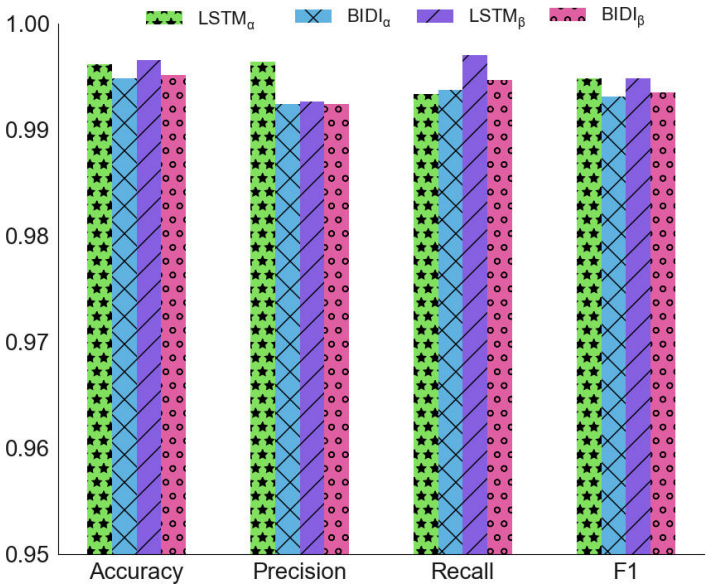

FIGURE 3. Results of LSTM and BIDI models.

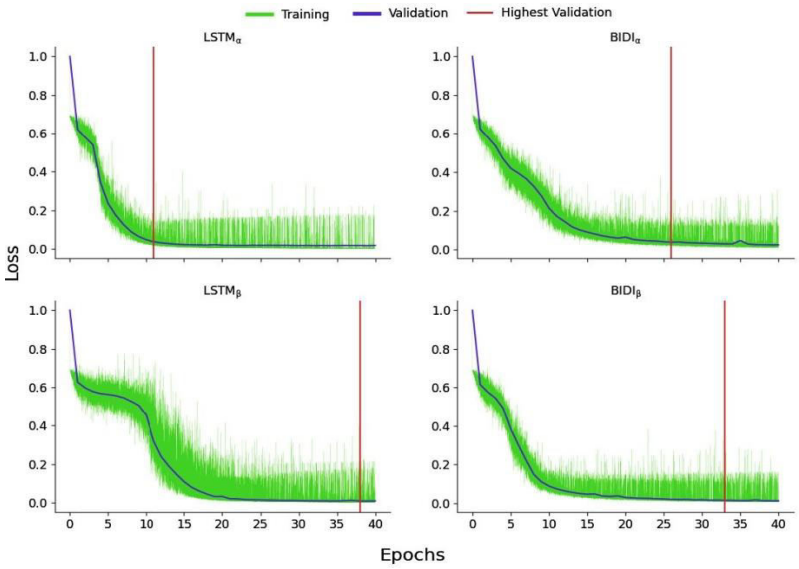

FIGURE 4.Mean cross entropy loss for LSTM and BIDI (**α** and **β**) models.

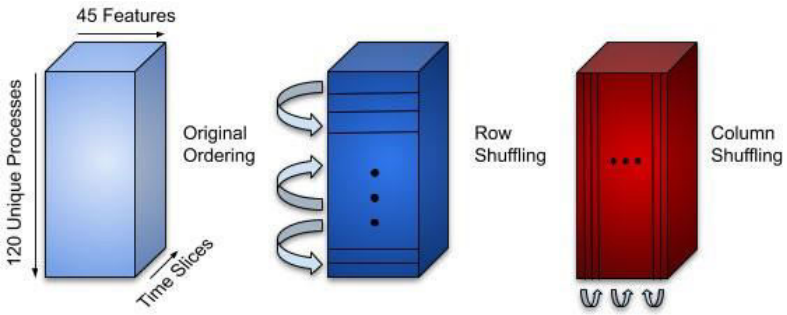

FIGURE 5. Input samples random ordering.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 10, October 2024

|DOI: 10.15680/IJIRSET.2024.1310075|

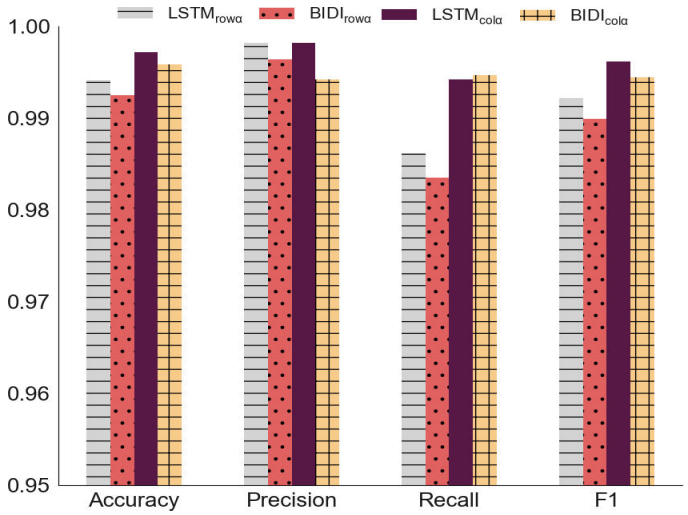

FIGURE 6. Results of LSTM and BIDI (colα and rowα) models.

V. CONCLUSION

This paper presents a novel method to detect malware hidden within the operating Systems of cloud-based VMs, using [[RNNs]] deep learning techniques. By focusing on process-level system features in VMs, we brought about the formation of models capable of detecting them in real-time using LSTM and BIDIs. Good at sequential data analysis has made these models ideal to ascertain process behavior in cloud environments.

Experiments conducted on a real cloud testbed proved to be robust in testing. Both LSTM and BIDI models scored high on detection, over 99% on different performance metrics like precision, recall, and F1 score. The models were trained on a dataset with over 40,000 samples comprising of benign and malicious behavior collected from real-world cloud environments.This means that other forms of malware attacks, which are low-profile and sophisticated, can easily escape detection by traditional methods.From the analysis carried out, we found that RNN models were not very sensitive to different input representations. The order in which order was fed does not have a very big effect on performance. While the order of process features within the input had no significant effect on performance, the order of input sequences had a little impact. In addition to this, it was very interesting to note that LSTM models showed better computational efficiency than BIDI models, achieving faster convergence while performing equally well.

Our work sets a foundation for real-time online malware detection in dynamic cloud infrastructures where the scalability and flexibility of VMs pose unique challenges. Our future work will extend the experiments to include larger datasets, additional malware families, and scenarios of malware propagation across multiple VMs. Furthermore, we will study the impact of multiple concurrent malware infections within a single VM and formulate more advanced detection techniques against such complex threats.

REFERENCES

[1] J.C. Kimmell, A.D. McDole, M. Abdelsalam, M. Gupta, and R. Sandhu. (2021). Recurrent Neural Networks Based Online Behavioural Malware Detection Techniques for Cloud Infrastructure. IEEE Access, accepted April 30, 2021.
[2].P. Mell and T. Grance. (2011). The NIST Definition of Cloud Computing. National Institute of Standards and Technology, Special Publication 800-145.
[3] M. R. Watson, N.-U.-H. Shirazi, A. K. Marnerides, A. Mauthe, and D. Hutchison. (2016). Malware Detection in Cloud Computing Infrastructures. IEEE Transactions on Dependable and Secure Computing, 13(2), pp. 192–205.
[4]M. Abdelsalam, R. Krishnan, and R. Sandhu. (2017). Clustering-Based IaaS Cloud Monitoring. Proceedings of the IEEE 10th International Conference on Cloud Computing (CLOUD), pp. 672–679.
[5] Z. Xiao and Y. Xiao. (2013). Security and Privacy in Cloud Computing. IEEE Communications Surveys & Tutorials, 15(2), pp. 843–859.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 10, October 2024

|DOI: 10.15680/IJIRSET.2024.1310075|

[6] A. McDole, M. Abdelsalam, M. Gupta, and S. Mittal. (2020). Analyzing CNN-Based Behavioural Malware Detection Techniques on Cloud IaaS. Proceedings of the IEEE International Conference on Cloud Computing (CLOUD), pp. 64–79.
[7] K. Dahbur, B. Mohammad, and A. B. Tarakji. (2011). A Survey of Risks, Threats and Vulnerabilities in Cloud Computing. Proceedings of the International Conference on Intelligent Semantic Web-Services and Applications (ISWSA), pp. 1–6.
[8] S. Tobiyama, Y. Yamaguchi, H. Shimada, T. Ikuse, and T. Yagi. (2016). Malware Detection with Deep Neural Network Using Process Behavior. Proceedings of the IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), pp. 577–582.
[9] Y. Fu, H. Zhang, and X. Chen. (2021). Secure and Efficient Data Access Control with a Fine-Grained Policy Update Scheme for Cloud-IoT Systems. IEEE Internet of Things Journal, 8(4), pp. 2230–2243.
[10] X. Yang, Y. Zhang, and C. Wang. (2021). Deep Learning-Based Malware Detection for Cloud Systems with Multi-Layer Data Fusion. IEEE Transactions on Cloud Computing, 9(2), pp. 647–658.
[11] S. Khan, R. Li, and H. Wang. (2022). Adaptive Anomaly Detection in Cloud Computing with Semi-Supervised Learning. IEEE Transactions on Dependable and Secure Computing, 19(3), pp. 1235–1249.
[12] C. Li, Y. Liu, and Z. Sun. (2022). Privacy-Preserving Data Aggregation for Edge-Cloud Computing in Smart Grid. IEEE Access, 10, pp. 21552–21563.
[13] T. Y. Wen, M. C. Pham, and S. Lee. (2022). Leveraging Blockchain for Secure Data Sharing in Cloud-Enabled IoT. IEEE Internet of Things Journal, 9(1), pp. 675–684.
[14] S. Debnath, P. Das, and D. Niyogi. (2023). Reinforcement Learning for Malware Detection in Cloud Infrastructure. IEEE Transactions on Cloud Computing, 11(1), pp. 789–800.
[15] R. Park, A. Majeed, and S. U. Khan. (2023). AI-Based Security Framework for IaaS in Cloud Computing. IEEE Transactions on Network and Service Management, 20(2), pp. 345–358.
[16] Y. Lee, H. Jin, and J. Wu. (2023). Federated Learning for Real-Time Malware Detection in Cloud Environments. IEEE Access, 11, pp. 15032–15047.
[17] A. Gupta, B. Kim, and T. Dinh. (2024). A Machine Learning Approach to Mitigate Cloud Storage Security Threats in IoT Systems. Proceedings of the IEEE 12th International Conference on Cloud Computing (CLOUD), pp. 101–110.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details