Detection of Zero-Day Attacks using CNN and LSTM in Networked Autonomous Systems 'IEEE CNS 23 Poster'

Hassan El alami and Danda B. Rawat
Department of Electrical Engineering and Computer Science
Howard University, Washington, DC 20059, USA
hassan.elalami@howard.edu, danda.rawat@howard.edu

2023 IEEE Conference on Communications and Network Security (CNS) | 979-8-3503-3945-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/CNS59707.2023.10288680

***Abstract*---In this paper, we propose a novel approach for detecting zero-day attacks on networked autonomous systems (AS). The proposed approach combines CNN and LSTM algorithms to offer efficient and precise detection of zero-day attacks. We evaluated the proposed approach's performance against various ML models using a real-world dataset. The experimental results demonstrate the effectiveness of the proposed approach in detecting zero-day attacks in networked AS, achieving better accuracy and detection probability than other ML models.**

***Index Terms*---Artificial Intelligence, Machine Learning, Autonomous Systems, Cybersecurity, Zero-Day Attacks**

I. Introduction and Motivation

Increasing cyberattacks in networked autonomous systems (AS) introduce new security challenges regarding zero-day attacks. Intrusion detection systems (IDS) are trained to detect specific attacks in order to protect the AS application, but the attacks that are not yet known to IDS (i.e., zero-day attacks) still pose challenges and concerns regarding the user's privacy and security in those applications. Mostly, anomaly detection approaches are based on artificial intelligence (AI). AI includes classical machine learning (ML) and deep learning (DL) techniques. As ML models have low prediction quality and detection rates for untrained data, DL techniques are able to overcome this problem, resulting in improved prediction accuracy with unknown data without overfitting [1], such as convolutional neural networks (CNNs) and long short-term memory (LSTM). A combination of CNN and LSTM is presented and demonstrated in this paper to be more effective at detecting zero-day attacks generated by adversaries than classical ML. Real-world dataset TON-IoT [2] is used in the proposed ML models, as well as scanning attack data, which is the most common type of attack in networked AS. A scenario containing backdoor attack data will be included in the testing phase of the selected models to provide the full picture of this evaluation.

II. Proposed Approach

In this section, we propose a new architecture of Convolutional LSTM (1D-ConvLSTM) for detecting zero-day attacks on networked autonomous systems (AS). ConvLSTM integrates conventional neural network (CNN) structures with LSTM units, making it adept at handling spatiotemporal data, a trait valuable in cybersecurity. For zero-day attack detection, ConvLSTM processes sequenced network traffic data, with its convolution operations identifying spatial patterns and its LSTM architecture capturing long-term temporal dependencies. ConvLSTM can identify zero-day attacks even without a known signature, by recognizing deviations from networked AS behavior. Spatial-temporal analysis enhances the capability of detecting stealthy and previously unknown malicious behavior associated with zero-day attacks. Within the evolved 1D-ConvLSTM architecture, the convolutional processes are used similarly to the approach in [3]. In contrast to the classical LSTM, 1D-convLSTM surpasses in capturing spatiotemporal connections in the input with a reduced model parameter count. This approach forecasts a cell's impending state by taking into account its antecedent states and the contributions from its local neighbors. The designed architecture employs LSTM modules, where every module possesses a memory state denoted as c_t at a specific time t. In Fig. 1, three gates control the ability to read or alter the data stored within a memory cell: input gate (i), output gate (o), and forget gate (f). 1D-convLSTM cells determine what information to retain through gates that can be activated and deactivated, and when to initiate reads, writes, or removals. At every time t, the LSTM takes in inputs from two distinct sources: the present frame x_{t+1} and the previous hidden states from all LSTM units in the identical layer h_t, for each of the four connection points. These inputs

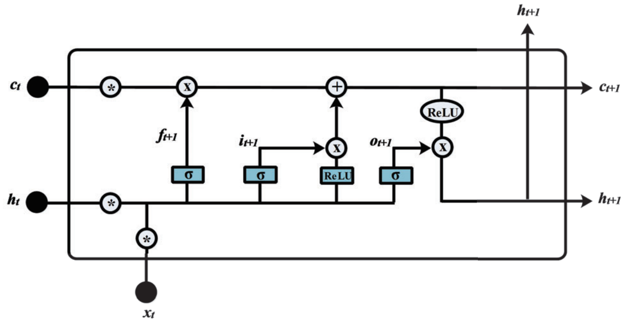

Figure 1: Internal architecture of 1D-ConvLSTM model.

979-8-3503-3945-1/23/$31.00 ©2023 IEEE

are then aggregated, taking into account the bias factors b_f, b_i, b_c, and b_o. The gates' function is determined by channeling their collective input through logistic functions. Furthermore, the modification procedures for a layer of 1D-convLSTM units can be encapsulated in the following equations:

$$\begin{aligned}
i_{t+1} &= \sigma\left(w_{xi} * x_{t+1} + w_{hi} * h_t + w_{ci} \odot c_t + b_i\right) \\
f_{t+1} &= \sigma\left(w_{xf} * x_{t+1} + w_{hf} * h_t + w_{cf} \odot c_t + b_f\right) \\
c_{t+1} &= f_{t+1} \odot c_t + i_{t+1} \odot \max(0, w_{xc} * x_{t+1} + w_{hc} * h_t + b_c) \\
o_{t+1} &= \sigma\left(w_{xo} * x_{t+1} + w_{ho} * h_t + w_{co} \odot c_{t+1} + b_o\right) \\
H_{t+1} &= o_{t+1} \odot \max(0, c_{t+1})
\end{aligned}$$

Where, w_h and w_x represent the matrices for hidden-to-hidden and input-to-hidden weights, respectively, while b denotes the bias matrix. The distinction between 1D-convLSTM and LSTM lies in x, h, and c. These represent the matrices of spatiotemporal input, hidden state, and cell memory, respectively.

III. Performance Evaluation

A. Dataset description

Evaluating the proficiency of anomaly-based IDS is of paramount importance. Commonly, this is achieved by leveraging data from renowned datasets and replicating real-world conditions for both training and testing the models. As such, the use of the dataset must reflect the study's unique demands. The use of current networked AS traffic data, diverse attack data points, and adequate standard traffic data is imperative in our study. It appears that the TON-IoT dataset described in [2] meets these criteria, making it a suitable test dataset for assessing the selected ML models. Furthermore, data is directly prepared for training models. To evaluate the effectiveness of various AI-based cybersecurity applications for networked AS, we use the TON_IoT dataset. After removing the unwanted features, we selected *date*, *timelatitude*, *longitude*, *label*, and *type* as features for the module training.

B. Results and Discussion

As the study aims to detect zero-day attacks, we use training phase data that contains scanning attack traffic. The testing phase includes data related to backdoor attacks. Further, we evaluated the proposed approach in comparison with Linear Regression (LR), Naive Bayes (NB), AdaBoost, CNN, and LSTM using a variety of performance metrics, including accuracy, recall, F1 score, probability of detection (PD), and receiver operating characteristic (ROC). Fig. 2 shows logistic regression achieves 83.16% accuracy, indicating a reliable detection rate; however, its ROC of 70.98% and probability of detection at 60.82% suggest it might struggle with differentiating true attacks from false ones. While AdaBoost achieves a higher accuracy at 89.45%, it stands out with an exceptional ROC of 99.97%, indicating superior discriminatory ability. In spite of the slightly lower accuracy of 77.54%, the NB model has a high recall of 99.98%, thus capturing almost all attacks, although potentially false alarms may occur. CNN model balances with 87.81% accuracy and an ROC of 87.81%, implying consistent detection and discrimination capabilities. By exhibiting a near-perfect accuracy of 98.41%, the LSTM model is most adept at detecting zero-day attacks. The 1D-ConvLSTM model is able to detect zero-day attacks more accurately than any other model, with an accuracy of 99.09%, which implies that it is nearly always able to make correct predictions. With a recall of 99.99%, it almost never misses genuine attacks, and its ROC of 98.63% indicates its exceptional ability to distinguish between malicious and benign traffic. Despite the high recall, the PD of 64.34% suggests differences in metric definitions or nuances in the model's behavior.

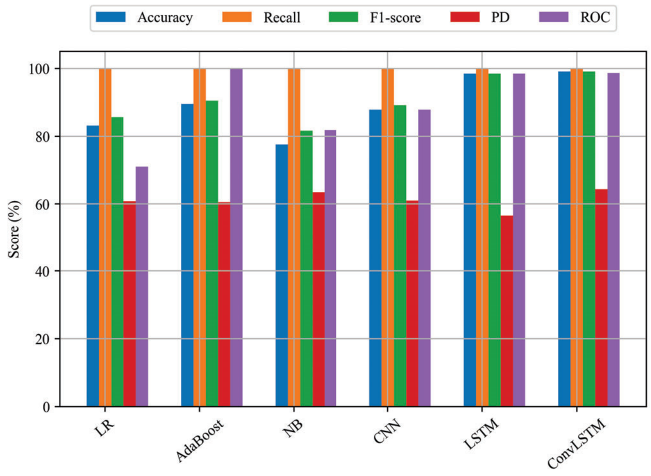

Figure 2: Evaluation metrics for the selected ML models.

IV. Conclusion

This paper presents a novel approach to detecting zero-day attacks in networked AS. Our approach combines CNN and LSTM algorithms, resulting in efficient and accurate zero-day attack detection. To evaluate its performance, we conducted simulations using a real-world dataset, measuring accuracy, recall, F1-score, PD, and ROC. The results demonstrate the high effectiveness of our approach-based detector in detecting zero-day attacks in networked AS.

Acknowledgment

This work was supported in part by the DoD Center of Excellence in AI and Machine Learning (CoE-AIML) at Howard University under Contract W911NF-20-2-0277 with the U.S. Army Research Laboratory, VMware Inc. Research Gift Funds and Microsoft Corp. Research Gift Funds.

References

[1] Y. Xiao, C. Xing, T. Zhang, and Z. Zhao, ''An intrusion detection model based on feature reduction and convolutional neural networks,'' *IEEE Access*, vol. 7, pp. 42 210--42 219, 2019.

[2] N. Moustafa, ''A new distributed architecture for evaluating ai-based security systems at the edge: Network ton_iot datasets,'' *Sustainable Cities and Society*, vol. 72, p. 102994, 2021.

[3] P. Mansourian, N. Zhang, A. Jaekel, and M. Kneppers, ''Deep learning-based anomaly detection for connected autonomous vehicles using spatiotemporal information,'' *IEEE Transactions on Intelligent Transportation Systems*, 2023.