Module Code: CSMRS16

Assignment Report Title: Literature Review and Critique of Findings

Student Number: 28812368

Date (when the work completed): 8-Nov-2021

Actual hrs spent for the assignment: 8 hours (excluding the weekly learnings and practical hours)

Assignment Evaluation:

- A Dual Learning Intrusion Detection System using CNN-DRL Algorithm for Cloud Computing
- Training and Testing with adequate samples with benchmark datasets / real-time simulation
- A six sigma standard accuracy results and less false positive rate

I. Introduction

In today's world, DATA has become an organization's most valuable and critical asset to the organization's business. Everything an organization does involves using data in some way or another. More and more data are being stored and processed on network-based computers, and as the enterprise data is growing exponentially, the majority of the organizations move away from traditional methods of computing and storing data to "Cloud storage" and "Cloud computing". In contrast, the on-premise data warehouse is not cost-effective, may not support advanced analytics solutions and is not capable of rapid information processing. By moving to Cloud computing, organizations can make data-driven business decisions proficient by leveraging the cost-benefit, flexibility, and advancement of analytical data support. Organizations need this level of flexibility to grow in a crowded market.

A vast majority of the enterprise workloads and data is already on the Cloud for various advantages like cost and rapid information processing; cybersecurity is growing day by day is being an organization's high risk and considerable concern. A Firewall only may not be able to defend Cloud computing against cybersecurity threats sufficiently. For example, it cannot detect insider attacks, either over a physical or virtual network. There are various challenges in Cloud computing, such as Data Breaches, Data Loss, Insecure Access Control Points, Denial of Service (DOS) and Distributed Denial of Service (DOS). A DOS attacks are the most common type of attack in Cloud computing. Attackers create colossal traffic to force the target routers and network to consume its bandwidth and resources with the virtually created network, thus causing overloading to prevent the server from serving or even shutting them down. Thus, the target system cannot provide services to its legitimate users, resulting in an authorized denial of service. Therefore, the second line of defence control after the firewall to detect cyber-attacks is required mandatory.

There is a constant need to design and implement a more accurate intrusion detection system and reduce false positives to prevent attacks efficiently on the network in a Cloud computing environment and protect Cloud data centres. As a result, there is enormous research in intrusion detection using deep learning that usually checks all incoming and outgoing packets of a particular network and determines whether each packet has signs of an intrusion. The intrusion detection can be considered two main categories based on operational logic: signature-based intrusion system (SIDS), where the network traffic is compared against known threats; anomaly-based intrusion system (AIDS), which inspects the traffic based on the behaviour of activities.

A well-designed Intrusion Detection System (IDS) can identify the characteristics of most intrusion activities and automatically respond to them. [1] Both CNN and DRL are the most representative neural networks in deep learning technology research. The CNN is a feed-forward neural network with profound structure and convolution calculation. Intrusion detection accuracy is determined using a binary and multiclass classification and 5-label classification identifying whether it is normal or any one of the other four attack types: Denial of Service (DOS), User to Root (U2R), Probe (Probing) and Root to Local (R2L). Deep Reinforcement Learning (DRL) is a universal framework for learning sequential decision-making tasks. A reinforcement learning problem is defined using a Markov decision process consisting of five entities, i.e., state, action, reward, policy, and value. Here, the agent learns its behaviour based on feedback from the environment and subsequently improve its action. The basis of solving the reinforcement learning problem is to find a policy, i.e. mapping from state to action.

II. Background

According to Cyberthreat Defense Report (CDR), 2021 by CyberEdge group, cyberattacks on a global network of large-scale enterprise organization has been increasing; a record 86% of organizations suffered from a successful cyberattack in 2021 (Figure 1), four out of five organizations over the globe prefer security products that feature machine learning (ML) and artificial intelligence (AI) technology.

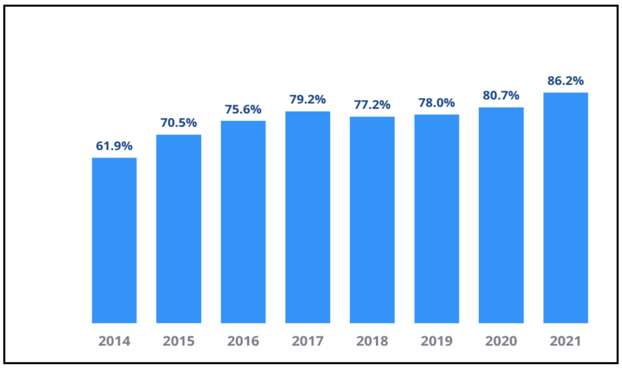


Fig. 1. Percentage of organization compromised by atleast one successful attack

In addition, the Forbes council have stated that Cloud adoption is already becoming mainstream. As a result, the usage of Cloud services across the globe is increasing, and 37.5% of UK organizations deliver their services via Cloud computing services (Figure 2), including servers, storage, databases, networking, software, analytics and intelligence - over the Internet ("the Cloud") and these services are provided to the economies of scale by large-scale enterprises pioneers such as Microsoft Azure, IBM, Google, and Amazon.

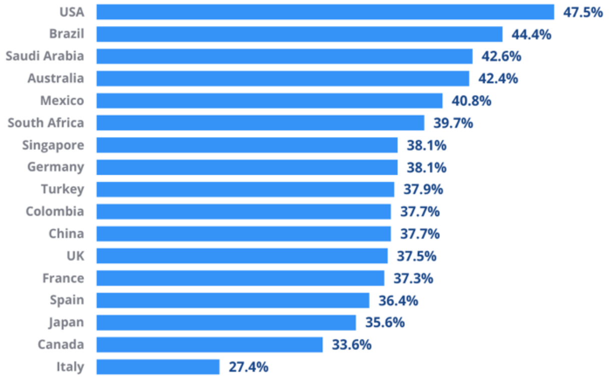
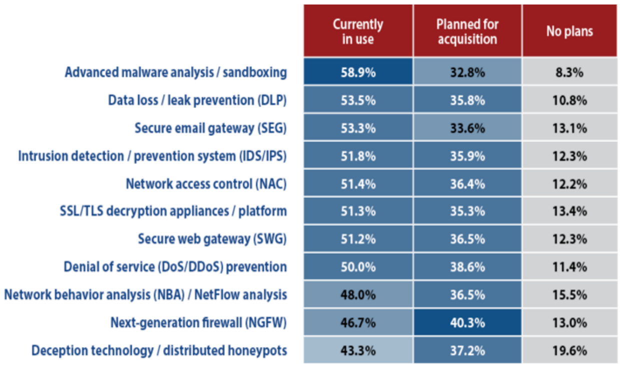


Fig. 2. Percentage of Secuity applications and services delivered via the Cloud, by Country

A significant percentage, 38.6%, of network security planned for acquisition in DoS / DDoS prevention (Table I) and the COVID-19 have brought major reprioritisation of new IT security investment by an organisation across the globe (Figure 3).



	Currently in use	Planned for acquisition	No plans
Advanced malware analysis / sandboxing	58.9%	32.8%	8.3%
Data loss / leak prevention (DLP)	53.5%	35.8%	10.8%
Secure email gateway (SEG)	53.3%	33.6%	13.1%
Intrusion detection / prevention system (IDS/IPS)	51.8%	35.9%	12.3%
Network access control (NAC)	51.4%	36.4%	12.2%
SSL/TLS decryption appliances / platform	51.3%	35.3%	13.4%
Secure web gateway (SWG)	51.2%	36.5%	12.3%
Denial of service (DoS/DDoS) prevention	50.0%	38.6%	11.4%
Network behavior analysis (NBA) / NetFlow analysis	48.0%	36.5%	15.5%
Next-generation firewall (NGFW)	46.7%	40.3%	13.0%
Deception technology / distributed honeypots	43.3%	37.2%	19.6%

TABLE I. NETWORK SECURITY TECHNOLOGIES IN USE AND PLANNED FOR ACQUISITION

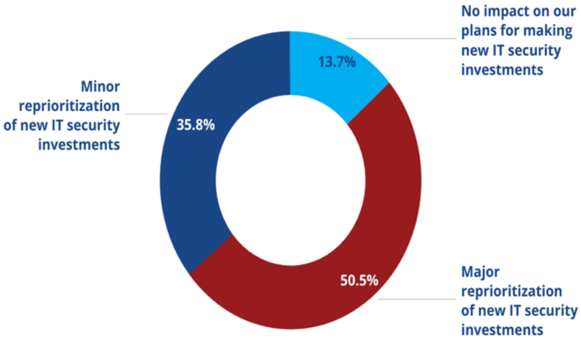


Fig. 3. Effects of the COVID-19 pandemic on IT security spending priorities

The study of Forbes Article, Cybersecurity Report, and the Survey [1] is a critical driver of choosing the research papers for the Literature Review and Critique Analysis on the various existing intrusion detection model in deep learning architectures and Cloud computing.

III. RESEARCH PROBLEM

The main problem of the DoS/DDoS attack is the debasement of network protocol. The Intrusion Detection System (IDS) in Cloud Computing presents various challenges such as identifying attacks from encrypted traffic, difficulty to analyze high-speed network, and challenges in detecting an insider attack and preventing attacks masked by evasion techniques. This coursework focuses on creating a dual model based on the combination of CNN-DRL, ensuring that not approved traffic prevents reaching the Cloud servers, studying the performance of binary and multiclass classification, and comparing the performance with other available models replicating real-time sample size. Further, the accuracy of detection shall be evaluated based on binary and multiclass classification using benchmark datasets such as CICIDS 2018, KDDcup99 and NLS-KDD datasets based on the following standard performance measures as

Accuracy: Accuracy measures how accurate the IDS is in detecting normal or anomalous traffic behaviour. Accuracy defined as the percentage of all those correctly predicted instances to all instances

- Accuracy = TP+TN/(TP+TN+FP+FN)

Actual Positive Rate (APR): APR is the ratio between correctly predicted intrusion and the total number of intrusions. If all intrusions and anomalies are detected, then the APR is 100% accurate. It is known as Detection Rate (DR). APR represented as

- APR=TP/(TP+FN)

False Positive Rate (FPR): FPR is the ratio between the number of legitimate instances falsely evaluated as an attack and the total number of legitimate instances

- FPR=FP/(FP+FN)

False Negative Rate (FNR): False-negative means when a detector fails to identify an anomaly and classify it as legitimate

- FNR=FN/(FN+TP)

IV. Literature review

A. Related Work

In recent years, numerous studies using machine learning and AI methods for intrusion detection have taken primary focus and surpassed traditional intrusion detection methods. Hizal Et al. [2] proposed a combined approach using a lightweight model based on convolutional and recurrent layers. The performance of the CNN-RNN based IDS model is tested on the NSL-KDD dataset, and the performance is evaluated using both binary and multiclass label classification and depicted an accuracy of 99.86%. Yin C Et al. [3] proposed another deep learning method called Recurrent Neural Network. The RNN-IDS model is analysed based on binary and multi-category classification on the NSL-KDD dataset, depicted the accuracy and detection rate with a low false-positive rate, and the performance is compared with other traditional classification methods such J48, native Bayesian, and random forest. The RNN-IDS model is also evaluated with various learning rates and hidden nodes. Xu C Et al. [4] propose a DNN model to identify the network intrusion detection and uses gated recurrent units (GRU) as the central memory unit along with multilayer perceptron (MLP) and SoftMax modules to increase the performance of intrusion detection systems. They use the well-known KDD and NSL-KDD datasets. The experiments were compared on LSTM and GRU, and according to the experimental results, the overall detection rate was 99.42% on KDD 99, and 99.31% on NSL-KDD and the detection rates for DOS attacks were 99.98% on KDD99 and 99.55% on NSL-KDD. According to the test results, it has been shown that the bidirectional GRU system has better results than LSTM for intrusion detection systems. Sethi K Et.al [5] proposed an intrusion detection using Deep Reinforcement Learning for Cloud Infrastructure using UNSW-NB15 dataset and evaluating higher accuracy and less false positive rate (FPR). There are five different classifiers, namely, Random Forest (RF), AdaBoost (ADB), Gaussian Naive Bayes (GNB), K-Nearest Neighbours (KNN), and Quadratic Discriminant Analysis (QDA) studied on UNSW-NB15 dataset, and the combination of classifiers C1, C2, C3 and C4 with various Q-threshold values is applied in obtaining a balance between high accuracy and less false positive rate. Adequate records samples are considered for measuring the performance (testing dataset having 1,75,341 records/samples, and training dataset having 82,332 records /samples).

B. Critique of Findings

[2] The accuracy and the detection rate in the CNN-RNN IDS model are produced with sample records of only 20% of the total size of the benchmark NSL-KDD data set. The research does not cover sufficient samples for all types of attacks. U2R performance results are abysmal as samples are considered very low (training = 52 samples, test = 200 samples). [3] The RNN-IDS Training model consumed 1765 seconds for 20 hidden nodes at the learning rate of 0.1, and epochs are 50 without GPU acceleration. The research does not depict which nodes take more time and which node takes less time. As the number of hidden nodes grows, the learning rate and the time taken to train the dataset increases; these are some of the limitations. [4] The DDN model based on GRU relies on theoretical verification and does not replicate real-time network data. The detection of R2L and U2R were not ideal as the sample of records considered for evaluation of these two types of attacks were too small. The smaller number of records makes the learning algorithm ineffective. In the KDD 99 dataset, the proportions of U2R and R2L were 0.01% and 0.23%, respectively. In the NSL-KDD dataset, as a revised version of KDD 99, the proportions of U2R and R2L increased to 0.04% and 0.79%, respectively. [5] The performance (high accuracy and less false positive rate) of the Deep Reinforcement Learning IDS model for Cloud Infrastructure is measured with the limitation to mandatorily use either GNB or QDA classifiers in every combination of classifiers such as C1, C2, C3 or C4.

V. Conclusion

A. Concluding Remark

The various research discussed in this coursework has several advantages and delimitations. First, the proposed models are tested with benchmark datasets; however not adequately tested and trained with a considerable size of samples/records for all types of classifications and networks with optimal nodes and network size. Second, the GPU acceleration is not considered [3] to improve the learning rate and detection accuracy.

The performance (high accuracy and less false positive rate) of various proposed models does not meet six sigma standards and does not address real-time complexity.

The various researcher considered in this critique review does not provide confidence if the models can be implemented, especially suitable for demanding applications in modern data networks that require a rapid response, high accuracy and reduce the false-positive rate.

B. Reflection (use "I")

To overcome the limitations of intrusion detections using various deep learning algorithms detailed in this literature review, a dual learning intrusion detection system for Cloud Security is proposed to optimally train and adequately test for all types of classification with optimal nodes, convolutional layers, and Deep-Q-Network (combination of CNN and DRL Algorithms) on well-known benchmark datasets such as UNSW-NB15, CICIDS 2018, KDDcup99 and NLS-KDD. Six Sigma standard intrusion detection is the goal of a new proposed dual model suitable for demanding applications in modern data networks that require a rapid response.

REFERENCES

[1] Jauro F, Chiroma H, Gital AY, Almutairi M, Shafi'i M A, Abawajy J H "Deep learning architectures in emerging Cloud computing architectures: Recent development, challenges and next research trend" - Applied Soft Computing. 2020 Nov 1; 96:106582.

[2] Hizal S, Cavusoglu U, Akgun D. "A new Deep Learning Based Intrusion Detection System for Cloud Security" In 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) 2021 Jun 11 (pp. 1-4). IEEE

[3] Yin C, Zhu Y, Fei J, He X. "A deep learning approach for intrusion detection using recurrent neural networks". IEEE Access. 2017 Oct 12;5:21954-61

[4] Xu C, Shen J, Du X, Zhang F. "An intrusion detection system using a deep neural network with gated recurrent units". IEEE Access. 2018 Aug 28;6:48697-707

[5] Sethi K, Kumar R, Prajapati N, Bera P. "Deep reinforcement learning based intrusion detection system for Cloud Infrastructure". In 2020 International Conference on Communication Systems & Networks (COMSNETS) 2020 Jan 7 (pp. 1-6). IEEE