A Seminar Report

On

"DEVELOPING THE SECURITY THREAT DETECTION MODEL FOR THE WEB SERVICE USING DEEP NEURAL NETWORK"

Submitted in partial fulfilment requirements for the award of the Degree

BACHELOR OF ENGINEERING

IN

INFORMATION SCIENCE AND ENGINEERING

By

H SWASTHIK SOMAYAJI 4NM18IS044

Under the Guidance of

Dr. Bola Sunil Kamath

Assistant Professor Gd III

Mr. Santhosh S

Assistant Professor Gd.II

Department of Information Science and Engineering

Department of Information Science and Engineering

NMAM Institute of Technology, Nitte 2021– 2022

NITTE EDUCATION TRUST | NMAM INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that **H SWASTHIK SOMAYAJI** , **4NM18IS044** , a bonafide student of NMAM Institute of Technology, Nitte has submitted the seminar report for the seminar entitled "**DEVELOPING THE SECURITY THREAT DETECTION MODEL FOR THE WEB SERVICE USING DEEP NEURAL NETWORK**" in partial fulfilment of the requirements for the award of Bachelor of Engineering in Information Science andEngineering during the year 2021-22. It is verified that all corrections / suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The technical seminar report has been approved as it satisfies the academic requirements in respect of seminar work prescribed by Bachelor of Engineering degree.

Signature of the guide	**Signature of the coordinator**	**Signature of the HOD**
Mr. Santhosh S	Dr Bola Sunil Kamath	Dr Karthik Pai

NITTE EDUCATION TRUST | NMAM INSTITUTE OF TECHNOLOGY

DECLARATION

We hereby declare that the entire work embodied in this Seminar report titled "**DEVELOPING THE SECURITY THREAT DETECTION MODEL FOR THE WEB SERVICE USING DEEP NEURAL NETWORK**" has been carried out by us at NMAM Institute of Technology, Nitte under the supervision of **Dr Bola Sunil Kamath** and **Mr. Santhosh S** for Bachelor of Engineering in Information Science and Engineering. This report has not been submitted to this or any other University for the award of any other degree.

H SWASTHIK SOMAYAJI(4NM18IS044)
Department of ISE, NMAMIT, Nitte

TABLE OF CONTENTS

ABSTRACT

The co-evolution of broadband networks and intelligent information system development ushers present golden days of web service. However, cyber attackers find loopholes easily for security threats under the web service environment. Detection of web service attacks requires to develop a new security threat counter-measure differentiated from the existing either signature-based or anomaly-based algorithms. In this regard, this research introduces a state-of-the-art intrusion detection model specialized in cyber-attacks on the web-service using the combination of deep neural network algorithms. Then, we evaluate the intrusion detection performance of the proposed deep neural network models with the Big Data of real time network traffic. Our research helps improve the limitation of existing intrusion detection systems and to overcome web service vulnerability on cyber threats..

LIST OF FIGURES

1.INTRODUCTION

The anomaly detection approach, as an alternative, detects an attack or an abnormal behavior by judging the traffic deviating substantially from a rule for a predetermined normal behavior. Although anomaly detection model is an effective method for discovering new attack types, it is impossible to predefine all the normal activity rules and network protocols [12]. In order to overcome this problem, previous studies have used machine-learning techniques. Recently deep neural network-based studies have also been conducted [5-7] however; they found it challenging to obtain Big Data for model learning. Further improvement in the processing speed and accuracy is required for real-time intrusion detection. In addition, those researches concentrate on the specific types of attacks, such as Distributed Denial of Service (DDoS) or information system scanning, which occurs mainly in a lower (network) layer of the TCP/IP model. In this regard, this research introduces the deep neural network-based intrusion detection models to deal with the threats on the application layer, which are difficult to detect due to the complex syntax of HTTP (Hypertext Transfer Protocol) of the web service protocol. To achieve the research objectives, we collected the big data of real-time network traffic from the web service server farm, and propose an intrusion detection method for a web service application to identify security threats that bypass the signature-based security systems. This research showed that the deep neural network technique provided excellent performance for the detection of web application intrusion that are not detected by signature-based intrusion detection system.

Unlike connection-oriented Internet services such as telnet, FTP, and e-mail, web services are connectionless open services. Even though web services require user authentication through login, most of the web services also have service pages for outsiders or non-members. In addition, web services have complex systems with hierarchical architecture [8]. A large number of web application programs that use scripting languages embedded in HTML, PHP, ASP, and JAVA are connected to the database and retrieve stored data on the website. Web contents may contain privacy such as a credit card number and personal information [9]. An attacker attempts to manipulate, destroy, or leak information without access authorization using a variety of web scripting languages.

1.PROPOSED TECHNOLOGY

2.1 Intrusion detection analysis technique

The Intrusion Detection System (IDS) is designed to detect malicious activities that may threaten the reliability and security of computer systems [10]. The existing IDS has either the signature-based analysis or anomaly-based analysis as the intrusion detection method [11]. The signature-based analysis technique finds a specific pattern of a known attack threat to analyze the list of already stored signatures by comparing the corresponding string with a regular expression. This technique is very successful if it keeps the database of signature patterns up to date, but it cannot detect unknown attacks or new malware such as zero-day attacks [11]. Compared to the signature-based analysis technique, abnormal-based analysis technique can detect attacks deviated from standard traffic patterns even for attacks whose signatures are not defined, and quickly finds new types of attacks [12]. However, the performance of an anomalybased detection system depends on how well it is executed and tested on all protocols. Since the definition of the standard traffic patterns variable according to the protocols generated by specific vendor, it is challenging to define detection rules and to describe protocols [12].

2.2 Trends in research on AI-based intrusion detection

Recently, studies on artificial intelligence algorithms have been actively conducted to solve the problems of both signature-based and anomaly based intrusion detection system

2.2.1 Machine Learning

Machine learning based methods enable the detection of new and subtle attacks occurring at the moment without extensive human-oriented inspection or intervention[2]. As shown in Table 1 representative machine learning models include Decision Tree, Bayesian network, SVM (Support Vector Machine), GA (Genetic Algorithm), and KNN (k-nearest neighbor). Although studies above showed more than 90% accuracy, but most of them used test data sets or designed to detect already known specific type of attacks

3.IMPLEMENTATION

The deep neural network-based intrusion detection systems have attracted considerable interest both in industry and in academia. Kim et al. (2016) applied the LSTM (Long-term and Short-term Memories) architecture to the RNN and trained the intrusion detection system using the KDDCup'99 dataset [7]. Compared to other intrusion detection classifiers, LSTM-RNN achieved an accuracy of 96.93% and a detection rate of 98.99%. Notably, a recent study by Arnaldo et al. (2017) compared the network intrusion detection performances of FFNN with RNN (LSTM), and CNN models by training them with the log data collected from a corporate security system. However, their study only focused on the network intrusion detection through the attributes of lower layers (IP address, etc.) of the TCP/IP model [13-15]. Unlike to network layer, web servers and applications are very complex systems, which increase the probability that vulnerability exists and makes it challenging to detect cyber threats. Also, a desirable intrusion detection system for web service needs to process noisy data with a high computation speed and accuracy since the noise level in a data set increases with a data set size [12]. Single technique has a limit to obtaining high performance considering these issues on the web service attack. To deal with these issues, it needs to compare and analyze the intrusion detection performances of hybrid deep neural network models with unstructured letter and number-based syntax structures

3.1RESEARCH METHOD

3.1.1 Data collection and preprocessing

3.1.1.1 The Architecture for Network Traffic Collection

We collected and preprocessed real-time network traffic that had flown into the public website of the NEC (National Election Commission) in Korea. Figure 1 is a conceptual diagram of the network for building the data set. The firewall plays the role of primary access control for IP and the transport protocol (TCP, UDP, etc.), which is the third and fourth layer respectively. Is also responsible for defense against DDoS attacks that overload the homepage server. Traffic that passes through the firewall undergoes the second access control or the IPS (Intrusion Prevention System). Based on pre-defined detection rules, IPS detects and blocks intrusion threats such as the inflow of malicious code, abnormal protocol, and a DDoS attack.

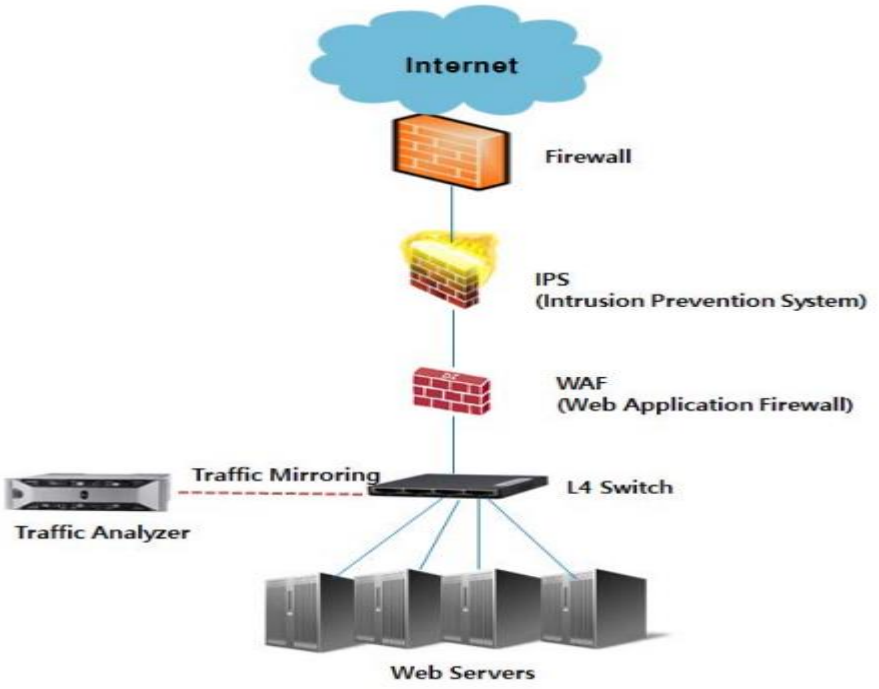

Figure 1: The architecture for network traffic collection

3.1.1.2 Data collection and classification

We collect the traffic from 5 pm to 8 pm on May 5, 2017, and the amount of the collected traffic data reached about 13 Terabyte. All network traffic collected in real time is raw packet data. Of 14,215 data records in total, 13,942 records were classified as normal behaviors and 273 records as attacks. After 90% of the dataset was extracted as the learning set and the remaining 10% as the validation set.

Additive attributes that provide status information, such as the data size and messages, during the process of transmitting and receiving data between the client and the web server. They are not directly related to intrusion detection, but they are helpful when detailed traffic analysis is needed.

3.2 Intrusion Detection System development

Figure 3 shows the analysis model of the deep neural network-based intrusion detection system. System components largely divided into the traffic classifier, preprocessor, and intrusion detector. Bro 2.5, an open-source intrusion detection platform, was used as the traffic classifier. It collects all real-time traffic of the target to be protected, removes unnecessary information from it, and classifies the traffic data according to each service protocol. The preprocessor processes data through modification, deletion, and addition of the result data of the traffic classifier, extracts security threat information and constructs the input data set to the intrusion detector.

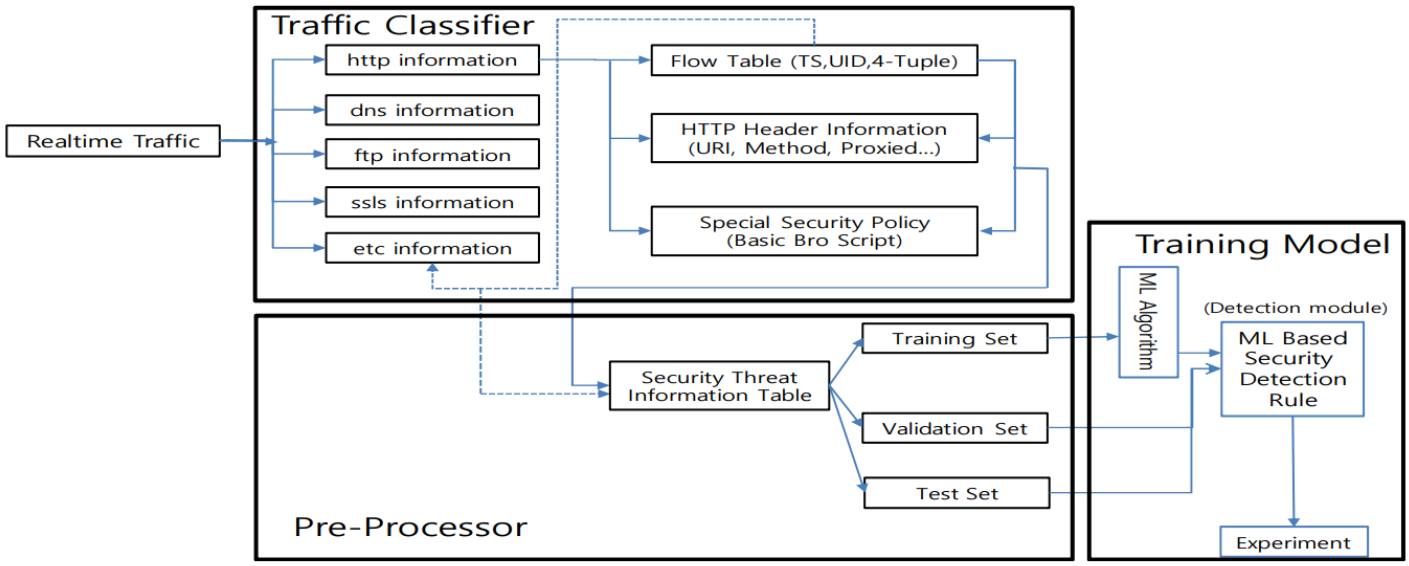

Figure 3: The data analysis model of Deep Neural Network

3.3 Training Model for Intrusion Detector

The syntax of the text format (HTTP header message information) in the dataset converted into vector values through data preprocessing. After the vector input values passed through the models of LSTM-RNN, CNN, and CLSTM, we compare each result.

The comparison with machine learning models such as Decision Tree, SVM, etc. was excluded because the size of input data was variable depending on the length of the phrase or the number of words. Instead, it is a reasonable approach to compare deep neural network model in that the value of an input variable is not meaningful but the whole meaning of the text phrase must be mechanically understood.

4.ANALYSIS

The target variable has the binary classification system that categorizes a normal behavior as ‘0’ and an attack as ‘1’. We used precision, recall, and F-score as well as the ROC curve as the indices using the confusion matrix that is used to evaluate intrusion detection performance. The results of the detection performance of each model based on the evaluation indices. The analysis results were obtained by dividing the total of 14,215 data records into the sets of 256 input records (batch size = 256) to be learned at a time. This process was repeated 20 times (epoch = 20), and the results of each step (the total number of steps = 989) were averaged. As shown by the results, the LSTM model showed better overall performance than the other models, with a recall of 0.966, an accuracy of 0.997, and an F1 score of 0.898. In terms of precision, the CNN model showed excellent performance with 0.899 precision. On the other hand, the overall performance of CLSTM in recall, accuracy, and precision was lower compared to other models. Table 5 shows the ROC curve, which represents the accuracy and loss of each step for each model. CNN, LSTM-RNN, and C-LSTM models showed the excellent performance for the intrusion detection in the web service envrironments.

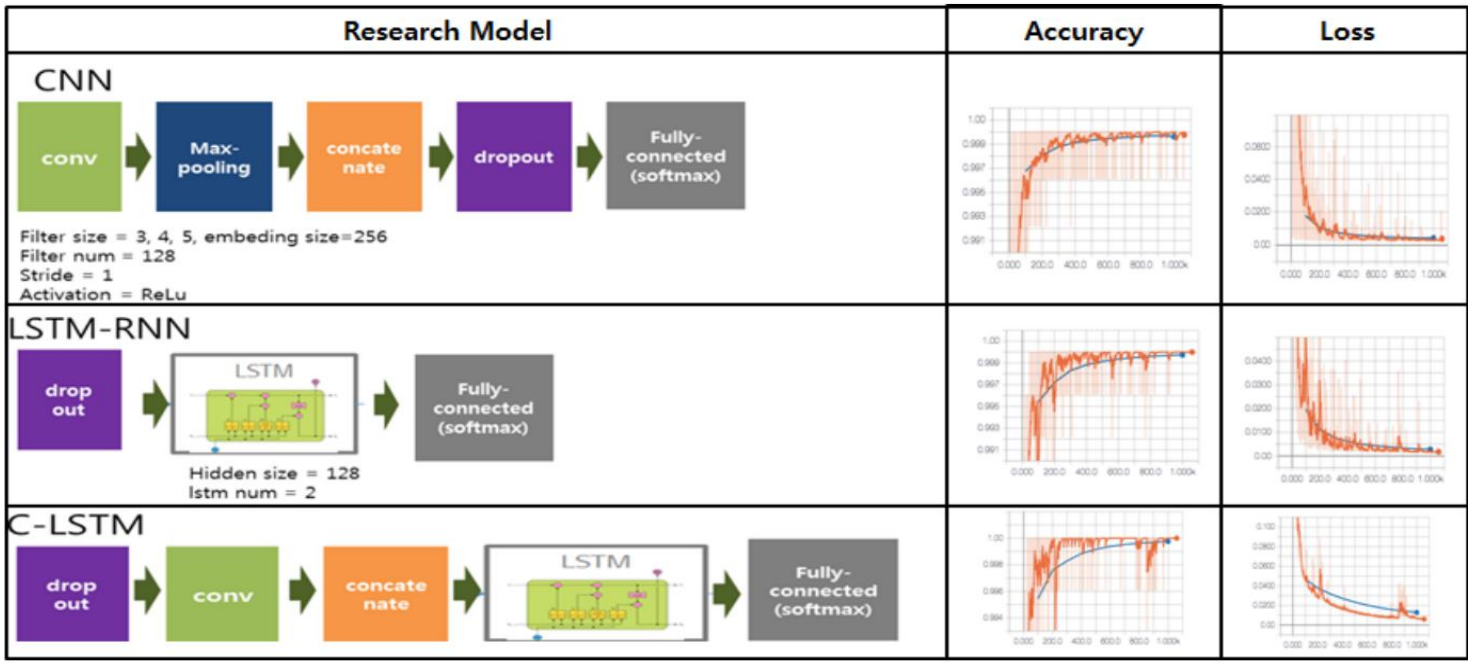

Figure 4: Research Model for Intrusion Detection

5.CONCLUSION

Through the real-time website traffic data analysis, this study showed that it is possible to conduct a big data collection as well as analysis in the presentation and application layer in the TCP/IP protocol. In addition, it was also shown that the deep neural network technique, which has an excellent performance in the ranking of images or sentences, also shows excellent performance for the detection of web application intrusions, which are not possible, by signature-based intrusion detection systems. Signature-based technique have not been able to detect all types of attacks if the signature list of intrusion detection systems did not contain the right signature. To the best of our knowledge, the hybrid deep neural network model of CNN and LSTM has not been introduced in the field of web service intrusion detection but has proven to be an excellent classifier. Thus, there is no need to stack multiple hidden layers in the deep neural networks, so the burden of processing performance is low, which makes it possible to develop and commercialize an intrusion detection system with excellent performance. However, since it was not possible to control all of the variables in the model training process due to the nature of artificial neural networks, there is a possibility that different results will be obtained depending on the operating environment of the web service and the experimental setting. Therefore, in order to generalize the results of this study, there is a need to verify them using the bigger size of web service data.

REFERENCES

[1] Storey, V.C. and I.-Y. Song, Big data technologies and management: What conceptual modeling can do. Data & Knowledge Engineering, 2017. 108: p. 50-67.

[2] Hodo, E., et al., Shallow and deep networks intrusion detection system: A taxonomy and survey. arXiv preprint arXiv:1701.02145, 2017.

[3] Nadav Avital, N.E. The State of Web Application Vulnerabilities in 2017. 2017 [cited 2018 09-28]; Available from: https://www.imperva.com/blog/2017/12/thestate-of-web-application-vulnerabilities-in2017.

4] Noel, S., D. Wijesekera, and C. Youman, Modern intrusion detection, data mining, and degrees of attack guilt, in Applications of data mining in computer security. 2002, Springer. p. 1-31.

[5] Fiore, U., et al., Network anomaly detection with the restricted Boltzmann machine. Neurocomputing, 2013. 122: p. 13-23.

[6] Gao, N., et al. An intrusion detection model based on deep belief networks. in Advanced Cloud and Big Data (CBD), 2014 Second International Conference on. 2014. IEEE.

[7] Kim, J., et al. Long short-term memory recurrent neural network classifier for intrusion detection. in Platform Technology and Service (PlatCon), 2016 International Conference on. 2016. IEEE.

[8] Yongho Lee, et al., Counter-measure development by web application threat analysis for safe web service, Korea Institute of Information Security and Cryptology, 2004. 14(4): p. 1-9.