Towards Network Traffic Monitoring Using Deep Transfer Learning

Bindushree K P , Anusha K H, Chethana jain R , Nisarga L, Girish L

Department of Computer Science & Engineering, Visveswaraya Technological University,
Channabasaveshwara Institute of Technology, Gubbi, Tumkur,
Karnataka, India
bindushreekp5@gmail.com, nisargamoulya@gmail.com

Abstract- **Organization traffic is developing at a dominated speed internationally. The cutting edge network foundation makes exemplary organization interruption discovery strategies wasteful to characterize an inflow of tremendous organization traffic. This paper means to introduce an advanced methodology towards building an organization interruption discovery framework (NIDS) by utilizing different profound learning techniques. To additionally work on our proposed conspire what's more, make it compelling in genuine settings, we utilize profound exchange learning methods where we move the information learned by our model in a source area with ample computational and information assets to an objective space with inadequate accessibility of both the assets. Our proposed technique accomplished 98.30% arrangement exactness score in the source space and a worked on 98.43% grouping exactness score in the objective area with a lift in the characterization speed utilizing UNSW-15 dataset. This examination exhibits that profound move learning procedures make it conceivable to build huge profound learning models to perform network characterization, execution and further develop their arrangement speed in spite of the restricted availability of assets.**

1. INTRODUCTION

A computer that is able to be reached over a network is called a host. This can either be a client, server or any other type of computer however, a network is a collection of computers, servers, mainframes, network devices or other devices connected to one another to allow the sharing of data.

Information and communications technology (ICT) systems and networks handle various sensitive user data that are harmed by various attacks from both external and internal intruders [1]. These attacks can be both manual and machine generated and are gradually advancing in confusions resulting in undetected data violations. Cyber attacks are continuously evolving with highly complicated algorithms with the development of software, hardware, and network topologies including the latest developments in the Internet of Things (IOT) [2]. Spiteful cyber-attacks that constitute a crucial security issues that demand the need for a novel, flexible and more trust worthy intrusion detection system (IDS). An IDS is a bold intrusion detection tool which is used to classify and detect intrusions, attacks, or violations of the security policies accordingly at network and host level infrastructure without delay. Intrusion detection is classified into network-based intrusion detection system (NIDS) and host-based intrusion detection system (HIDS) [3] based on intrusive behaviors. Network behavior used IDS is called as NIDS. These network behaviors are collected using network equipment via mirroring by networking devices, such as switches, routers, and network taps and analysed in order to identify attacks and possible threats hidden within in network traffic. Program function used IDS which is in the form of various log files that run on a local computer to identify an attack is called HIDS. It also checks the file system status or ram it contains data which we expected or not. Local sensors collects log files, Where NIDS scans the contents of each package for network flow, HIDS relies on log information that includes sensor logs, system logs, software logs, file systems, disk resources, user account details. Many companies use a combination of both NIDS and HIDS. Network traffic flow analysis was performed using unmodified detection, anomaly detection and explicit protocol analysis. Incorrect detection uses previously defined signatures and filters to detect attacks. It depends on personal input to update the signature database regularly. This method is unambiguous for detecting attacks which is already known but does not work perfectly in the case of attacks which is unknown. Anomaly detection uses problem solving methods for detecting unknown dangerous activities. In many cases, anomaly detection reveals a positive false positive result [3]. To remedy this issue, many companies use combination of both abuse and misuse in their trading solutions. Formal protocol analysis is more capable compared to visual cues stated because the precise protocol analysis works in the network layer, application layer and transport layer. This make use of pre-defined vendor specification settings to detect deviations from appropriate agreements with applications. Although depth learning methods are recently considered to increase the creativity of such entry-level techniques, there is a lack of study to measure such electronic reading skills with publicly available databases. The common issues in existing solutions based on machine learning models are: first, models produce a high level of false positives [4], [3] with a wide variety of attacks; second, the models cannot be generalized as existing studies have used mainly one database to report the performance of the machine learning model; third, the models studied so far did not fully reflect the current large network traffic; and finally solutions are needed to withstand network size of rapidly growing speed, speed and power. These challenges for this work form the main motivation by focusing on the research of various classical classifiers and deep neural networks (DNNs) used in NIDS and HIDS.

In the current world, rapid technological advances have experience the entire companies to embrace the combination of information and communication technology (ICT). If the security of the ICT system is compromised, an environment is created where all actions are delivered in such a way that the organization is at risk. So, this is a multi-line detection and security system that could manage a novel attack of the system and be able to adapt independently to new data.

There are so many programs that can be put to use to protect such ICT systems from danger, namely anonymous detection and IDS. The failure of uncontrolled detection systems is the difficulty involved in the process of interpreting the rules. Every protocol examined should be defined, used and tested for accuracy. Another mud associated with misdiagnosis is that hazardous activity that falls into a normal pattern of use is not detected, so the need of an IDS that can be adapt to the latest novel attacks and can be trained and deployed using informal distribution data sets is needed. Intrusion Detect Systems (IDS) has a variety of technologies based on cyber security originally designed to detect exposure and deed on the target person. The use of an IDS is to only discover threats. It is therefore available without a band on the network infrastructure and is not in a real-time communication location between the sender and the data receiver. Rather, solutions often use TAP or SPAN ports to analyze a copy of a traffic line and will attempt to predict an attack based on a pre-trained algorithm, which is why making the need for human intervention less meaningful [21]. The rest of the paper is organized as follows: 2. Related work, 3. Methodology, 4. Model, 5. Dataset, 6. Conclusion

2. RELATED WORK

Since the birth of computer systems, research on safety issues related to NIDS and HIDS has been around. In current days, the use of machine-based learning solutions to NIDS and HIDS has become important for safety researchers and specialists. A detailed study of the available solutions based on machine learning is discussed in more detail at [3]. This section discusses the panorama of the largest research to date that explores the field of machine learning and the in-depth learning methods used to develop NIDS and HIDS.

The use of ML techniques and solutions in complete IDS has become commonplace in recent days, but the training details on hand are limited and used for benchmark purposes only. DARPA datasets [26], are one of the most complete data sets available to the public. The tepdump data provided by the 1998 DARPA ID Evaluation network of 1998 was refined and used for the 1999 KDD Cup contest at the 5th International Conference on Information Access and Data Mining. The task was to edit the connection records already processed in standard traffic, or in one of the following categories of attacks: 'DoS', 'Probing', 'R2L' and 'U2R'. The main reason for the popularity of ML-based methods is because of its ability to withstand flexible and varied threats to obtain an acceptable false standard of ID at the appropriate cost of calculation. In the first sections, [27] he used the PN rule method found in P rules and N rules to determine the presence and absence of a class respectively. This is beneficial due to the improvement in the detection rate of other types of attacks outside the U2R category. In addition to the promotion of traditional feed forward networks (FFN) in the biodiversity aircraft, there is a network called the Convolutional Neural Network (CNN). In the early stages, CNN was used to process images using standard 2D layers, combining 2D layers and fully integrated layers[28]. IDS studies with KDDCup's'99 'database and compared the results with many other bleeding algorithms. After extensive analysis, they now hold CNN's superiority over other algorithms. A study of the use of the Long Short-Term Memory (LSTM) separator was performed by [4] with the same database. It has been said that because of LSTM's ability to see past and present consecutive connecting records it is useful for intrusion detection system.

In the real world, the Intrusion Detection System (IDS) plays a key role in detecting intrusion. There are two types of detection, that is anomaly-based and misuse-based [29] [30]. To find different functions in the established patterns of users, we use the uncontrolled based. Over the last few years, probably the use of the Intrusion Detection System is used to strike an existing pattern to detect it. Soft computing is one of the strategies that helps reduce costs. There are few soft computer simulations on IDS. For example Artificial Neural Network, decision tree, mysterious concept. It is used to build resources in the field of access to access due to learning and flexibility. Through the soft computer techniques, the neural network approach is popularly used in modern research. Using SOM, Haywood et al. elevated the hierarchical neural network to detect intrusion [31]. The feed forward neural network is used to create IDS using the Back Propagation algorithm training, suggested by J. Shum et al. [32]. Mukkamala et al. published an alternative to the interaction between the neural network and the SVM [33]. Another method that has transformed the recurrent Jordan neural network is presented by Xue et al. [34]. It also successfully installed a recurrent Jordan neural network to expose SQL-based attacks [35]. Although the neural network is quite good to apply in this field, in-depth learning is another way to gain the accuracy of getting better than previous methods. In 2015, our study used Recurrent Neural Network and Hessian-Free Optimization to train a DARPA data set [36]. We found a 95.37% acquisition rate. We continue to use another method in deep learning to detect modern attacks and malwares. For this function, we use Long Short Term Memory and Recurrent Neural Networks in IDS.

DATA NETWORK INTRUSION DETECTION

An open database, we use UNSW-NB15 database, which it is a database for broadband network access. UNSW-NB15 is designed for limited testing of NIDS [37]. In particular, this was intended to put back the KDD Cup 99 and NSL-KDD data sets, which have been famous NIDS databases throughout the years, but don't reflect the freshly developed network behaviors and this was described in [37], in order to demonstrate modern hacking behavior, attacks at UNSW-NB15 were carried out and make use of the IXIA Perfect Storm, which could mimic a major attack on the CVE website. After setting up the test site in the company of attack generator, the traffic was replaced by the TCP dump. After that the ending database is created by running the feature with tools like Bro and Argus Gown et al.:

k-nearest neighbour, and help Vector Machine [38]. During testing, the J48 and K-NN algorithms were introduced as the most-worthy models for high efficiency and accuracy. Moustafa et al. An anomaly-based detection method was developed based upon geometrical analysis using trapezoidal spatial measurements [37]. At the same time, Papamarztivanos et al. has introduce a novel approach to NIDS in the company of a genetic algorithm and a decision-making drug [39]. In their job, they have used a genetic algorithm to generate discovery order that form the model of deciduous trees. The developed model was tested by UNSW-NB15, and showed major work in detecting all common and unusual attacks on the database. Recently, Tama et al. to test the effectiveness of deep neural networks (DNNs) NIDS in UNSW-NB15 [40]. In addition, VinayaKumar et al. a comparative analysis was performed of DNN models and a classical algorithm for machine learning [41]. After conducting a thorough performance parameter search, they finished that DNNs were worthy for IDS development.

3. METHODOLOGY

The Deep neural networks(DNNs)arises the Artificial Neural Network (ANN) with many structures built within the input layers.
Below we cover about simple DNNs and the use of Linear Rectified Units and why it is present over additional activation functions.
1.The Deep Neural Network (DNN)
2.Uses of rectified linear units
3.Data network intrusion detection
4.Method network Intrusion Detection

1. The Deep Neural Network(DNN)

The Deep neural networks are subjected to growing forces of gravity and extraction. The input layer is detected by the input layer and transferred to the first hidden layer. Such hidden layers make up the numerical properties of one's input.

2. Uses of rectified linear units

Rectified Units are very effective and have the power to speed up the whole training process completely. Typically, Neural networks use the function of sigmoidal activation. these functions are usually the end of the gradient problem. Gradient disappearance take place when the lower layers of DNN have almost empty gradients due to the upper layer units are almost filled with asymptotes of tanh activity.

3. Data network intrusion detection

An open database, we use the UNSW-NB15 database, which it is a database for broadband network access. UNSW-NB15 is designed for limited testing of NIDS [44]. Aims to put back the KDD Cup 99 data and NSL-KDD, which have been famous NIDS databases throughout years, but does not communicate the latest trends in network attacks.

4. Method network Intrusion Detection

The IDS carries two ways of finding out just as the definitions of risky activity[43]. The signature-based detection method describes malicious activities, and detects behavioral-like behavior. In contrast, the detection-based method of detection defines normal function, and detects deviant behavior.

DNNs are used to forecast attacks on the Network Intrusion Detection System (N-IDS). DNN along with a reading rate of 0.1 is used and operated with 1000 epochs and the KDDCup-'99 'database is used for training and network marking.

Long Short-Term Memory (LSTM) is a structure developed by Hochreiter and Schmidhuber[42].

The recurrent Neural Network is the most popular model for training data sequencing. Standard RNN has a problem when it is utilized for training in step-by-step size. In this part, we briefly talk over the RNN system and the vanishing issue. After that, we define Short Memory to deal with this issue.

The diagram shows one LSTM cell. We also define statistics to calculate three gate values and cell status.

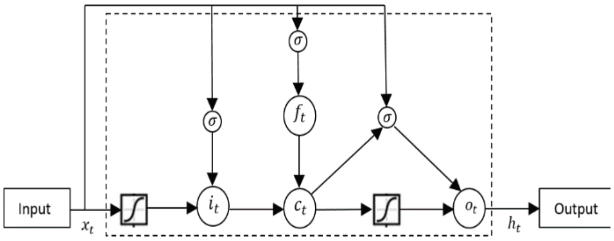


Fig 3.1: LSTM cell

The logical sigmoid function σ, and f, I, o and c respectively the, forget gate, input gate, output gate and cell state. Wci, Wcf and Wco are described as peephole weight matrics communication. By utilizing LSTM, we can answer the problems of extinction including explosion of the gradient due to three gates. In the LSTM-RNN formation, a hidden duplicate layer is restore by LSTM cell.

3.2 EXPERIMENT

At this stage, we take two tests. First test is about finding the highest parameter values to determine the best performance of the IDS model. Second test is about measuring performance by the parameter values obtained from first test.

Typically, False Alarm Rate (FAR) and Detection Rate (DR) are used as IDS test ratings.DR shows the rate of entry conditions obtained by the IDS model. FAR is a measure of standard conditions that are not well defined. Depending on the matrix of confusion(TP-True positive, TN-True negative, FP-False positive, FP-False negative)

$DR = TP / (FN+TP)$

$FAR = FP / (FP+TN)$

As FAR decreases and DR increases, performance improves. So, we utilize one metric, for efficiency. Utilizing this metric, we can easily analyze the IDS model.

Efficiency = DRFAR

We F1-score (F1) and accuracy (AC), AC and F1 calculated respectively

$AC = TN + TP / TP + TN + FP + FN$ $F1 = 2P * R / P + R$

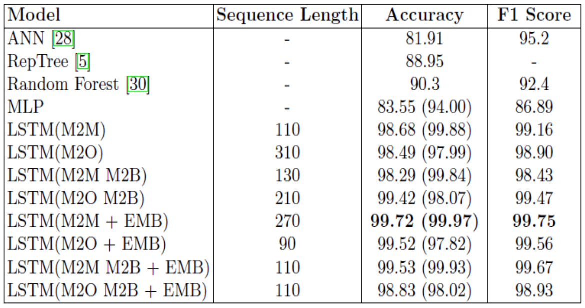
where P and R stand with precision and memory

$P = TP / TP + FP$ $R = TP / TP + FN$

We are testing multiple combinations of training preparation in LSTM for inclusion. initial, the LSTM model is trained in two ways as report above. One learns from the mistake of all output (M2M) and the other only reads from the mistake of the final output (M2O).

Moreover, in binary categories, we include `multiple division in binary division '(M2B) which trains the multi-division model and change all negative labels and model effects into the same label` attack'. Finally, embedded input (EMB) is applicable to all models.

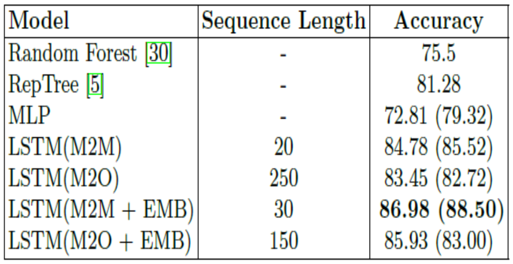
- Binary-classification result for LSTM Model. Verification solution are in parenthesis.



Model	Sequence Length	Accuracy	F1 Score
ANN [28]	-	81.91	95.2
RepTree [5]	-	88.95	-
Random Forest [30]	-	90.3	92.4
MLP	-	83.55 (94.00)	86.89
LSTM(M2M)	110	98.68 (99.88)	99.16
LSTM(M2O)	310	98.49 (97.99)	98.90
LSTM(M2M M2B)	130	98.29 (99.84)	98.43
LSTM(M2O M2B)	210	99.42 (98.07)	99.47
LSTM(M2M + EMB)	270	**99.72 (99.97)**	**99.75**
LSTM(M2O + EMB)	90	99.52 (97.82)	99.56
LSTM(M2M M2B + EMB)	110	99.53 (99.93)	99.67
LSTM(M2O M2B + EMB)	110	98.83 (98.02)	98.93

Fig 3.2.1: Binary Classification table

- Multi-classification result for LSTM Model.



Model	Sequence Length	Accuracy
Random Forest [30]	-	75.5
RepTree [5]	-	81.28
MLP	-	72.81 (79.32)
LSTM(M2M)	20	84.78 (85.52)
LSTM(M2O)	250	83.45 (82.72)
LSTM(M2M + EMB)	30	**86.98 (88.50)**
LSTM(M2O + EMB)	150	85.93 (83.00)

Verification solution are in parenthesis.

Fig 3.2.2 Multiple Classification table

72.81% and 83.55% multiple divisions and binary options, respectively. The F1 rate for a binary case is 86.89%. LSTM models show more than 98% accuracy in binary categories and 83% in most categories. Moreover, our LSTM models outperform the previous works [45,46,47]

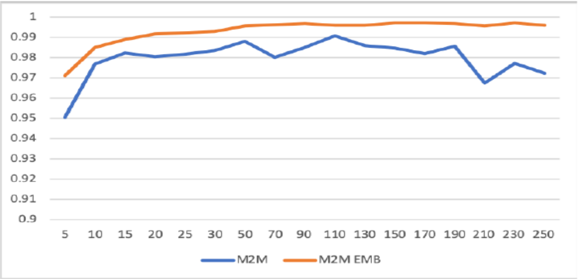


Fig 3.2.3: Classical accuracy graphs

Classical accuracy graphs for verification data: M2M embedding and M2M. The horizontal axis shows the length of the sequence. Among the LSTM models, theEMB+M2M model reach the highest performance of all binary and multi-segment functions.

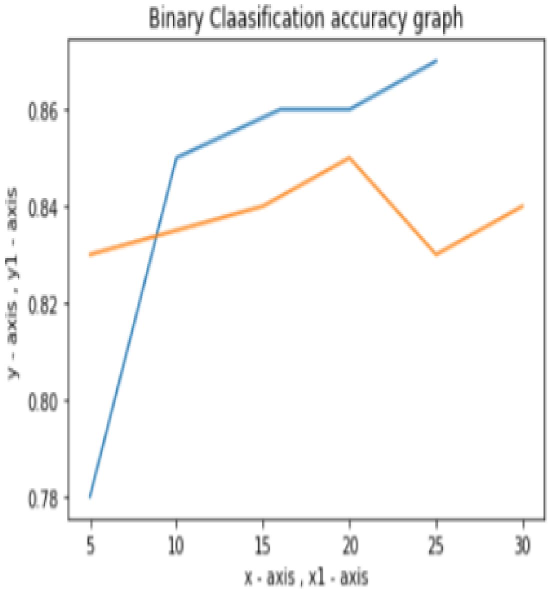


Fig 3.2.4: Multiple-category graphs

Multi-category graphs for verification data: M2M embedding and M2M. The horizontal axis shows the length of the sequence. Moreover, in binary categories, M2B could be used, but it does not make a significant impact on performance. The solution for the non-M2B and M2B models are almost identical.

For practical consideration, we examined the predictive time of different sequence lengths in the model where we can determine if the predictive time corresponds to the sequence length.

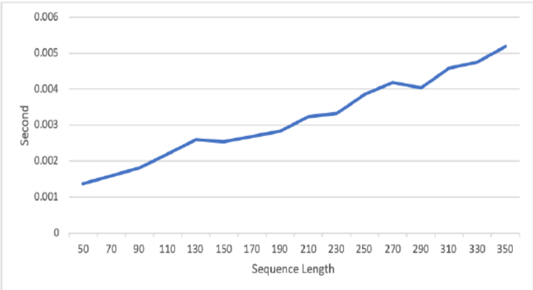


Fig 3.2.5: predictability time in seconds

Predictability time in seconds for each sequence along with different sequence lengths.

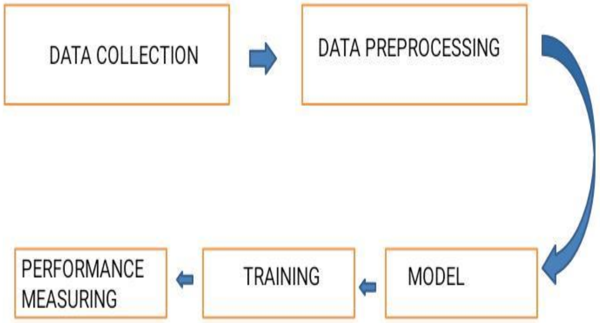


Fig 3.2.6: KDD modelling process

Data collection isa process ofMeasuring and gathering details on targeted variable in established system. It is utilized to perform research components in all study area including business social science etc.

Data preprocessing is necessary step in data mining method which is utilize to convey the raw data in a useful and efficient format.

MODEL

Proposed Architecture:

An outline of proposed DNNs engineering for all utilization cases is appeared in Fig. 1. This includes a covered up layer tally of 5 and a yield layer. The information layer comprises of 41 neurons. The neurons in input-layer to covered up layer and covered up to yield layer are associated totally. Back-engendering component is utilized to prepare the DNN organizations. The proposed network is made out of completely associated layers, inclination layers and dropout layers to make the organization more hearty.

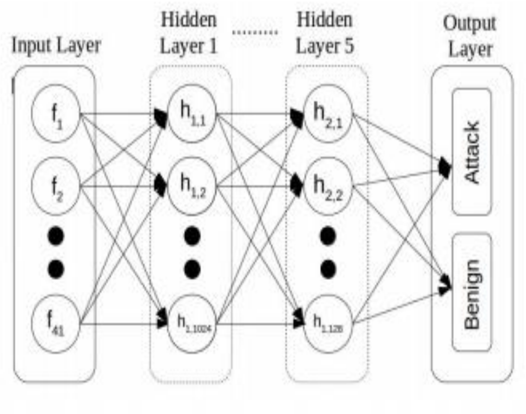


Fig. 1. Proposed architecture

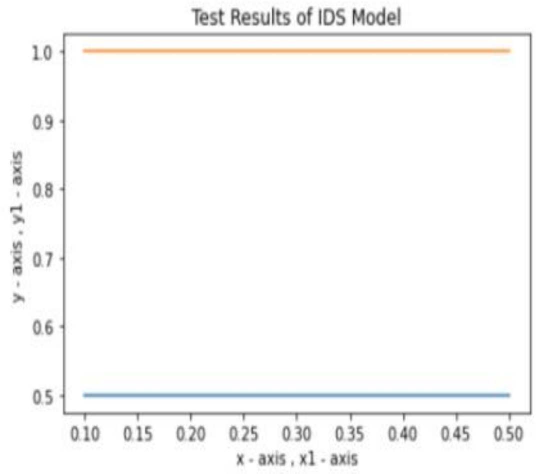
Info and covered up layers: This layer comprises of 41 neurons. Secret layers use ReLU as the non-direct initiation work. At that point loads are added to take care of them forward to the following secret layer. The neuron include in each secret layer is diminished consistently from the first to the yield to make the yields more precise and simultaneously decreasing the computational expense.

Regularization: To make the entire cycle effective and efficient, Dropout (0.01). The capacity of the dropout is to unplug the neurons arbitrarily, making the model more vigorous and consequently keeping it from over-fitting the preparation set.

Yield layer and grouping: The out layer comprises just of two neurons Attack and Benign. Since the 1024 neurons from the past layer should be changed over into only 2 neurons, a sigmoid actuation work is utilized. Because of the idea of the sigmoid capacity, it returns just two yields, henceforth preferring the twofold order that was planned in this paper.

The IDS model set up:

We standardized all occurrences from 0 to 1. The info vector is 41 highlights and the yield vector is made out of 4 assaults and 1 non assault. Along these lines, the information measurement is 41 and the yield measurement is 5. Furthermore, we apply LSTM design to the secret layer. The time step size, cluster size and age are 100, 50, 500 separately. We use softmax for the yield layer and stochastic slope good (SGD) for an analyzer. Also, the misfortune work is mean squared blunder (MSE).



d1 d2 d3 d4 d5 d6 d7
d-->data

Test Data

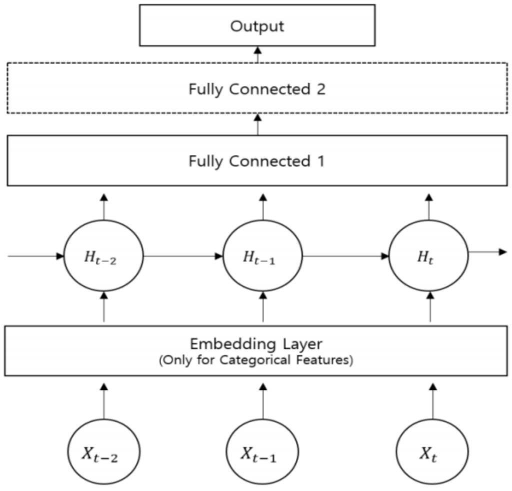
Detection Rate False Alarm Rate

IDS BASED ON LSTM:

Model Architecture

Model Architecture: installing, LSTM, and completely associated layers. 'Completely Connected 2' is utilized distinctly for paired arrangement.

Our model is made out of 3 kinds of layers: installing, LSTM, and completely associated layers. The installing layer is just for ostensible highlights of an info, and nonstop highlights are saved. 3 ostensible highlights (proto, administration, and state) are planned to 5, 3, and 2 dimensional vectors, individually. These yield vectors are connected to ceaseless highlights and travel to the following layer in the model. The LSTM layer is made out of covered up state with 100 hubs. The completely associated layer is of size 50 with dropout. As actuation work, cracked ReLU is applied for non-direct change. If there should be an occurrence of parallel grouping, the second completely associated layer is added with size of 10 hubs. the dabbed line demonstrates the layer working just in the event of twofold characterization.



Among the LSTM models, the M2M+EMB model accomplished the best for both twofold and numerous grouping errands. It is on the grounds that straight out highlights incorporates recognizable data and highlight implanting is proficient to catch the data for neural organizations .As a matter of fact, when contrasting EMB models with comparing non-EMB models, the EMB models have better execution (around 1% higher for parallel characterization and 2% higher for multi-order) and more steady outcomes as demonstrated in Figs. Organization Intrusion Detection dependent on LSTM and Feature Embedding.

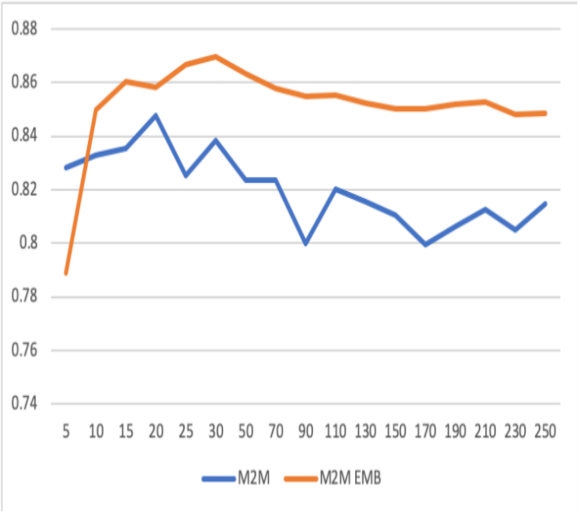


Fig: Multi-grouping precision diagrams on the approval information: M2M, and M2M with installing. The flat hub demonstrates the length of grouping.

Dataset

The DARPA's program for ID assessment of 1998 was overseen and arranged by Lincoln Labs of MIT.

The KDD interruption location challenge's dataset of 1999 was a very much refined variant of this.

Redresses Linear units has ended up being more productive and have an itemized report and significant inadequacies of the given manufactured informational collection like KDDCup-'98' and KDDCup-'99" were talked about by [49].

The principle judgment was that they neglected to approve their informational index a reenactment of genuine organization traffic profile. Regardless of every one of these reactions, the dataset of KDDCup-'99' has been utilized as a viable dataset by numerous scientists for seat denoting the IDS calculations throughout the long term. Rather than the investigation about the formation of the dataset, [50] has uncovered a nitty-gritty examination of the substance, distinguished the non-consistency and reenacted the antiquities in the mimicked network traffic information. The purposes for why the AI classifiers have restricted ability in distinguishing the assaults that have a place with the substance classifications R2L, U2R in KDDCup-'99' datasets have been talked about by [51].

The recreated assaults were ordered comprehensively as given underneath :

- Denial-of-Service-Attack (DOS): Interruption where an individual intends to make a host distant to its real reason by momentarily or some of the time forever upsetting administrations by flooding the objective machine with tremendous measures of solicitations and consequently over-burdening the host [52].

- User-to-Root-Attack (U2R): A category of commonly used maneuver by the perpetrator start by trying to gain access to a user's pre-existing access and exploiting the holes to obtain root control.

- Remote-to-Local-Attack (R2L): The intrusion in which the attacker can send data packets to the target but has no user account on that machine itself, tries to exploit one vulnerability to obtain local access cloaking themselves as the existing user of the target machine.

- Probing-Attack: The type in which the perpetrator tries to gather information about the computers of the network and the ultimate aim for doing so is to get past the firewall and gaining root access.

CONCLUSION

DNNs models on cutting edge equipment through conveyed approach. Because of broad computational expense related with complex DNNs structures, they were not prepared in this examination utilizing the benchmark IDS dataset. In this paper, we proposed a half and half interruption recognition ready framework utilizing an exceptionally adaptable structure on product equipment worker which has the ability to dissect the organization and host-level exercises. The system utilized conveyed profound learning model with DNNs for taking care of and investigating enormous scope information progressively. The DNN model was picked by exhaustively assessing their exhibition in contrast with old style AI classifiers on different benchmark IDS datasets. What's more, we gathered host-based and network-based highlights continuously and utilized the proposed DNN model for distinguishing assaults and interruption.

The expansion in huge variations of profound learning calculations requires a general assessment of these calculations with respect to its viability towards IDSs. This will be one of the bearing towards IDS exploration can travel and subsequently will stay as a work of future.

For the purpose of reference, other classical ML algorithms have been accounted and compared against the results of DNN. The publicly available KDDCup-'99' dataset has been primarily used as the benchmarking tool for the study, through which the superiority of the DNN over the other compared algorithms have been documented clearly. For further refinement of the algorithm, this paper takes into account of DNNs with different counts of hidden layers and it was concluded that a DNN with 3 layers has been proven to be effective and accurate of all.

Bibliography

[1] Mukherjee, B., Heberlein, L. T., & Levitt, K. N. (1994). Network intrusion detection. IEEE network, 8(3), 26-41.

[2] Venkatraman, S., Alazab, M. "Use of Data Visualisation for Zero-Day Malware Detection," Security and Communication Networks, vol. 2018, Article ID 1728303, 13 pages, 2018. https://doi.org/10.1155/2018/1728303

[3] Mishra, P., Varadharajan, V., Tupakula, U., & Pilli, E. S. (2018). A detailed investigation and analysis of using machine learning techniques for intrusion detection. IEEE Communications Surveys & Tutorials.

[4] Staudemeyer, R. C. (2015). Applying long short-term memory recurrent neural networks to intrusion detection. South African Computer Journal, 56(1), 136-154.

5. Beaugnon, A., Chifflier, P.: Machine learning for computer security detection systems: Practical feedback and solutions. In: 2018 Computer & Electronics Security Application Rendez-vous (2018)

6. Shah, S.A.R., Issac, B.: Performance comparison of intrusion detection systems and application of machine learning to snort system. Future Generation Computer Systems 80, 157 – 170 (2018). https://doi.org/10.1016/j.future.2017.10.016

7. Northcutt, S., Zeltser, L., Winters, S., Kent, K., Ritchey, R.W.: Inside Network Perimeter Security (2Nd Edition) (Inside). Sams, Indianapolis, IN, USA (2005)

8. Laskov, P., Rieck, K., Muller, K.R.: Machine Learning for Intrusion Detection, pp. 366–373. IOS press (09 2008)

9. Garc´ıa-Teodoro, P., D´ıaz-Verdejo, J., Maci´a-Fern´andez, G., V´azquez, E.: Anomalybased network intrusion detection: Techniques, systems and challenges. Comput. Secur. 28(1-2), 18–28 (Feb 2009). https://doi.org/10.1016/j.cose.2008.08.003

10. The zeek network security monitor. https://docs.zeek.org/en/stable/intro/ index.html, accessed: 2019-08-01

11. Wenke Lee, Stolfo, S.J., Mok, K.W.: A data mining framework for building intrusion detection models. In: IEEE Symposium on Security and Privacy (Cat. No.99CB36344). pp. 120–132 (May 1999). https://doi.org/10.1109/SECPRI.1999.766909

12. Tang, T.A., Mhamdi, L., McLernon, D., Zaidi, S.A.R., Ghogho, M.: Deep recurrent neural network for intrusion detection in sdn-based networks. In: 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft). pp. 202– 206 (June 2018). https://doi.org/10.1109/NETSOFT.2018.8460090

13. Moon, T., Choi, H., Lee, H., Song, I.: Rnndrop: a novel dropout for rnns in asr. In: ASRU. pp. 65–70 (12 2015). https://doi.org/10.1109/ASRU.2015.7404775

14. Bahdanau, D., Cho, K., Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate. In: International Conference on Learning Representation (ICLR) (2015), http://arxiv.org/abs/1409.0473

15. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation 9(8), 1735–1780(Nov1997). https://doi.org/10.1162/neco.1997.9.8.1735

[16] C. Elkan. ”Results of the KDD’99 classifier learning”. SIGKDD explorations newsletter, vol. 1, pp. 63 64, 2000. DOI http://dx.doi.org/10.1145/846183. 846199.

17. Greff, K., Srivastava, R.K., Koutn´ık, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems 28(10), 2222–2232 (2017)

18. Pennington, J., Socher, R., Manning, C.D.: GloVe: Global Vectors for Word Representation. In: Empirical Methods in Natural Language Processing. pp. 1532–1543 (2014). https://doi.org/10.3115/v1/D14-1162

19. Choi, H., Cho, K., Bengio, Y.: Context-dependent word representation for neural machine translation. Computer Speech and Language 45, 149–160 (2017)

20. Mikolov, T., Kombrink, S., Deoras, A., Burget, L., Cernock´y, J.: RNNLM - Re- ˇ current Neural Network Language Modeling Toolkit. In: ASRU. pp. 1–4 (2011)

[21] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), p.436.

[22] Sommer, R. and Paxson, V., 2010, May. Outside the closed world: On using machine learning for network intrusion detection. In Security and Privacy (SP), 2010 IEEE Symposium on (pp. 305-316). IEEE

[23] Liao, Yihua, and V. Rao Vemuri, Use of k-nearest neighbor classifier for intrusion detection, Computers & Security 21.5, pp.439-448, 2002

[24] Sung, Andrew H., and Srinivas Mukkamala, Identifying important features for intrusion detection using support vector machines and neural networks, Applications and the Internet, 2003

[25] Sabhnani, Maheshkumar, and Grsel Serpen. , Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context, MLMTA, 2003

579 595, 2000. DOI http://dx.doi.org/10.1016/S1389-1286(00)00139-0.