International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Bidirectional LSTM Based Approach for Network Intrusion Detection

Praveen Kumar Kollu, R. Satya Prasad

Abstract*: *In today's digital age the number of people that have access to the internet services have grown in leaps and bounds. It has been directly associated with the increasing security threats. Traditionally it has been difficult to handle these security threats, but with the recent advancements in recurrent neural network architectures have paved a path for effective threat identification. Intrusion detection system has been a widely researched area in security analysis and evaluation. In this paper, we are leveraging the Bidirectional LSTM (Long Short Term Memory) networks for Intrusion detection model. Bidirectional LSTM uses all the available information in the network and also provides context to the network than the normal LSTM. NSL-KDD dataset is used to validate the proposed model. To evaluate our proposed model, we have compared our model against vanillaRNN, normal LSTM and GRU networks which are most popular models for network intrusion detection. They are compared in terms of accuracy, precision, recall and F1-measure.

Index Terms*: *Intrusion Detection, KDD dataset, Network Security, Neural Networks

I. INTRODUCTION

Network security has become significant requirement for institutions and organizations alike. With the popularity and growth of cloud based services, cloud security has also received prominent attention. It is estimated to be worth $9 Billion USD by 2020[1]. Intrusion detection is one of the most widely researched areas in network security. Most of the traditional approaches for intrusion detection has become redundant as they fail find the subtle relations in the data, which are most important for effective intrusion detection. It is especially difficult for cloud security as often the control of the cloud resources is shared between the user and the service provider.

Intrusion detection system is a passive system that detects the unwanted perpetrators in the network that does not have the permission to access those resources in the network. By accessing the network without authorization perpetrators can steal sensitive information, tamper the network integrity and even bring down the entire network. if there are vulnerabilities present in the network they can be exploited by the authorized users of the network [2]. A robust intrusion detection system analyses the data from the network to detect any intrusions, finds vulnerable parts of the network & tests the overall integrity [2,3]. Most of the traditional approaches for intrusion detection have been based on the widely popular machine learning approaches. Tree based algorithms such as J48 Decision trees, random forests are used [4,5,6]. Other approaches include distance based kNN (K Nearest Neighbors), discriminative based SVM (Support Vector Machines) [7,8]. The drawbacks with these approaches is that the feature extraction is often a tedious process and feature selection is not complete. They also cannot remember previous state of the network as it is very important to keep the state information so that change can be easily identified.

With the advancements in the deep learning technologies, RNN (Recurrent neural networks) have become computationally feasible. They can have memory module at each layer so that they can save the previous state information they deem important. This feature is highly suitable for intrusion detection as keeping track of the state changes at different times is necessary to detect precise intrusions. In this paper, we are proposing a Bidirectional LSTM based approach for intrusion detection system (IDS). The bidirectional models add extra context information to the model which can in turn increase the efficiency and learning of the model. NSL-KDD [9] dataset is evaluated on the proposed model as it has been one of the benchmark dataset for network intrusion detection. The performance of our proposed model is measured in terms of accuracy, precision, recall and F1-measure and is compared against the normal LSTM and GRU networks.

II. RELATED WORK

Recently there has been a lot of research applying RNN architectures in network intrusion detection. Jihyun Kim et al. [10] applied LSTM architecture to the KDD Cup 1999 dataset. They have performed two experiments, finding hyper-parameter values and performance measurement. Their proposed approach is evaluated using Detection Rate (DR) and False Alarm Rate (FAR). Chuanlong Yin et al. [11] approach is to use normal RNN for intrusion detection system. NSL-KDD cup was used to train the model and the results were compared against J48, ANN, Randomforest and SVM. The results have shown that their approach has outperformed the rest in binary classification and multi class classification. Tuan A Tang et al. [12] research focused on intrusion detection in software defined networks (SDN). Gated Recurrent Network architecture was used for developing IDS and was tested using the NSL-KDD dataset. [13] used LSTM RNN to detect network traffic anomalies. It predicts the communications between two IPS. ISCX IDS dataset is used to evaluate the models performance.

Revised Manuscript Received on 30 May 2019.
***** Correspondence Author
Praveen Kumar Kollu*, Research Scholar, Acharya Nagarjuna University, Guntur, India.
R. Satya Prasad, Professor, Department of CSE, Acharya Nagarjuna University, Guntur, India.

III. METHODOLOGY

A. Dataset

NSL-KDD dataset was used to implement our proposed model. The dataset has been a popular benchmark dataset for research on intrusion detection system. NSL-KDD improves on the KDD'99 dataset [9] especially it does not have the redundant data that is found in the older dataset. It also has no duplicate records which makes it suitable for training models. The dataset contains around 42 feature attributes of 125973 records. Out of the 42 attributes 41 are independent variables and one attribute is dependent variable used for classification. The dataset contains 9 attributes that are basic features of the TCP connections, 13 are Content features, 9 features are traffic features and final 10 attributes are host features. The list of attributes in the dataset are given in the Table 1. The final attribute is used for binary classification between intrusion network and normal network.

Table 1. Attributes in NSL KDD Dataset

S.NO	Attribute Name
1	duration
2	protocol type
3	service
4	src_bytes
5	dst bytes
6	ag
7	land
8	wrong fragment
9	urgent
10	hot
11	num failed logins
12	logged in
13	num compromised
14	root shell
15	su attempted
16	num root
17	num _le creations
18	num shells
19	num access _les
20	num outbound cmds
21	is hot login
22	is guest login
23	count
24	serror rate
25	rerror rate
26	same srv rate
27	di_ srv rate
28	srv count
29	srv serror rate
30	srv rerror rate
31	srv di_ host rate
32	dst host count
33	dst host srv count
34	dst host same srv rate
35	dst host di_ srv rate
36	dst host same src port rate
37	dst host srv di_ host rate
38	dst host serror rate
39	dst host srv serror rate
40	dst host rerror rate
41	dst host srv rerror rate

B. Pre-Processing

As the network only accepts numerical data as input data has to be pre-processed before it was trained. Initially non-numerical data has to be converted in to numerical data. This can be achieved by converting the non-numerical data in to corresponding individual attributes and the values of those attributes are taken as binary vectors, where in the value of that particular record, the corresponding attribute will have 1 and rest of the remaining attributes will have 0. The next step is to normalize all the attribute values. In this normalization is done record wise instead of attribute wise so that each component is rescaled between 0-1 independent of other records. The architecture for pre-processing is shown in Figure 1.

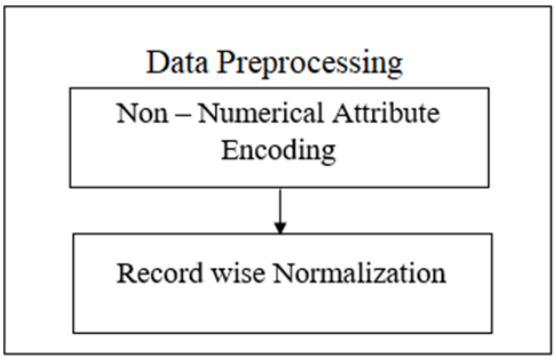

Figure 1. Architecture of Data Pre-Processing

C. vanillaRNN

vanillaRNN or simple RNN are recurrent neural networks that does not have any gates. They only have hidden layers and input at a specific time step. The input is multiplied by the hidden layer which is also the output from the previous RNN cell. It has only a tanh activation function.

D. LSTM

Long Short Term Memory (LSTM) [14] networks are huge improvement over normal vanillaRNN networks. They can remember longer context information than RNN and also by having different cell states it can decide what information is important and not important. LSTM contains different gates and a cell state as shown in the Figure 2. The forget gate is used to decide whether to keep the information or not. It contains a sigmoid function which outputs a values between 0 and 1. If the value of the activation function in 0, the information is forgotten and if the value is 1, then the information is kept.

The cell state is updated by the input gate. Input gate takes the previous hidden state and current input. It contains tanh and sigmoid activation function and also has their multiplied values. Now the cell state is calculated by performing pointwise addition with the output of the input gate. Finally, the output gates decide the next hidden state value.

E. GRU

Gated Recurrent Unit (GRU) is similar to the LSTM network but it contains only the update gate which decides whether to pass the previous output to the next cell or not. It also does not have a specific forget gate as it is replaced by an additional mathematical operation.

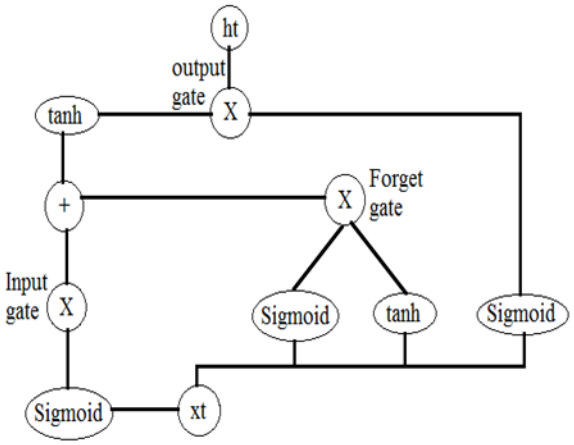

Figure 2. Single LSTM Cell

F. Bidirectional LSTM

Bidirectional LSTM add additional context information to the network. This can result in faster learning and more information propagation. A Bidirectional LSTM contains a two LSTM layers side by side. One LSTM is called the forward pass and the other LSTM is reverse of the first one. The forward pass contains a positive time dimension and the backward pass contains a negative time dimension. The outputs of the two LSTMs are merged by using concatenation. The Bidirectional LSTM is shown in Figure 3. It shows two LSTM layers one with the forward pass and other one reverse to the first one. It also shows the concatenation operation of the outputs from the two layers.

Having two LSTM passes can help in applications like intrusion security due to the dynamic nature of the network. Continuous flow of data is highly difficult to classify with a single pass of the LSTM network. It increases the false alarm rate. But having two passes can reduce the false alarm rates and also classify intrusion which are previously unnoticed. This type of true positive classification is desirable because even a minimal misclassification can cause the network to become vulnerable. Our proposed model can add the additional context to minimize the misclassifications by having two passes in the network.

IV. EXPERIMENTS

A. Implementation Details

Keras framework is used to implement the proposed model [15]. As the proposed model has a binary classification problem binary cross entropy is used as loss function. Adam optimizer [16] is used in the network as it provides faster learning. The model was trained for around 300 epochs with a 20% validation split and it achieves convergence around 200 epochs.

B. Evaluation Metrics

For evaluating the proposed model following evaluation measures are used.

Accuracy: It is the percentage of true records detected over total number of records. It is calculated as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \#(1)$$

Precision: The ratio of correctly predicted anomalies to actual anomalies.

$$Precision = \frac{TP}{TP + FP} \#(2)$$

Recall: The ratio of correctly predicted anomalies to all the instances available.

$$Recall = \frac{TP}{TP + FN} \#(3)$$

F1-measure: It is the weighted average between the precision and recall.

$$F1 - Measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \#(4)$$

where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

C. Results

Our proposed model has been compared with normal LSTM and GRU networks. It has been evaluated on Accuracy, Precision, Recall and F1-Measure. Table 2 shows the accuracy measure between normal RNN namely vanillaRNN, LSTM, GRU and the proposed Bidirectional LSTM. It shows that the proposed model outperforms the remaining models and also trains faster. Figure 4 shows the training plot of Bidirectional LSTM and Figure 5 plots the loss.

Table 3, 4, 5 shows the Precision, Recall and F1-measure respectively. They also show the similar results where the proposed model has superior performance than the rest of the models.

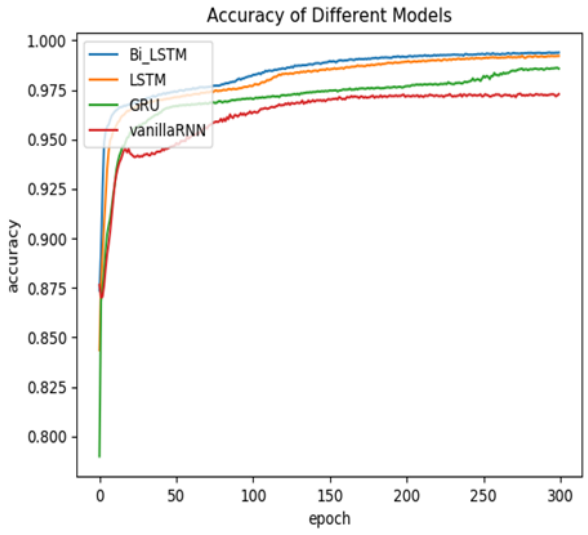

Figure 4. Accuracies of Models

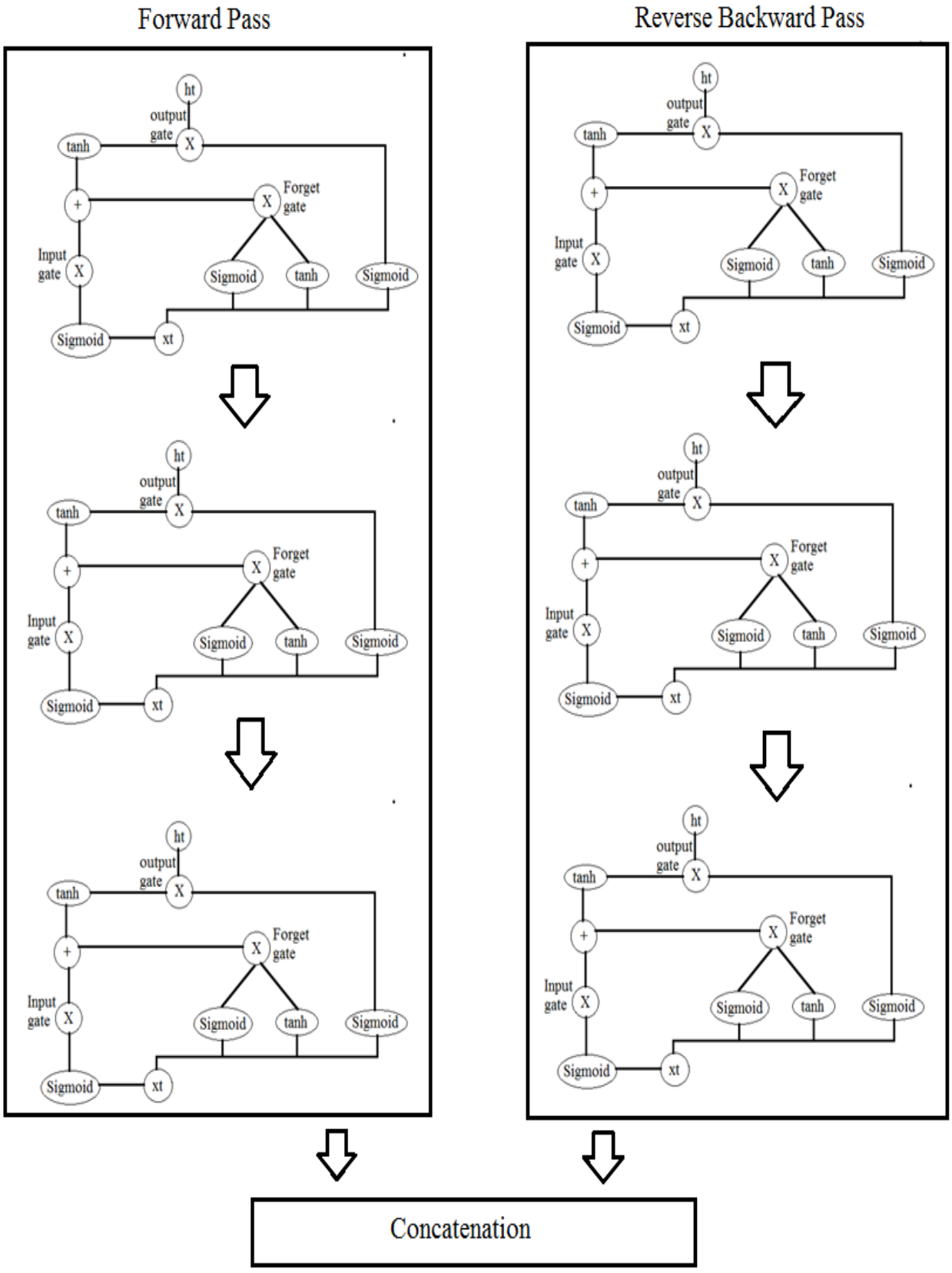

Figure 3. Bidirectional LSTM Architecture

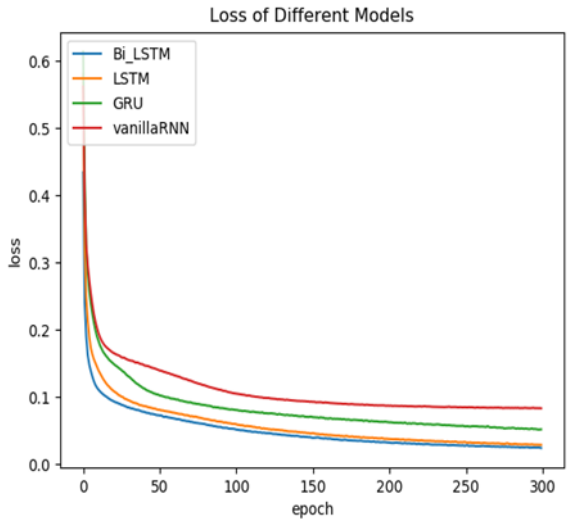

Figure 5. Loss of Models

Table 2. Accuracy

Model	Accuracy
vanillaRNN	95.45
LSTM	97.62
GRU	98.55
Bi-LSTM	99.80

Table 3: Precision Measure

Model	Precision
vanillaRNN	97.89
LSTM	98.30
GRU	98.35
Bi-LSTM	98.86

Table 4: Recall Measure

Model	Recall
vanillaRNN	94.80
LSTM	97.30
GRU	98.56
Bi-LSTM	99.52

Table 5: F1-Measure

Model	F1-Measure
vanillaRNN	96.32
LSTM	97.79
GRU	98.45
Bi-LSTM	99.18

V. CONCLUSION

In this research, we have proposed a new model for intrusion detection system. The proposed model adds more context to the network which can result in faster learning and accurate results. NSL-KDD dataset was used to train and evaluate the model. NSL-KDD has been widely used benchmark dataset for IDS research. Experiments have shown that our model outperforms other RNN based model in terms of all the evaluation metrics. Our model can pave a path for more robust Intrusion Detection Systems which can utilize the additional context information and time sensitivity.

REFERENCES

1. Gartner Inc. (2017), “Gartner Forecasts Worldwide Cloud-Based Security Services to Grow 21 Percent in 2017,” Available: https://www.gartner.com/en/newsroom/press-releases/2017-06-13-gartner-forecasts-worldwide-cloud-based-security-services-to-grow-21-percent-in-2017 accessed 29 Mar. 2019.
2. L. Dali et al., "A survey of intrusion detection system," 2015 2nd World Symposium on Web Applications and Networking (WSWAN), Sousse, 2015, pp. 1-6. doi: 10.1109/WSWAN.2015.7210351.
3. F. Sabahi and A. Movaghar, "Intrusion Detection: A Survey," 2008 Third International Conference on Systems and Networks Communications, Sliema, 2008, pp. 23-26. doi: 10.1109/ICSNC.2008.44.
4. M. Almseidin, M. Alzubi, S. Kovacs and M. Alkasassbeh, "Evaluation of machine learning algorithms for intrusion detection system," 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, 2017, pp. 000277-000282. doi: 10.1109/SISY.2017.8080566.
5. S. Sahu and B. M. Mehtre, "Network intrusion detection system using J48 Decision Tree," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, 2015, pp. 2023-2026. doi: 10.1109/ICACCI.2015.7275914.
6. Y. Chang, W. Li and Z. Yang, "Network Intrusion Detection Based on Random Forest and Support Vector Machine," 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou, 2017, pp. 635-638. doi: 10.1109/CSE-EUC.2017.118.
7. Liao, Yihua, and V. Rao Vemuri, Use of k-nearest neighbor classifier for intrusion detection, Computers & Security 21.5, pp.439-448, 2002.
8. Wun-Hwa Chen, Sheng-Hsun Hsu, Hwang-Pin Shen, Application of SVM and ANN for intrusion detection, Computers \& Operations Research, Volume 32, Issue 10, 2005, Pages 2617-2634, ISSN 0305-0548, https://doi.org/10.1016/j.cor.2004.03.019.
9. M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.
10. J. Kim, J. Kim, H. L. T. Thu and H. Kim, "Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection," 2016 International Conference on Platform Technology and Service (PlatCon), Jeju, 2016, pp. 1-5. doi: 10.1109/PlatCon.2016.7456805.
11. C. Yin, Y. Zhu, J. Fei and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," in IEEE Access, vol. 5, pp. 21954-21961, 2017. doi: 10.1109/ACCESS.2017.2762418.
12. T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi and M. Ghogho, "Deep Recurrent Neural Network for Intrusion Detection in SDN-based Networks," 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft), Montreal, QC, 2018, pp. 202-206. doi: 10.1109/NETSOFT.2018.8460090.
13. Radford, Benjamin & M. Apolonio, Leonardo & J. Trias, Antonio & A. Simpson, Jim. (2018). Network Traffic Anomaly Detection Using Recurrent Neural Networks.CoRR abs/1803.10769 (2018).
14. Gers, F., Schraudolph, N., & Schmidhuber, J. (2002). Learning precise timing with LSTM recurrent networks. Journal of Machine Learning Research, 3, 115–143.
15. F. Chollet et al. (2015). Keras. [Online]. Available: https://github.com/fchollet/keras
16. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

AUTHORS PROFILE

Praveen Kumar Kollu is a research Scholar in department of CSE, Acharya Nagarjuna University and working as Sr. Assistant Professor in Computer Science and Engineering Department, Velagapudi Ramakrishna Siddhartha Engineering College. He is having 15+ experiences in Teaching. His areas of interest are Intrusion detection, deep learning, Machine learning, bioinformatics and cyber security.

Dr. R. Satya Prasad is now employed as a professor in computer science and engineering department, Acharya Nagarjuna University. Guntur. He has around 70 publications in reputed national and international journals. He is also the recipient of Dr. Abdul Kalam Life Time Achievement Award for his efforts in Teaching and Research. His research interests include Digital Image Processing, Databases and Software Engineering.