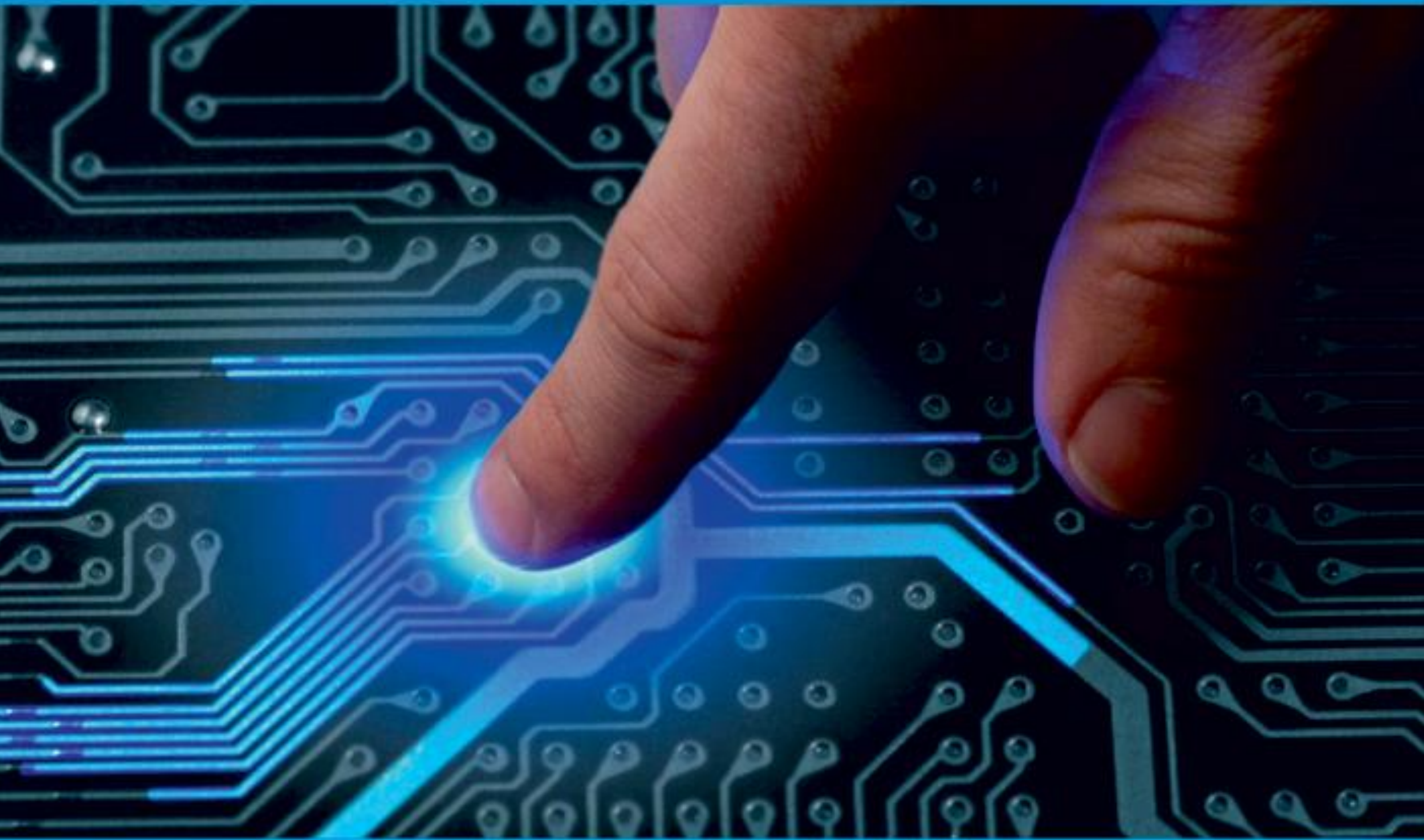
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 4, April 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004147|

Attack Recognition System using Deep Learning and Machine Learning Algorithm

Katrajkar Manas Nitin ,Kamble Sayali Suresh , Nagpure Mayuri Sunil , Hange Jijabai Bandu , Prof.N.B.Pokale

Department of Computer Engineering, TSSM's Bhivarabai Sawant College of Engineering & Research, (Savitribai Phule Pune University), Pune, India

ABSTRACT - Redundant and in applicable features in data have precipitated a long haul problem in network traffic classification. In recent years, one of the foremost focuses inside NIDS studies has been the application of machine learning and shallow learning knowledge of techniques. This paper proposes a novel deep learning model to enable NIDS operation within modern networks. The model shows a combination of deep and shallow learning, capable of correctly analyzing a wide-range of network traffic. The novel approach proposes non-symmetric deep autoencoder (NDAE) for unsupervised feature learning. Moreover, additionally proposes novel deep learning classification display built utilizing stacked NDAEs. Our proposed classifier has been executed in Graphics processing unit (GPU)-empowered TensorFlow and assessed utilizing the WSN Trace dataset. The performance evaluated network intrusion detection analysis dataset, particularly WSN Trace dataset. The contribution work is to implement intrusion prevention system (IPS) contains IDS functionality but more sophisticated systems which are capable of taking immediate action in order to prevent or reduce the malicious behavior.

KEYWORDS: Deep learning; Anomaly detection; Auto-encoders; Neural Network; Network security.

I. INTRODUCTION

One of the major challenges in network security is the provision of a robust and effective Network Intrusion Detection System (NIDS). Despite the considerable advances in NIDS system, the majority of solutions still operate using less-successful signature-based techniques, rather than anomaly detection strategies. The current issues are the existing techniques leads to ineffective and inaccurate detection of attacks. There are three main limitations like, volume of network data, in-depth monitoring and granularity required to improve effectiveness and accuracy and finally the number of different protocols and diversity of data traversing. The main focus of NIDS research has been the application of machine learning and shallow learning techniques. The initial deep learning research has demonstrated that its superior layer-wise feature learning can better or at least match the performance of shallow learning techniques. It is able to facilitating a deeper evaluation of network data and faster identification of any anomalies. In this paper, proposes a novel deep learning version to enable NIDS operation inside modern networks.

Thus refer the papers [1], [2], [3] and [4] and study all the details about how to work Deep Belief Network (DBN) and Restricted Boltzmann Machine (RBM) for intrusion detection. The contribution of intrusion prevention system refers the paper [12] using for rule generation technique and according to take the action against it.

A. Motivation

- A new NDAE technique for unsupervised feature learning, which unlike typical auto-encoder approaches provides non-symmetric data dimensionality reduction. Hence, our technique is able to facilitate improved classification results when compared with leading methods such as Deep Belief Networks (DBNs).
- A novel classifier model that utilizes stacked NDAEs and the RF classification algorithm. By combining both deep and shallow learning techniques to exploit their respective strengths and reduce analytical overheads. We are able to better or at least match results from similar research, whilst significantly reducing the training time.

II. RELATED WORK

The paper [2] focuses on deep learning methods which are inspired by the structure depth of human brain learn from lower level characteristic to higher levels concept. It is because of abstraction from multiple levels, the Deep Belief Network (DBN) helps to learn functions which are mapping from input to the output. The technique of getting to know does not dependent on human-crafted features. DBN makes use of an unsupervised learning algorithm, a Restricted Boltzmann Machine (RBM) for every layer. Advantages are: Deep coding is its ability to adapt to changing contexts concerning data that ensures the technique conducts exhaustive data analysis. Detects abnormalities in the system that includes anomaly detection, traffic identification. Disadvantages are: Demand for faster and efficient data assessment.

In [3] paper, a deep learning approach for anomaly detection using a Restricted Boltzmann Machine (RBM) and a deep belief network are implemented. This method uses a one-hidden layer RBM to perform unsupervised feature reduction. The resultant weights from this RBM are exceeded to any other RBM producing a deep belief network. The pre-trained weights are handed into a exceptional tuning layer consisting of a Logistic Regression (LR) classifier with multi-class soft-max. Advantages are: Achieves 97.9% accuracy. It produces a low false negative rate of 2.47%. Disadvantages are: Need to improve the method to maximize the feature reduction process in the deep learning network and to improve the dataset.

The paper [4] proposes a deep learning based approach for developing an efficient and flexible NIDS. A sparse autoencoder and soft-max regression based NIDS was resolved. Uses Self-taught Learning (STL), a deep learning based technique, on NSL-KDD - a benchmark dataset for network intrusion. Advantages are: STL achieved a classification accuracy rate more than 98% for all types of classification. Disadvantages are: Need to implement a real-time NIDS for actual networks using deep learning technique.

In [5] paper choose multi-core CPU’s as well as GPU’s to evaluate the performance of the DNN based IDS to handle huge network data. The parallel computing capabilities of the neural network make the Deep Neural Network (DNN) to effectively look through the network traffic with an accelerated performance. Advantages are: The DNN based IDS is reliable and efficient in intrusion detection for identifying the specific attack classes with required number of samples for training. The multi core CPU’s was faster than the serial training mechanism. Disadvantages are: Need to improve the detection accuracies of DNN based IDS.

In [6] paper, proposes a mechanism for detecting large scale network-wide attacks using Replicator Neural Networks (RNNs) for creating anomaly detection models. Our approach is unsupervised and requires no labeled data. It also accurately detects network-wide anomalies without presuming that the training data is completely free of attacks. Advantages are: The proposed methodology is able to successfully discover all prominent DDoS attacks and *SYN Port* scans injected. Proposed methodology is resilient against learning in the presence of attacks, something that related work lacks. Disadvantages are: Need to improve proposed methodology by using stacked autoencoder deep learning techniques.

Based on the flow-based essentiality of SDN, we propose a flow-based anomaly detection system using deep learning. In [7] paper, apply a deep learning approach for flow-based anomaly detection in an SDN environment. Advantages are: It finds an optimal hyper-parameter for DNN and confirms the detection rate and false alarm rate. The model gets the performance with accuracy of 75.75% which is quite affordable from just using six simple network features. Disadvantages are: It will not work on real SDN environment.

The paper [8] proposes a deep learning approach for intrusion detection using recurrent neural networks (RNN-IDS). The RNN model basically has a one-way flow of data from the input units to the hidden units, and the synthesis of the one-way data flow from the previous temporal concealment unit to the current timing hiding unit. Advantages are: It has a strong modeling ability for intrusion detection. It has higher accuracy than the other machine learning methods. Disadvantages are: It will spend more time for training dataset.

Using all the 41 features in the NSL-KDD dataset to evaluate the intrusive patterns may leads to time consuming and it also reduce performance degradation of the system [9]. CFS Subset is utilized to reduce the dimensionality of the dataset. Advantages are: The Random Forest algorithm shows the highest accuracy compared with rest of the algorithms by considering with and without feature reduction. Random Forest has the high speed for classification. Disadvantages are: Need to improve the Random Forest algorithm to build an efficient intrusion detection system.

The paper [10] proposes a deep learning based approach to implement such an effective and flexible NIDS. We use Self-taught Learning (STL), a deep learning primarily based approach, on NSL-KDD - a benchmark dataset for network intrusion. Self-taught Learning (STL) is a deep learning technique that consists of two stages for the classification. First, a good feature representation is learnt from a large collection of unlabeled data, known as Unsupervised Feature Learning (UFL). In the second stage, this learnt representation is applied to labeled data, and used for the classification task. Although the unlabeled and labeled data may come from different distributions, there must be relevance among them. Advantages are: NIDSs based on this approach achieved very high-accuracy and less false-alarm rates. Disadvantages are: Need to work on stacked auto-encoders in deep belief network.

In [11] paper, anomaly-based network intrusion detection techniques are a valuable technology to protect target systems and networks against malicious activities. The main A-NIDS technologies, together with their general operational architecture, and provides a classification for them according to the type of processing related to the ‘‘behavioral’’ model for the target system. There are three types of Anomaly detection: statistical, knowledge and machine learning-based techniques. Advantages are: Machine learning A-NIDS is the use of a flexible and robust global search method. Disadvantages are: High resource consumption.

III. OPEN ISSUES

The current network traffic data, which can be regularly large in size, present a major task to IDSs. These “big data” slow down the whole detection process and might result in unsatisfactory classification accuracy due to the computational difficulties in coping with such data. Machine learning technologies have been usually used in IDS. However, most of the traditional machine learning technologies refer to shallow learning; they cannot effectively solve the enormous intrusion data classification issue that arises in the face of a real network application environment. Additionally, shallow learning is incompatible to intelligent analysis and the predetermined requirements of high-dimensional learning with enormous data.

Disadvantages are**:**

Computer systems and internet have become a primary part of the critical system. The current network traffic data, which may be frequently huge in size, present a major task to IDSs. These “big data” slow down the whole detection process and might result in unsatisfactory classification accuracy due to the computational difficulties in coping with such data. Classifying a large amount of data usually causes many mathematical difficulties which then lead to higher computational complexity.

IV. SYSTEM OVERVIEW

The paper [1] proposes a novel deep learning model to enable NIDS operation within modern networks. The model proposes is a combination of deep and shallow learning, capable of correctly analyzing a wide-range of network traffic. More specifically, combine the power of stacking our proposed Non-symmetric Deep Auto-Encoder (NDAE) (deep learning) and the accuracy and speed of Random Forest (RF) (shallow learning). This paper proposes NDAE, which is an auto-encoder featuring non-symmetrical multiple hidden layers. NDAE may be utilized as a hierarchical unsupervised feature extractor that scales properly to deal with excessive-dimensional inputs. It learns non-trivial features using a similar training approach to that of a regular auto-encoder. Stacking the NDAEs offers a layer-wise unsupervised representation learning algorithm, which will allow our model to learn the complex relationships between different features. It also has feature extraction capabilities, so it is able to refine the model by prioritizing the most descriptive features.

Fig. 1 shows the proposed system architecture of Network Intrusion Detection and Prevention System (NIDPS). The input traffic data is uses for WSN Trace dataset with 12 features. The training dataset contains data preprocessing which includes two steps: Data transformation and data normalization. After uses two NDAEs arranged in a stack, which uses for selecting number of features. After that apply the Random Forest Classifier [9] for attack detection. Intrusion Prevention Systems (IPS) contains IDS functionality but more sophisticated systems which are capable of taking immediate action in order to prevent or reduce the malicious behavior. The intrusion prevention system is implementing with the help of Rule Status Monitoring Algorithm [12]. There are 8 rule actions when the attack detected or not, the system will take the action using following list:

- ALERT - Generate an alert using the selected ALERT method, and then log the packet
- LOG - Log the packet
- PASS - Ignore the packet
- ACTIVATE - Alert and then turn on another dynamic rule
- DYNAMIC - Remain idle until activated by an activate rule , then act as a log rule
- DROP - Block and log the packet
- REJECT - Block the packet, log it, and then send a TCP reset if the protocol is TCP or an ICMP port unreachablemessage if the protocol is UDP
- SDROP - Block the packet but do not log it.

A. Architecture

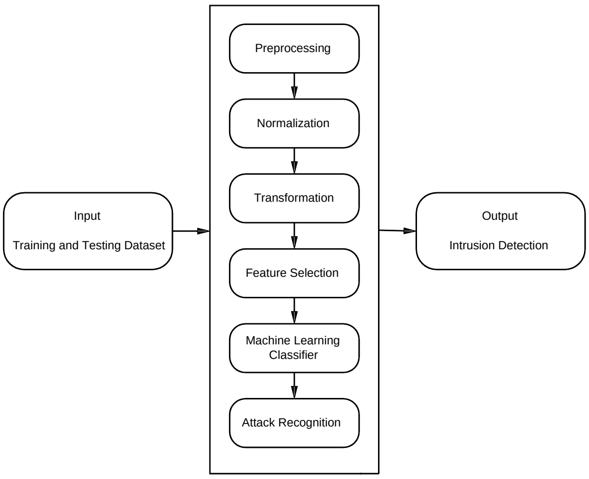


Fig. 1 Proposed System Architecture

B. Mathematical Model

1. Preprocessing:

In this step, training data source (T) is normalized to be equipped for processing by using following steps:

$$T_{norm} = \{\frac{T-\mu_T}{\sigma_T}, \sigma_T \neq 0\ andT - \mu_T, \sigma_T = 0 \qquad (1)$$

Where,

$$T = \{x_{i,j} | i = 1,2, \dots, mandj = 1,2,3, \dots, n\}$$
$$\mu_T = \{\mu_j | j = 1,2,3, \dots, n\}$$
$$\sigma_T = \{\sigma_j | j = 1,2,3, \dots, n\}$$

T is m samples with n column attributes; x_{ij} is the jth column attribute in ith sample, $\mu_T$and μ_T are 1*n matrix which are the training data mean and standard deviation respectively for each of the n attributes. Test dataset (TS) which is used to measure detection accuracy is normalized using the same μ_T and μ_T as follows:

$$TS_{norm} = \frac{(TS-\mu_T)}{\sigma_T}, \sigma_T \neq 0\ and TS - \mu_T, \sigma_T = 0 \qquad (2)$$

2. Feature Selection:

NDAE is an auto-encoder featuring *non-symmetrical* multiple hidden layers. The proposed NDAE [1] takes an input vector $x \in R^d$and step-by-step maps it to the latent representations $h_i \in R^d$(here d represents the dimension of the vector) using a deterministic function shown in (3) below:

$$h_i = \sigma(W_i.h_{i-1} + b_i);\ i = \overline{1, n}, \qquad (3)$$

Here, $h_0 = x,\ \sigma$is an activation function (in this work use sigmoid function $\sigma(t) = 1/(1 + e^{-t})$) and n is the number of hidden layers. Unlike a conventional auto-encoder and deep auto-encoder, the proposed NDAE does not contain a decoder and its output vector is calculated by a similar formula to (4) as the latent representation.

$$y = \sigma(W_{n+1}.h_n + b_{n+1}) \qquad (4)$$

The estimator of the model $\theta = (W_i, b_i)$can be obtained by minimizing the square reconstruction error over m training samples $(x^{(i)}, y^{(i)})_{i=1}^{m}$, as shown in (5).

$$E(\theta) = \sum_{i=1}^{m}(x^{(i)}, y^{(i)})^2 \quad (5)$$

V. RESULT AND DISCUSSIONS

WSN-Trace is the wireless dataset for researchers. The WSN Trace dataset.

For experimental set up, use Windows 7 operating system, Intel i5 processor, 4 GB RAM, 200GB Hard disk, Eclipse Luna JDK 8 tool and Tomcat server. To calculate the results, WSN Trace dataset is used. WSN Trace dataset is a real-time wireless dataset which gets information from router which contains node details and packet information. In training and testing dataset, there are 5 types of attack which are subtypes of normal, probing, dos, u2r and r2l attacks.

VI. CONCLUSION

In this paper, mentioned the problems confronted by previous NIDS techniques. In response to this proposed the novel NDAE approach for unsupervised feature learning. After then built upon this by proposing a novel classification model constructed from stacked NDAEs and the CNN classification algorithm. Also implemented the Intrusion prevention system. The result shows that given approach offers high levels of accuracy, precision and recall together with reduced training time. The proposed NIDS system is improved only 5% accuracy. So, there is need to further improvement of accuracy. And also further work on real-time network traffic and to handle zero-day attacks.

REFERENCES

[1] Nathan shone , tran nguyen ngoc, vu dinh phai , and qi sh, "a deep learning approach to network intrusion detection",ieee transactions on emerging topics in computational intelligence, vol. 2, no. 1, february 2018

[2] B. Dong and X. Wang, "Comparison deep learning method to traditional methods using for network intrusion detection," in Proc. 8th IEEE Int.Conf. Commun. Softw. Netw, Beijing, China, Jun. 2016, pp. 581–585.

[3] K. Alrawashdeh and C. Purdy, "Toward an online anomaly intrusion detection system based on deep learning," in Proc. 15th IEEE Int. Conf. Mach. Learn. Appl., Anaheim, CA, USA, Dec. 2016, pp. 195–200.

[4] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in Proc. 9th EAI Int.Conf. Bio-Inspired Inf. Commun. Technol., 2016, pp. 21–26. [Online]. Available: http://dx.doi.org/10.4108/eai.3-12-2015.2262516

[5] S. Potluri and C. Diedrich, "Accelerated deep neural networks for enhanced intrusion detection system," in *Proc. IEEE 21st Int. Conf. Emerg. Technol. Factory Autom.*, Berlin, Germany, Sep. 2016, pp. 1–8.

[6] C. Garcia Cordero, S. Hauke, M. Muhlhauser, and M. Fischer, "Analyzing flow-based anomaly intrusion detection using replicator neural networks," in *Proc. 14th Annu. Conf. Privacy, Security. Trust*, Auckland, New Zeland, Dec. 2016, pp. 317–324.

[7] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, “Deep learning approach for network intrusion detection in software defined networking,” in *Proc. Int. Conf. Wireless Netw. Mobile Commun.*, Oct. 2016, pp. 258–263.

[8] 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.

[9] Revathi, S &Malathi, A. (2013). A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection. International Journal of Engineering Research & Technology (IJERT). 2. 1848-1853.

[10] N. Shone, T. N. Ngoc, V. D. Phai and Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41-50, Feb. 2018.

[11] Anomaly-based network intrusion detection: Techniques, systems and challenges Garcia-Teodoro P. Diaz-Verdejo J., Macia-Fernandez G., Vazquez E. (2009) *Computers and Security*, 28 (1-2), pp. 18-28.

[12] Claude Turner*, Rolston Jeremiah, Dwight Richards, Anthony Joseph, “A Rule Status Monitoring Algorithm for Rule-Based Intrusion Detection and Prevention Systems”, Procedia Computer Science, ISSN: 1877-0509, Vol: 95, Page: 361-368, 2016

[13] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, “Deep learning and its applications to machine health monitoring: A survey,” Submitted to IEEE Trans. Neural Netw. Learn. Syst., 2016. [Online]. Available: http://arxiv.org/abs/1612.07640

[14] H. Lee, Y. Kim, and C. O. Kim, “A deep learning model for robust wafer fault monitoring with sensor measurement noise,” IEEE Trans. Semicond. Manuf., vol. 30, no. 1, pp. 23–31, Feb. 2017.

[15] L. You, Y. Li, Y. Wang, J. Zhang, and Y. Yang, “A deep learning based RNNs model for automatic security audit of short messages,” in Proc. 16th Int. Symp. Commun. Inf. Technol., Qingdao, China, Sep. 2016, pp. 225–229.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com