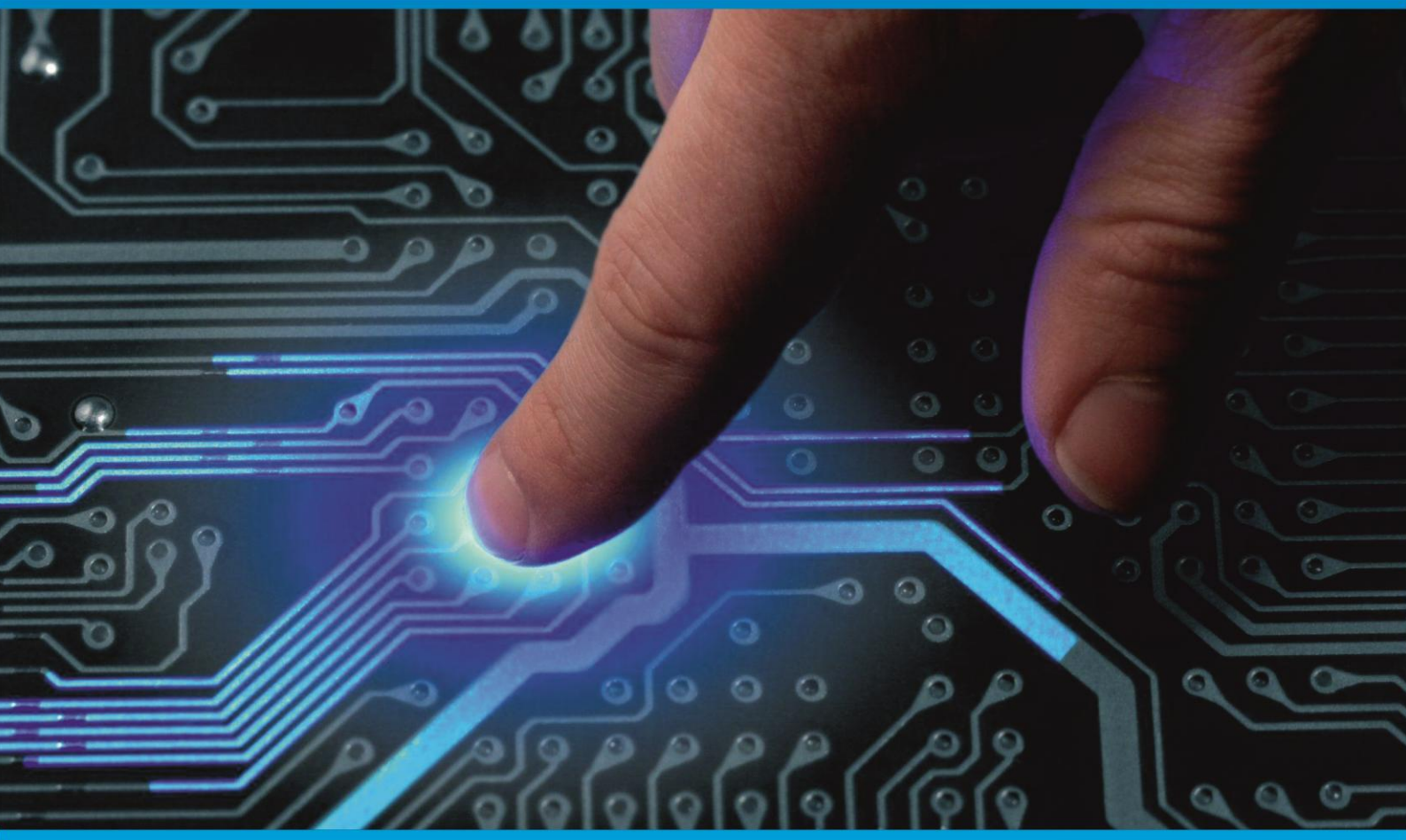
IJIRCCE

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



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903050|

Insider Threat Detection Using Ml And Dl For Cyber Security

Mrs.S.Kavitha[1], Devi S[2], Mahalakshmi M[3], Nivethitha [N4]

Assistant Professor, Dept. of CSE., Velammal College of Engineering and Technology, Viraganoor, Madurai, India[1]

UG Student, Dept. of CSE., Velammal College of Engineering and Technology,Viraganoor, Madurai, India[2,3,4]

ABSTRACT: The development of the Internet, cyber-attacks are changing rapidly and the cyber security situation is not optimistic. This survey report describes key literature surveys on machine learning (ML) and deep learning (DL) methods for network analysis of intrusion detection and provides a brief tutorial description of each ML / DL method. Projects representing each method were indexed, read, and summarized based on their temporal or thermal correlations. Because data are so important in ML / DL methods, we describe some of the commonly used network datasets used in ML / DL, discuss the challenges of using ML / DL for cybersecurity and provide suggestions for research directions.

I.INTRODUCTION

The ML and DL methods covered in this project are applicable to intrusion detection in wired and wireless networks. Readers who wish to focus on wireless network protection can refer to essays such as Soni et al, which focuses more on architectures for intrusion detection systems that have been introduced for MANETs.Security breaches include external intrusions and internal intrusions.

There are three main types of network analysis for IDSs: misuse-based, also known as signature-based, anomaly based, and hybrid. Misuse-based detection techniques aim to detect known attacks by using the signatures of these attacks . They are used for known types of attacks without generating a large number of false alarms. However, administrators often must manually update the database rules and signatures. New (zero-day) attacks cannot be detected based on misused technologies. Anomaly-based techniques study the normal network and system behavior and identify anomalies as deviations from normal behavior. They are appealing because of their capacity to detect zero-day attacks.

Another advantage is that the profiles of normal activity are customized for every system, application, or network, therefore making it difficult for attackers to know which activities they can perform undetected. Additionally, the data on which anomalybased techniques alert (novel attacks) can be used to define the signatures for misuse detectors. The main disadvantage of anomaly-based techniques is the potential for high false alarm rates because previously unseen system behaviors can be categorized as anomalies. Hybrid detection combines misuse and anomaly detection. It is used to increase the detection rate of known intrusions and to reduce the false positive rate of unknown attacks. Most ML / DL methods are hybrids.

II.RELATED WORK

A network security system consists of a network security system and a computer security system. Each of these systems includes firewalls, antivirus software, and intrusion detection systems (IDS). IDSs help discover, determine and identify unauthorized system behavior such as use, copying, modification and destruction.The purpose of this project is for those who want to study network intrusion detection in ML/DL.Thus, great emphasis is placed on a thorough description of the ML/DL methods, and references to seminal works for each ML and DL method are provided. Examples are provided concerning how the techniques were used in cyber security.Similarly to ML methods, DL methods also have supervised learning and unsupervised learning. Learning models built under different learning frameworks are quite different. The benefit of DL is the use of unsupervised or semi-supervised feature learning.The performance of most ML algorithms

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

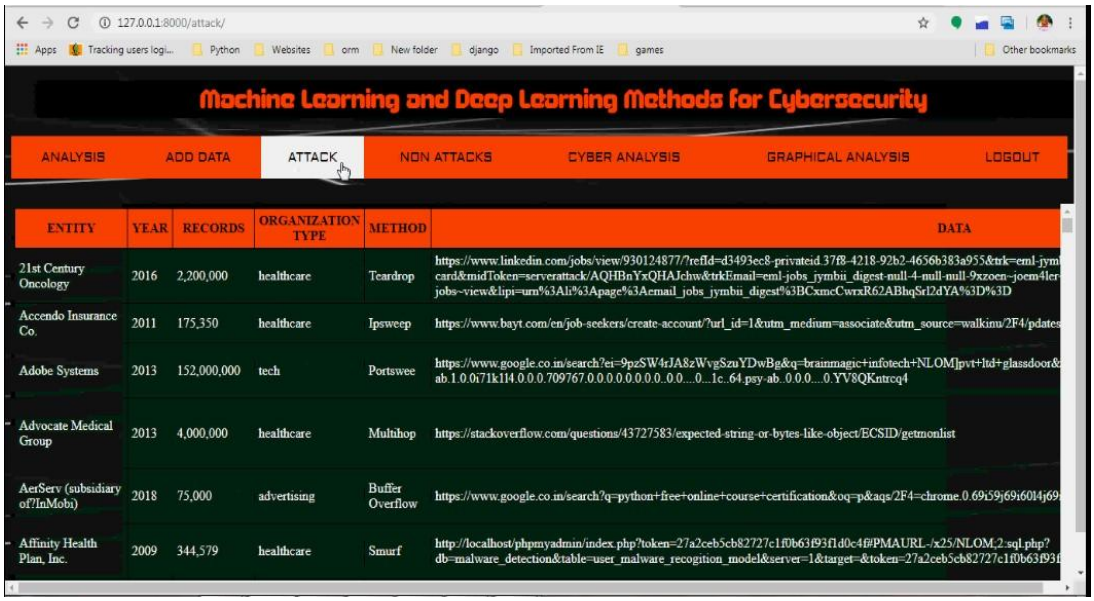
|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903050|

depends upon the accuracy of the features extracted.Dataset Intrusion Detection Dataset was applied to the 3rd International Knowledge Discovery and Data Mining Tools Contest. This model identifies features between intrusive and normal connections for building network intrusion detectors. In the NSL-KDD dataset, each instance has the characteristics of a type of network data. It contains 22 different attack types grouped into 4 major attack types.Dos back, Neptune, smurf, teardrop, land, pod Probe Satan, portsweep, ipsweep, nmap. R2L Warezmaster, warezclient, ftpwrite, guesspassword, imap, multihop, phf, spy U2R Rootkit, butteroverflow, loadmodule, perl.Here, we compare the classification performance of SVM with other popular machine learning algorithms. We have selected several popular classification algorithms. For all algorithms, we attempt to use multiple sets of parameters to maximize the performance of each algorithm. Using SVM algorithms classification for malware bag-of-words weightage.

III.DATASET DESCRIPTION

The statistical analysis showed that there are Important issues in the data set which highly Affects the performance of the systems, and Results in a very poor estimation of anomaly Detection approaches. To solve these issues, A new data set as,NSL-KDD is Proposed, which consists of selected records Of the complete KDD data set. The Advantage of NSL KDD dataset are 1.No redundant records in the train set, So the classifier will not produce any Biased result 1.No duplicate record in the test set Which have better reduction rates.The number of selected records from Each difficult level group is inversely Proportional to the percentage of Records in the original KDD data set.The training dataset is made up of 21 Different attacks out of the 37 present in the test dataset. The known attack types are those present in the training dataset While the novel attacks are the additional Attacks in the test dataset i.e. not Available in the training datasets. The Attack types are grouped into four Categories: DoS, Probe, U2R and R2L.



IV.PROPOSED ALGORITHM

The concept of DL was proposed by Hinton et al. Based on the deep belief network (DBN), in which an unsupervised greedy layer-by-layer training algorithm is proposed that provides hope for solving the optimization problem of deep structure. Then the deep structure of a multi-layer automatic encoder is proposed. In addition, the convolution neural network proposed by Lecun et al. Is the first real multi-layer structure learning algorithm that uses a space relative relationship to reduce the number of parameters to improve the training performance.Feature processing is the process of putting domain knowledge into a feature extractor to reduce the complexity of the data and generate patterns that make learning algorithms work better.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903050|

Feature processing is time-consuming and requires specialized knowledge. In ML, most of the characteristics of an application must be determined by an expert and then encoded as a data type. Features can be pixel values, shapes, textures, locations, and orientations.

The performance of most ML algorithms depends upon the accuracy of the features extracted.Dataset Intrusion Detection Dataset was applied to the 3rd International Knowledge Discovery and Data Mining Tools Contest. This model identifies features between intrusive and normal connections for building network intrusion detectors. In the NSL-KDD dataset, each instance has the characteristics of a type of network data. It contains 22 different attack types grouped into 4 major attack types.Dos back, Neptune, smurf, teardrop, land, pod Probe Satan, portsweep, ipsweep, nmap. R2L Warezmaster, warezclient, ftpwrite, guesspassword, imap, multihop, phf, spy U2R Rootkit, butteroverflow, loadmodule, perl.

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- ➢ What data should be given as input?
- ➢ How the data should be arranged or coded?
- ➢ The dialog to guide the operating personnel in providing input.
- ➢ Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903050|

3. Create document, report, or other formats that contain information produced by the system.
4. The output form of an information system should accomplish one or more of the following objectives.

❖Convey information about past activities, current status or projections of the

❖Future.

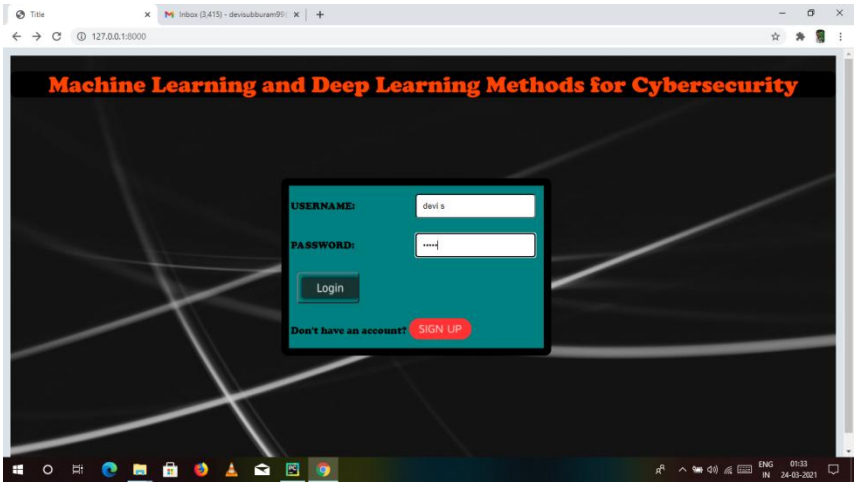
❖Signal important events, opportunities, problems, or warnings.

❖Trigger an action.

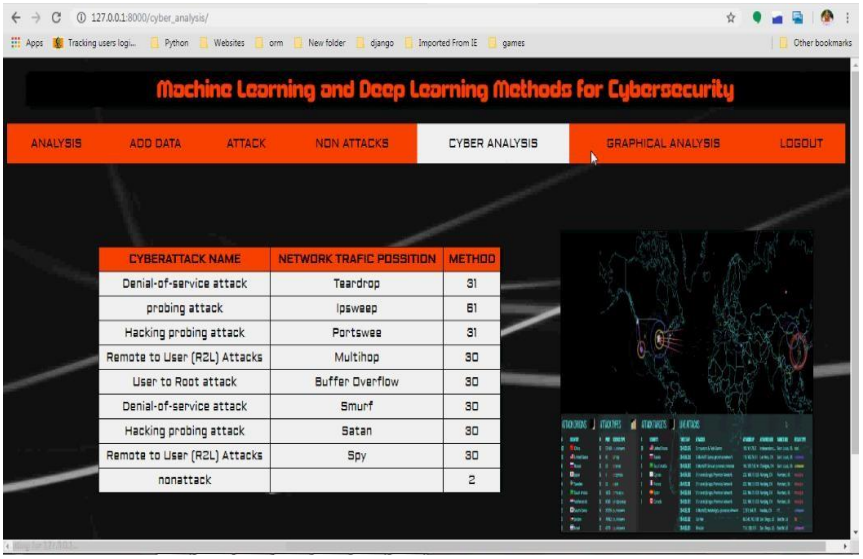
❖Confirm an action.

V.SIMULATION OF RESULTS

User login



Attack – cyber analysis

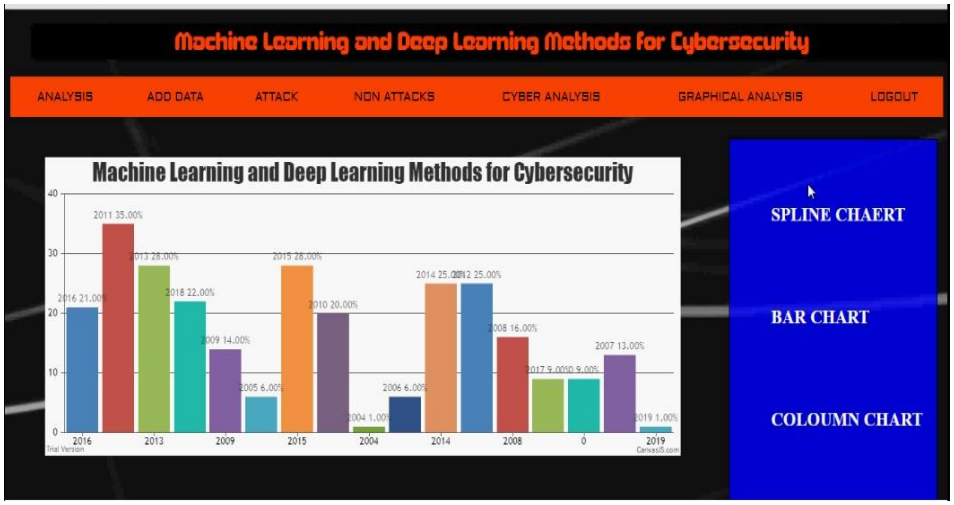


Graphical analysis

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903050|



VI.CONCLUSION AND FUTURE WORK

This project presents a literature review of ML and DL methods for network security. The project, which has mostly focused on the last three years, introduces the latest applications of ML and DL in the field of intrusion detection. Unfortunately, the most effective method of intrusion detection has not yet been established. Each approach to implementing an intrusion detection system has its own advantages and disadvantages, a point apparent from the discussion of comparisons among the various methods. Thus, it is difficult to choose a particular method to implement an intrusion detection system over the others. Datasets for network intrusion detection are very important for training and testing systems.

The ML and DL methods do not work without representative data, and obtaining such a dataset is difficult and time-consuming. However, there are many problems with the existing public dataset, such as uneven data, outdated content and the like. These problems have largely limited the development of research in this area. Network information update very fast, which brings to the DL and ML model training and use with difficulty, model needs to be retrained long-term and quickly. So incremental learning and lifelong learning will be the focus in the study of this field in the future.

REFERENCES

[1] S. Aftergood, "Cybersecurity: The cold war online," Nature, vol. 547, no. 7661, p. 30, 2017.
[2] A. Milenkoski, M. Vieira, S. Kounev, A. Avritzer, and B. D. Payne, "Evaluating Computer Intrusion Detection Systems:A Survey of Common Practices," AcmComput. Surv., vol. 48, no. 1, pp. 1–41, 2015.
[3] C. N. Modi and K. Acha, "Virtualization layer security challenges and intrusion detection/prevention systems in cloud computing: a comprehensive review," J. Supercomput., vol. 73, no. 3, pp. 1–43, 2016.
[4] E. Viegas, A. O. Santin, A. França, R. Jasinski, V. A. Pedroni, and L. S. Oliveira, "Towards an Energy-Efficient Anomaly-Based Intrusion Detection Engine for Embedded Systems," IEEE Trans. Comput., vol. 66, no. 1, pp. 163–177, 2017.
[5] A. Patcha and J. M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," Comput. Netw., vol. 51, no. 12, pp. 3448– 3470, 2007.
[6] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "Review: A survey of intrusion detection techniques in Cloud," J. Netw. Comput. Appl., vol. 36, no. 1, pp. 42–57, 2013.
[7] S. Revathi and A. Malathi, "A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection," in International Journal of Engineering Research and Technology, 2013.
[8] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL Detection using Machine Learning: A Survey," arXiv:1701.07179, 2017.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903050|

[9] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Commun. Surv. Tutor., vol. 18, no. 2, pp. 1153–1176, 2016.

[10] M. Soni, M. Ahirwa, and S. Agrawal, "A Survey on Intrusion Detection Techniques in MANET," in International Conference on Computational Intelligence and Communication Networks, 2016, pp. 1027–1032.

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details