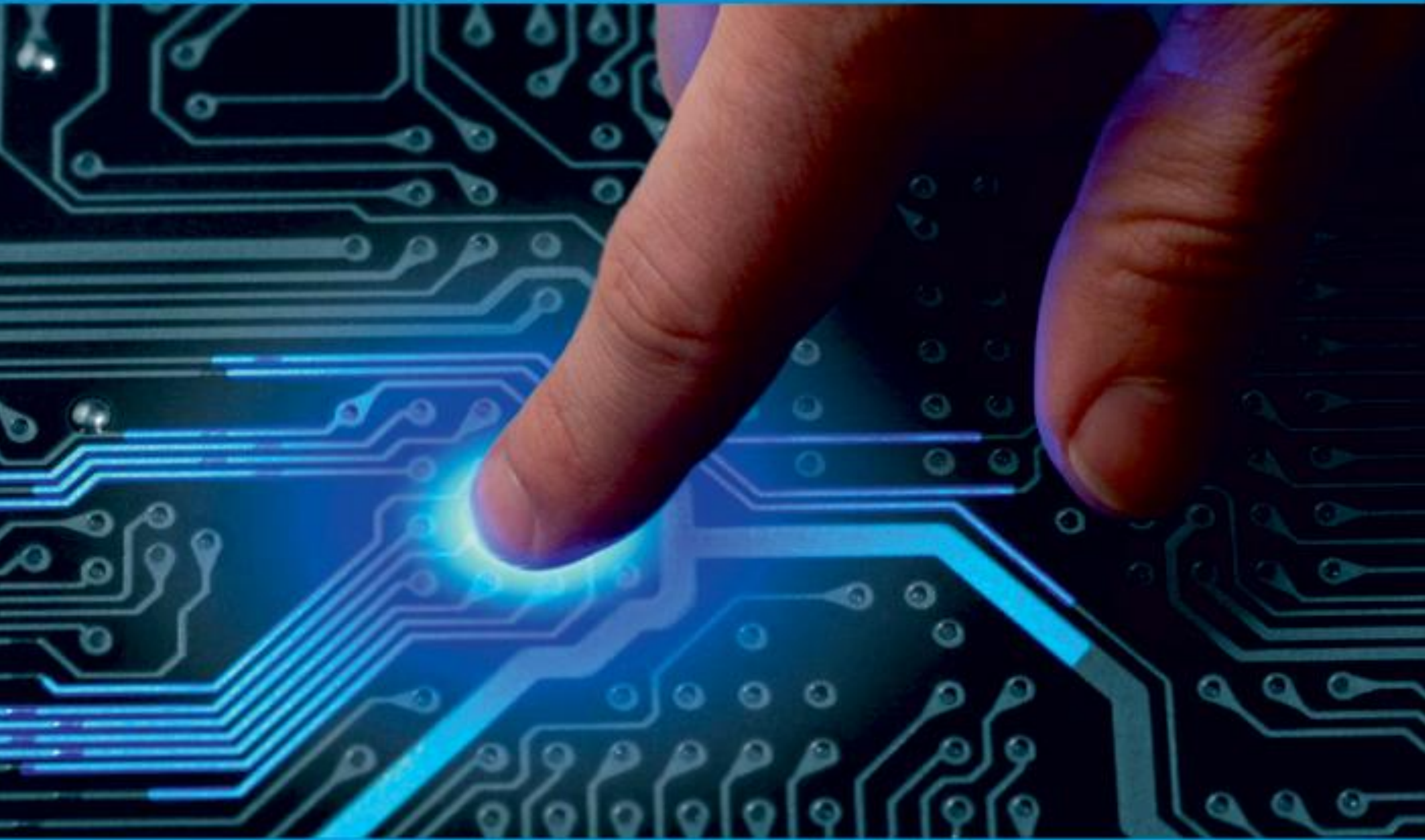
IJIRCCE

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.2022.1004145|

Detecting Phishing Website Using Machine Learning

Khandelwal Tushar Pradipkumar, Kumari Stuti S.N. Sharma, Aakash Haribhau Jamdhade

Sorte Manjusha Hiralal, Prof.S.A.Bahir

Department of Computer Engineering, Sinhgad Academy of Engineering Pune , India [1 2 3]

ABSTRACT: Phishing attacks are still a major threat to computer system defenders, as they are frequently the first step in a multi-stage attack.Although phishing detection has advanced significantly, some phishing emails appear to get past filters by making minor structural and semantic changes to the messages. We'll deal with it.Using a machine learning classifier trained on a large corpus of phishing and legitimate emails, researchers were able to solve the problem. We createSAFEPC (Semi-Automated Feature Generation for Phish Classification) is a system that extracts features, some of which are elevated to higher level features, in order to defeat common phishing email detection.strategies. To test SAFE-PC, we gathered a large number of phishing emails from a tier-1 university's central IT department.SAFE-implementation PC's. Phishing attacks continue to pose a major threat for computer system defenders, often forming the first step in a multi-stage attack. There have been great strides made in phishing detection; however, some phishing emails appear to pass through filters by making simple structural and semantic changes to the messages. We tackle this problem through the use of a machine learning classifier operating on a large corpus of phishing and legitimate emails. We design SAFEPC (Semi-Automated Feature generation for Phish Classification), a system to extract features, elevating some to higher level features, that are meant to defeat common phishing email detection strategies. To evaluate SAFE-PC, we collect a large corpus of phishing emails from the central IT organization at a tier-1 university. The execution of SAFE-PC on the dataset exposes hitherto unknown insights on phishing campaigns directed at university users. SAFEPC detects more than 70a state-of-the-art email filtering tool. It also outperforms Spam Assassin, a commonly used email filtering tool. We also developed an online version of SAFE-PC, that can be incrementally retrained with new samples. Its detection performance improves with time as new samples are collected, while the time to retrain the classifier stays constant

KEYWORDS: -Machine learning, Artificial intelligence

I. INTRODUCTION

Phishing is defined as impersonating a legitimate website in order to deceive users by stealing personal information such as usernames, passwords, account numbers, and social security numbers. Phishing scams are the most common type of cybercrime today. Phishing attacks can happen in a variety of places, including online. File hosting or cloud storage, payment sector, webmail, and financial institution numerous others Phishing attacks plagued the webmail and online payment sectors.

more than any other sector of the economy Phishing can take the form of email phishing scams or spear phishing, so users should be aware of the risks. Should not place their complete trust in common security software. Machine Learning is one of the most effective methods for detecting phishing because it eliminates drawbacks.

II. PROBLEM STATEMENT

Phishing is a new type of network attack in which an attacker creates a replica of an existing web page in order to trick users (for example, through specially designed e-mails or instant messages) into submitting personal, financial, or password data to what they believe is a legitimate website.

They believe it is the website of the service provider.URL phishing detection and prevention attacks.

III. MOTIVATION

Phishing is a new type of network attack in which an attacker creates a replica of an existing web page in order to trick users into submitting personal, financial, or password data to what they believe is a legitimate website (for example, through specially designed e-mails or instant messages).They believe it is the service provider's website.Detection and prevention of URL phishingattacks.

IV. OBJECTIVES

Victims suffer financial losses and identity theft as a result of these e-mails. In this study, "Anti Phishing Simulator" software was developed to provide information about the phishing detection problem and how to detect phishing emails. By examining the contents of emails, this software detects phishing and spam emails. A Bayesian algorithm is used to classify spam words added to the database.

V. LITERATURE REVIEW

Muhammet Baykara, Zahit Ziya Gurel,” - Detection of phishing attacks” [1]: Phishing is a form of cybercrime where an attacker imitates a real person / institution by promoting them as an official person or entity through e-mail or other communication mediums. In this type of cyber attack, the attacker sends malicious links or attachments through phishing e-mails that can perform various functions, including capturing the login credentials or account information of the victim. These e-mails harm victims because of money loss and identity theft. In this study, a software called ”Anti Phishing Simulator” was developed, giving information about the detection problem of phishing and how to detect phishing emails. With this software, phishing and spam mails are detected by examining mail contents. Classification of spam words added to the database by Bayesian algorithm is provided.

Vyacheslav Lyashenko, Oleg Kobylin, Mykyta Minenko “Tools for Investigating the Phishing Attacks Dynamics” [2], We are exploring new ways to analyze phishing attacks. To do this, we investigate the change in the dynamics of the power of phishing attacks. We also analyze the effectiveness of detection of phishing attacks. We are considering the possibility of using new tools for analyzing phishing attacks. As such tools, the methods of chaos theory and the ideology of wavelet coherence are used. The use of such analysis tools makes it possible to investigate the peculiarities of the phishing attacks occurrence, as well as methods for their identification effectiveness. This allows you to expand the scope of the analysis of phishing attacks. For analysis, we use real data about phishing attacks.

Narayana Darapaneni; Aruna Kumari Evoori; Vijaya Babu Vemuri; Thangaselvi Arichandrapandian; G Karthikeyan,” Automatic Face Detection and Recognition for Attendance Maintenance”,[3This study is interested in creating an attendance tracking system based on deep learning. Every second advances a fresh perspective in the modern world, which is focused on artificial intelligence. Artificial Intelligence is one of the most rapidly developing fields when it comes to face recognition (AI). Instead of employing the established methods for keeping track of attendance, we recommend implementing a facial recognition programme that classifies human faces based on their distinguishing facial features. In order to find and locate all faces, independent of their size, scale, orientation, illumination, expression, or other characteristics, the technique of face detection is necessary. To solve the problem of producing an improved attendance system with less maintenance, cheaper hardware costs, and increased performance and accuracy, we utilised YOLO, MTCNN, and FaceNet embeddings, as well as numerous augmentations, picture quality checks, and de-noise approaches.

Nathezhtha.T1 , Sangeetha.D2 ,Vaidehi.V3 “WC-PAD: Web Crawling based Phishing Attack Detection”,[4] Phishing is a criminal offense which involves theft of user’s sensitive data. The phishing websites target individuals, organizations, the cloud storage hosting sites and government websites. Currently, hardware based approaches for antiphishing is widely used but due to the cost and operational factors software based approaches are preferred. The existing phishing detection approaches fails to provide solution to problem like zero-day phishing website attacks. To overcome these issues and precisely detect phishing occurrence a three phase attack detection named as Web Crawler based Phishing Attack Detector(WC-PAD) has been proposed. It takes the web traffics, web content and Uniform Resource Locator(URL) as input features, based on these features classification of phishing and non phishing websites are done. The experimental analysis of the proposed WC-PAD is done with datasets collected from real phishing cases. From the experimental results, it is found that the proposed WC-PAD gives 98.9phishing and zero-day phishing attack detection

Ivan Ortiz-Garc ´ es´ , Roberto O. Andrade†, and Mar´ıa Cazares‡ ” Detection of Phishing Attacks with Machine Learning Techniques in Cognitive Security Architecture”[5] Ivan Ortiz-Garc ´ es´ , Roberto O. Andrade†, and Mar´ıa Cazares‡ Abstract: The number of phishing attacks has increased in Latin America, exceeding the operational skills of cybersecurity analysts. The cognitive security application proposes the use of bigdata, machine learning, and data analytics to improve response times in attack detection. This paper presents an investigation about the analysis of anomalous behavior related with phishing web attacks and how machine learning techniques can be an option to face the problem. This analysis is made with the use of an contaminated data sets, and python tools for developing machine learning for detect phishing attacks through of the analysis of URLs to determinate if are good or bad URLs in base of specific characteristics of the URLs, with the goal of provide realtime information for take proactive decisions that minimize the impact of an attack.

VI. SYSTEM ANALYSIS SYSTEM ARCHITECTIRE

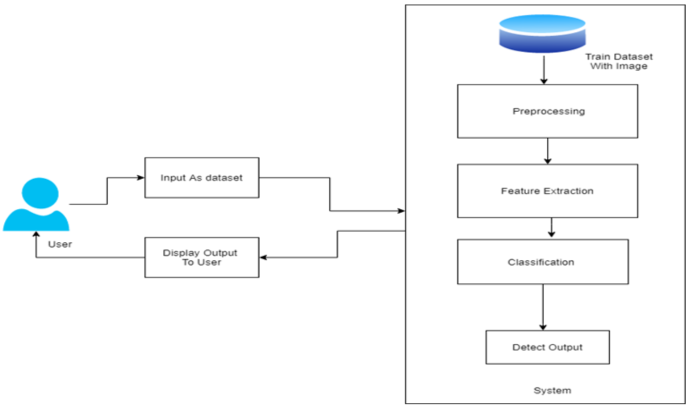


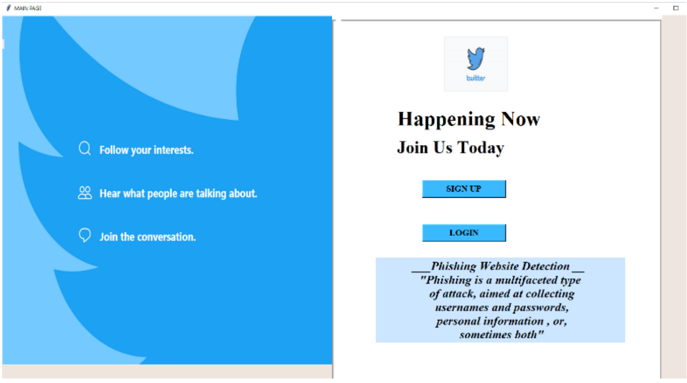
Fig. System Architecture

ALGORITHM

SVM:-

Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. “Support Vector Machine” (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems.

Results:



Follow your interests.

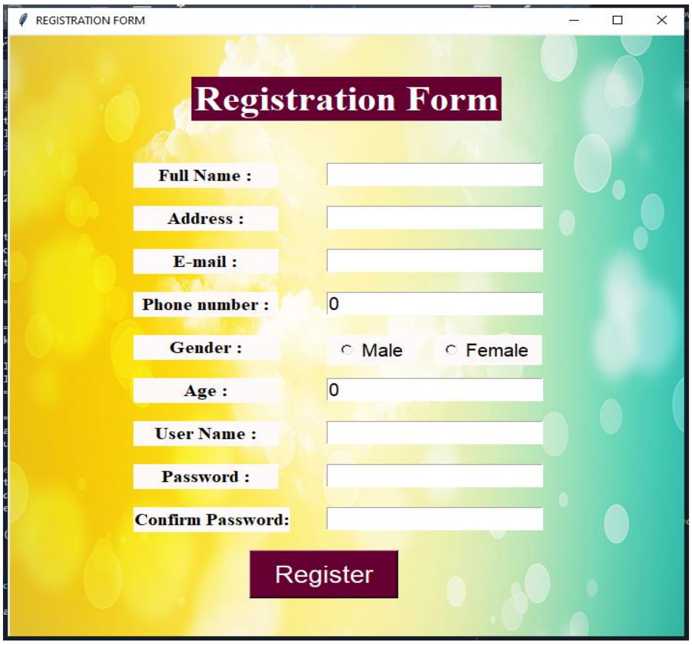
Hear what people are talking about.

Join the conversation.

Happening Now

Join Us Today

___Phishing Website Detection __
"Phishing is a multifaceted type
of attack, aimed at collecting
usernames and passwords,
personal information , or,
sometimes both"



Full Name :

Address :

E-mail :

Phone number : 0

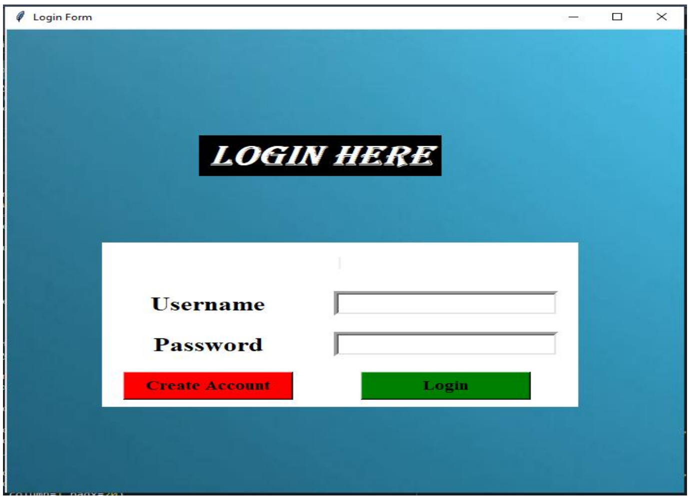
Gender : ○ Male ○ Female

Age : 0

User Name :

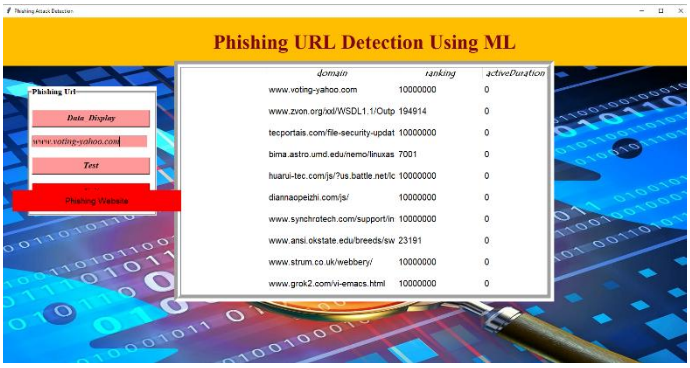
Password :

Confirm Password:



Username

Password



www.voting-yahoo.com

Phishing Website

domain	ranking	activeDuration
www.voting-yahoo.com	10000000	0
www.zvon.org/xxl/WSDL1.1/Outp	194914	0
tecportais.com/file-security-updat	10000000	0
bima.astro.umd.edu/nemo/linuxas	7001	0
huarui-tec.com/js/?us.battle.net/lc	10000000	0
diannaopeizhi.com/js/	10000000	0
www.synchrotech.com/support/in	10000000	0
www.ansi.okstate.edu/breeds/sw	23191	0
www.strum.co.uk/webbery/	10000000	0
www.grok2.com/vi-emacs.html	10000000	0

VII. CONCLUSION

Phishing has evolved into a serious network security issue that has resulted in financial losses for both consumers and e-commerce businesses.We discussed the implemented system in this paper using the linkguard algorithm.

REFERENCES

[1] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, ''The challenge of non-technical loss detection using artificial intelligence: A survey,'' Int. J. Comput. Intell. Syst., vol. 10, no. 1, pp. 760–775, 2016.
[2] M. A. Tebbi and B. Haddad, ''Artificial intelligence systems for rainy areas detection and convective cells' delineation for the south shore of Mediterranean Sea during day and nighttime using MSG satellite images,'' Atmos. Res., vols. 178–179, pp. 380–392, Sep. 2016.
[3] P. Dai, X. Wang, W. Zhang, P. Zhang, and W. You, ''Implicit relative attribute enabled cross-modality hashing for face image-video retrieval,'' Multimedia Tools Appl., vol. 77, no. 18, pp. 23547–23577, Sep. 2018.
[4] Q. W. Wang and Z. L. Ying, ''A face detection algorithm based on Haarlike t features,'' Pattern Recognit. Artif. Intell., vol. 28, no. 1, pp. 35–41, 2015.
[5] M. Sepandi, M. Taghdir, and A. Rezaianzadeh, ''Assessing breast cancer risk with an artificial neural network,'' Asian Pacific J. Cancer Prevention, vol. 19, no. 4, pp. 1017–1019, 2018.
[6] Y.-H. Lai and C.-K. Yang, ''Video object retrieval by trajectory and appearance,'' IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 6, pp. 1026–1037, Jun. 2015.
[7] A. K. G. Worner, ''Realtime quality monitoring of compressed video signals,'' SMPTE J., vol. 111, no. 9, pp. 373–377, Sep. 2002.
[8] D. Saravanan and S. Srinivasan, ''Video data mining information retrieval using BIRCH clustering technique,'' in Advances in Intelligent Systems and Computing, vol. 325. New Delhi, India: Springer, 2015, pp. 583–594.
[9] M. Okabe, Y. Dobashi, and K. Anjyo, ''Animating pictures of water scenes using video retrieval,'' Vis. Comput., vol. 34, no. 3, pp. 347–358, Mar. 2018.
[10] Y. Li, R. Wang, Z. Cui, S. Shan, and X. Chen, ''Spatial pyramid covariancebased compact video code for robust face retrieval in TV-series,'' IEEE Trans. Image Process., vol. 25, no. 12, pp. 5905–5919, Dec. 2016.
[11] S. Jairath, S. Bharadwaj, and M. Vatsa, ''Adaptive skin color model to improve video face detection,'' in Advances in Intelligent Systems and Computing, vol. 390. New Delhi, India: Springer, 2016, pp. 131–142.
[12] M. Chouchene, F. E. Sayadi, H. Bahri, J. Dubois, J. Miteran, and M. Atri, ''Optimized parallel implementation of face detection based on GPU component,'' Microprocessors Microsyst., vol. 39, no. 6, pp. 393–404, Aug. 2015.
[13] O. Dospinescu and I. Popa, ''Face detection and face recognition in Android mobile applications,'' Inf. Economica, vol. 20, no. 1, pp. 20–28, 2016.
[14] A. E. Sergeev, A. S. Konushin, and V. S. Konushin, ''Reducing background false positives for face detection in surveillance feeds,'' Comput. Opt., vol. 40, no. 6, pp. 958–967, 2016.
[15] Q. Chen, L. Yang, D. Zhang, Y. Shen, and S. Huang, ''Face deduplication in video surveillance,'' Int. J. Pattern Recognit. Artif. Intell., vol. 32, no. 03, Mar. 2018, Art. no. 1856001.
[16] A. Alotaibi and A. Mahmood, ''Deep face liveness detection based on nonlinear diffusion using convolution neural network,'' Signal, Image Video Process., vol. 11, no. 4, pp. 713–720, May 2017.
[17] K. Patel, H. Han, and A. K. Jain, ''Secure face unlock: Spoof detection on smartphones,'' IEEE Trans. Inf. Forensics Security, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.
[18] M. A. Haque, K. Nasrollahi, T. B. Moeslund, and R. Irani, ''Facial videobased detection of physical fatigue for maximal muscle activity,'' IET Comput. Vis., vol. 10, no. 4, pp. 323–330, Jun. 2016.
[19] S. Merat and W. Almuhtadi, ''Artificial intelligence application for improving cyber-security acquirement,'' in Proc. IEEE 28th Can. Conf. Electr. Comput. Eng. (CCECE), May 2015, pp. 1445–1450.
[20] M. Žarković and Z. Stojković, ''Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics,'' Electr. Power Syst. Res., vol. 149, pp. 125–136, Aug. 2017.

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