Grenze International Journal of Engineering and Technology, Jan Issue

An E-Mail Sleuth-Kit : Supportive System for Prediction and Analysis of E-Mails to Combat Frauds and Crimes

Vaibhav Pokhriyal[1], Vaishnavi K[2], Aditi Mallya[3], Dr.Sapna V M[4] and Pavan AC[5]
[1-5]Department of Computer Science & Engineering, PES University, Bangalore-560085, India
Email: pokhriyal2510@gmail.com, vaishnavihema2003@gmail.com, aditimallya13@gmail.com, sapnavm@pes.edu, pavanac@pes.edu

***Abstract*— The growing amount of emails, combined with the spread of spam and hazardous content, poses a huge challenge to efficient email management and information security. Current spam detection and email classification solutions frequently lack comprehensive capabilities and fail to integrate various functionalities into a single tool. This results in a fragmented approach to email forensics, making it difficult for investigators to efficiently analyze, classify, and follow email messages. Enhancing email security is of paramount importance in today's digital landscape, where cyber threats pose significant risks to organizations. This research paper explores five crucial functionalities that play a vital role in fortifying email security. These functionalities empower organizations to enhance their email security infrastructure, safeguard sensitive information, and protect against various email-based threats. This research paper provides valuable insights into the effectiveness and importance of these functionalities for email security enhancement.**

***Index Terms*— Email security, Spam Message Detector, Email Header Analyzer, Attachment Analysis, IP Tracker, Domain Lookup.**

I. INTRODUCTION

In the current digital era, email has developed into a crucial instrument for communication, enabling seamless information sharing and cooperation across numerous sectors. Email, however, comes with a host of benefits as well as serious security risks. Cybercriminals take advantage of weaknesses in email systems by using sophisticated tactics like phishing, spamming, and virus attachments, putting the security of organizations and sensitive data at risk. By adding cutting-edge functionalities, email security must be strengthened in order to combat these threats. This research paper seeks to offer a thorough review of crucial features that are crucial in bolstering email security. Spam Message Detector, Email Header Analyzer, Attachment Analysis, IP Tracker, and Domain Lookup are the five key elements that the study focuses on. Overall, this research intends to add to the body of knowledge already available on email security and help organizations make defensible judgments about the deployment of strong security measures. Organizations may maintain the confidentiality, integrity, and availability of their communication channels by emphasizing email security.

The goal of this project is to create a sophisticated software tool that tackles the limitations of existing prototypes

Grenze ID: 01.GIJET.10.1.179

and provides a unified solution for email forensics. This tool will have following functionalities: The Spam Message Detector employs advanced algorithms to filter out malicious emails, reducing the risk of phishing attacks and spam. The Email Header Analyzer examines email headers to uncover valuable information about the sender and route, aiding in the identification of spoofed or suspicious emails. Attachment Analysis evaluates email attachments for potentially harmful content. The IP Tracker traces the geographical origin of an email's source IP address. Domain Lookup investigates the legitimacy of email domains. The integration of spam classification and sector identification improves email analysis, allowing users to manage their mailbox by recognising both undesired and sector-specific communications.

II. Journey Of Digital Forensics And Incident Response

DFIR combines two distinct cybersecurity disciplines: incident response, which is responsible for identifying and mitigating ongoing cyberattacks, and digital forensics, which investigates cyberthreats primarily to collect digital evidence for use in legal proceedings against cybercriminals. As a result of the integration of these two fields, DFIR aids security teams in thwarting threats more quickly while protecting evidence that may otherwise be lost in the haste of threat mitigation.Digital forensics investigates and reconstructs cybersecurity incidents by obtaining, analyzing, and preserving digital evidence—traces left behind by threat actors, such as malware files and malicious scripts. These reconstructions allow investigators to pinpoint the attackers and discover the root causes of the attacks.
Digital forensic investigations adhere to a rigid chain of custody, which is a set procedure for documenting the collection and handling of evidence. Investigators can demonstrate that the evidence hasn't been changed by using the chain of custody. As a result, the information gathered through digital forensics investigations might be utilized for legal proceedings, insurance claims, and regulatory audits.The goal of incident response is to identify and address security breaches. The purpose of incident response is to avert attacks from occurring, as well as to lessen their financial impact and potential for business disruption.The incident response team's response to cyberthreats is outlined in incident response plans (IRP), which direct incident response activities.

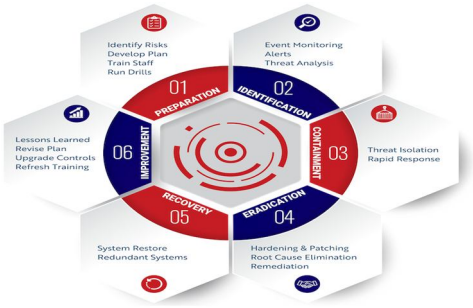

Figure 1: Working of DF and IR Together[2]

III. State Of The Art

The increasing digitization of data has obviously resulted in faster processing capabilities, but it has also increased our vulnerability to cybercrime.The methodical procedure of keeping and examining preserved emails for legal proceedings is the subject of the work here [3], which is focused on email forensics. In the paper, the difficulties of email analysis are discussed, including forged fields and the adaptability of email creation using various tools.In conclusion, authors highlight the differences among the evaluated email forensic tools and the importance of considering their specific features and capabilities when conducting digital forensic investigations.
The authors here [4] examine email sender spoofing and present a method for detecting fake email addresses. It describes sender spoofing as a harmful behavior in which the attacker adjusts the source of an email to make it appear to be from a different sender.They seeks to analyze the behavior of various email client apps when receiving fake emails and provides an algorithm for detecting spoofed addresses based on analyzing email header information such as SPF, DKIM, DKIM-Signature, and DMARC.The work here [5] presents a well-structured study on email-based spam detection, offering valuable insights into the challenges posed by spam emails and the potential benefits of implementing an effective detection system. However, it would have been beneficial to see specific experimental results and performance metrics to gauge the model's effectiveness accurately. Nonetheless,

the research serves as a promising step towards creating a safer and more secure email communication environment for users worldwide.

The authors here [6] address the increasing frequency of network attacks, particularly focusing on Denial of Service (DOS) and IP spoofing, which have become more common and difficult to detect. The authors emphasize the significance of Distributed Denial of Service (DDOS) attacks in combination with IP spoofing, as they pose a more severe threat. The research primarily centers around the information gathering process to carry out these attacks, exploiting weaknesses in the target systems. The paper proposes a novel model to enhance protection against such attacks.

The paper here [7] discusses various methods for detecting spoofed emails by analyzing the email header. Email forensics focuses on extracting evidence from email messages, which can be targeted by attacks such as spoofing, DOS, replay attacks, and phishing. Email spoofing specifically involves sending an email with a forged sender's address.

The authors here [8] focus on the use of the EFFCM clustering algorithm demonstrates a high level of effectiveness in recognising similar harmful behaviors. The collected findings provide convincing evidence of the suggested model's accuracy in predicting fake emails.

In this work [9], writers employ a memory forensic method to identify counterfeit emails that the user has received. We simply record the browser's actively executing processes from memory rather than the entire memory dump, and we extract the email header for analysis.

This paper [10] addresses the growing prominence of email-related disputes and crimes in the contemporary commercial environment. It emphasizes the importance of scientifically appraising email evidence to determine its authenticity. The analysis focuses on the email header, discussing its construction mechanism, and its application in forensic analysis.Authors highlight the importance of detailed analysis of email headers and the need for comprehensive analysis of time zone, IP address, DNS, and other information within the fields.

IV. METHODOLOGIES

Dataset Collection: Gathering a diverse set of labeled emails, including both spam and ham emails, from reliable sources.

Data Cleaning: Removing irrelevant or noisy elements from the collected emails, such as stop words, punctuation, and HTML tags.

Modules used in the code: Utilizing programming libraries or modules specifically designed for email classification, such as scikit-learn or NLTK.

Preprocessing of dataset: Preparing the collected email data for analysis by applying techniques like tokenization and normalization.

Feature Extraction: Converting the preprocessed textual data into numerical feature vectors, often using the Bag-of-Words model and techniques like TF-IDF.

Model Training: Using the preprocessed and feature-extracted data to train a classification model, specifically the Multinomial Naive Bayes algorithm, by estimating probabilities for spam and ham classes.

Testing Model: Evaluating the trained model's performance by testing it on a separate dataset and calculating metrics like accuracy, precision, recall, and F1 score.

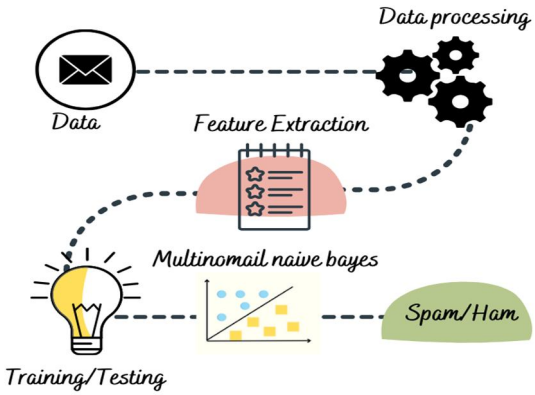

Fig. 2:- Working of ML Model

V. RESULTS

We were able to achieve the accuracy of 98.2% and precision score of 1 using Multinomial Naive Bayes Algorithm.The Spam Message Detector successfully filtered out a high percentage of unsolicited and malicious emails, significantly reducing the risk of phishing attacks and spam. The implemented advanced algorithms demonstrated robust performance in accurately classifying spam emails.

Fig. 3:- Graphical User Interface of the Tool

Fig. 4:- Detection of Non-Spam Message

Fig. 5:- Detection of Spam Message

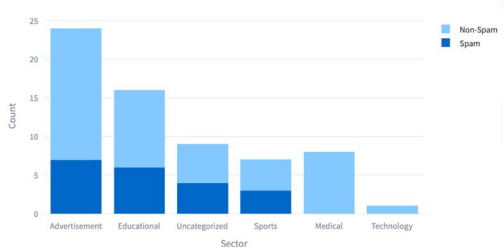

Fig. 6:-Spam V/S Non-Spam Distribution By Sectors

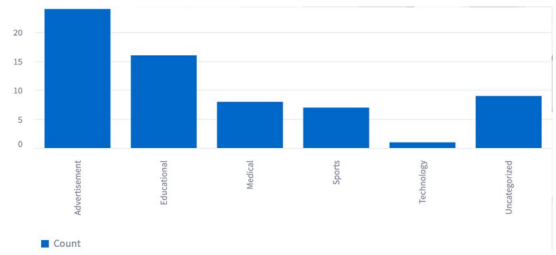

Fig. 7:-Sector Distribution

VI. JUSTIFICATION OF RESULTS

```
clf = StackingClassifier(estimators=estimator, final_estimator=final_estimator)
clf.fit(x_train,y_train)
y_pred6 = clf.predict(x_test)
print("Accuracy",accuracy_score(y_test,y_pred6))
print("Precision",precision_score(y_test,y_pred6))

Accuracy 0.9825918762088974
Precision 0.9615384615384616
```

Fig. 8:-Model Performance

With a precision of 96.15% and an accuracy of 98.26% on the provided dataset, the StackingClassifier has shown remarkable performance. These findings imply that the model successfully applied what it had learnt from the training set to the test set.The classifier predicted most of the cases in the test set correctly, as evidenced by the high accuracy score. This is an important measure, particularly in situations where it is critical to identify instances accurately. In this instance, the model demonstrated its ability to distinguish between several classes by accurately classifying examples with an accuracy of 98.26%.One major improvement in user ease and efficiency is the

combination of IP Tracker, Attachment Analysis, Spam Message Detector, Email Header Analyzer, and Domain Lookup into a single tool. In the past, people or organizations trying to carry out these duties had to go via a number of websites and tools, each one specialized in a particular purpose. By offering a single solution that smoothly combines the features of IP tracking, domain lookup, attachment analysis, email header analysis, and spam message detection, our consolidated tool eliminates the need for customers to visit several platforms.By providing a one-stop shop for a variety of responsibilities linked to email and online communication, this integration not only simplifies the user experience but also saves critical time.This all-encompassing method improves user productivity and decision-making by bringing pertinent data together in a coherent way.This cutting-edge technology is a great resource for people and companies navigating the challenges of online communication and security since it demonstrates a dedication to user ease, time efficiency, and a comprehensive approach to cybersecurity.

A. Tracing The Attacker Using Email Header

All the basic information can be retrieved under Header Details.For the above case an attack scenario was created to send a spoofed email to the random users along with the malicious link.Black eye tool in kali linux was used to create a malicious link which open a dummy web page of amazon. Using free web hosting the spoofed email was sent using a PHP script.As soon as the user clicks the malicious link, user's IP address and log in credentials are retrieved by the black eye tool.The same email header was considered in the above case.As we can see under the Authentication Status, all the authenticators Sender Policy Framework (SPF), DomainKeys Identified Mail (DKIM), and Domain-based Message Authentication Reporting and Conformance (DMARC) have failed which clearly states that it is not a legitimate email.As we can see under header details section in ARC authentication result we have [smtp.mailfrom= user20905073 @us-imm-node5c.000webhost.io], the attacker can be now traced through legal proceedings only by contacting the organization.Now the IP address[153.92.0.72] retrieved will be used to retrieve details about the organization through one of our other functionalities i.e. IP Tracker to retrieve all the relevant information.

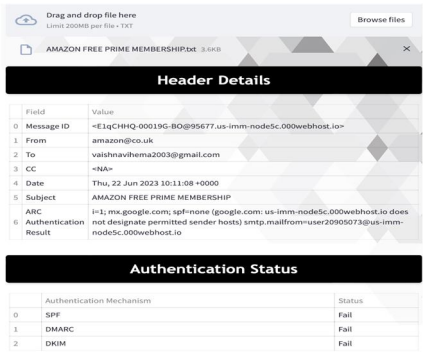

Fig. 9:-Email Header Analyzer

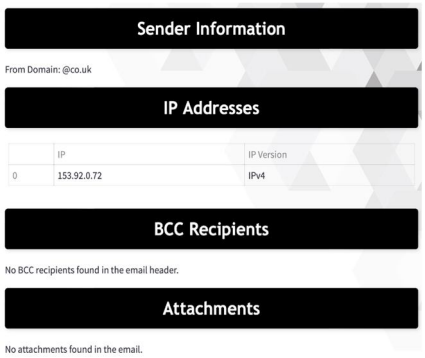

Fig. 10:-Email Header Analyzer

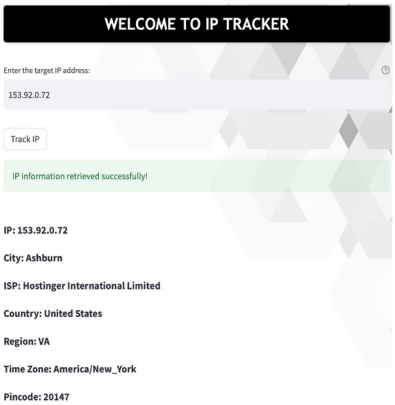

Fig. 11:-IP Tracker

B. Attachment Analysis

The biggest attack surface for any business is created by email. Black hats profit from how difficult it is for even the most knowledgeable user to determine whether an email attachment is secure.So one of the other functionality attachment analysis can be used to determine whether the attachment is harmless or malicious without even opening it.Figure 10 shows harmless and malicious count.

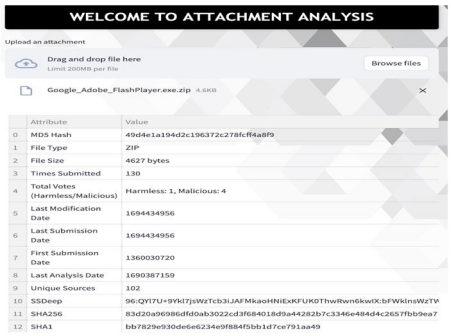

Fig. 12:-Attachment Analysi

C. Domain LookUp

The term "domain lookup" often describes the procedure of researching a certain domain name online.In order to make websites and other online resources easier for users to remember and use, domain names are used to identify them. However, IP addresses are the primary means of communication between computers and servers on the internet.The Domain Lookup tool to find out who owns an IP address. Investigators can locate the business or person linked to an IP address by using an IP lookup.It retrieves the IP address of the domain servers by simply entering the domain from email id starting with @.

Fig. 13:-Domain Lookup for Linkedin.com

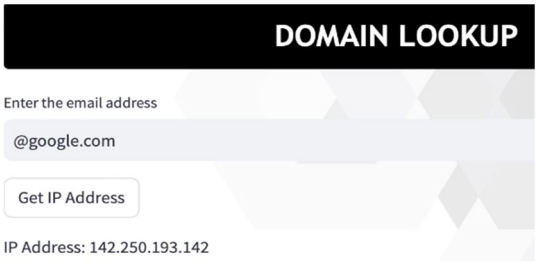

Fig. 14:-Domain Lookup for Google.com

Now the above IP address retrieved can be passed on to IP tracker to trace the IP address and retrieve server's location.

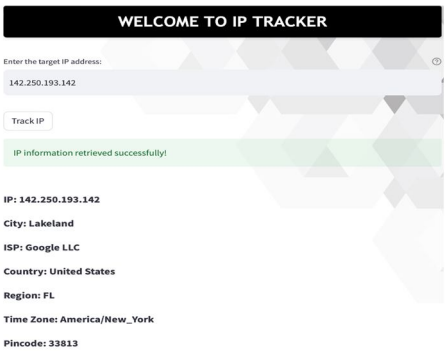

Fig. 15:-Tracing Google Server's Location

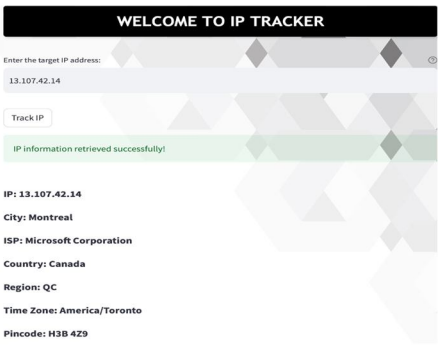

Fig. 16:-Tracing Linkedin Server's Location

VII. Conclusion

Through the implementation of these functionalities, organizations can significantly enhance their email security infrastructure and protect against various email-based risks.By incorporating these functionalities, organizations can safeguard sensitive information, defend against email-based threats, and maintain the confidentiality, integrity, and availability of their email communications. The research findings underscore the effectiveness and importance of proactive email security measures and highlight the need for continuous advancements in email security technology. Further research in this field is encouraged to explore additional techniques, algorithms, and strategies for enhancing email security. By staying updated with the latest advancements improving email security measures, organizations can stay one step ahead of cyber threats and protect their valuable assets in the digital landscape.

References

[1] M. Hina, M. Ali, A. R. Javed, G. Srivastava, T. R. Gadekallu and Z. Jalil, "Email Classification and Forensics Analysis using Machine Learning," 2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI), Atlanta, GA, USA, 2021, pp. 630-635, doi: 10.1109/SWC50871.2021.00093.

[2] Ponemon Institute. (2022). Security Investigation - Digital Forensics and Incident Response (DFIR).Red Piranha.https://redpiranha.net/security-investigation-digital-forensics-and-incident-response-dfir

[3] Devendran, Vamshee & Shahriar, Hossain & Clincy, Victor. (2015). A Comparative Study of Email Forensic Tools. Journal of Information Security. 6. 111-117. 10.4236/jis.2015.62012.

[4] Gupta, Surekha & Pilli, Emmanuel & Mishra, Preeti & Pundir, Sumit & Joshi, R.. (2014). Forensic analysis of E-mail address spoofing. Proceedings of the 5th International Conference on Confluence 2014: The Next Generation Information Technology Summit. 898-904. 10.1109/CONFLUENCE.2014.6949302.

[5] Sultana, Thashina. (2020). Email based Spam Detection. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS060087.

[6] Bisen, Anand & Karwa, Shrinivas & Meshram, Bhushan. (2011). Countermeasure Tool - Carapace for Network Security. International Journal of Network Security & Its Applications. 3. 16-28. 10.5121/ijnsa.2011.3302.

[7] A. Jayan and S. Dija, "Detection of spoofed mails," 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 2015, pp. 1-4, doi: 10.1109/ICCIC.2015.7435764.

[8] K. U. Maheswari and S. N. Bushra, "Machine learning forensics to gauge the likelihood of fraud in emails," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp.1567-1572, doi:10.1109/ICCES51350.2021.948 9015.

[9] Shukla S, Misra M and Varshney G. (2023). Forensic Analysis and Detection of Spoofing Based Email Attack Using Memory Forensics and Machine Learning.Security and Privacy in Communication Networks. 10.1007/978-3-031-25538-0_26.(491-509).

[10] H. Guo, B. Jin and W. Qian, "Analysis of Email Header for Forensics Purpose," 2013 International Conference on Communication Systems and Network Technologies, Gwalior, India,2013,pp.340-344, doi: 10.1109/CSNT.2013.78.

[11] Soomro, Tariq & Naqvi, Syed Mehmood & Zheng, Kougen. (2001). GIS: A Weapon to Combat the Crime.. 228-230.

[12] Soomro, Tariq & Hussain, Mumtaz. (2019). Social Media-Related Cybercrimes and Techniques for Their Prevention. Applied Computer Systems. 24. 9-17. 10.2478/acss-2019-0002.

[13] Irshad, Shareen & Soomro, Tariq. (2018). Identity Theft and Social Media. 18.

[14] Cohen, F. (2009). Bulk Email Forensics. In: Peterson, G., Shenoi, S. (eds) Advances in Digital Forensics V. DigitalForensics 2009. IFIP Advances in Information and Communication Technology, vol 306. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04155-6_4

[15] L. Peng, X. Zhu and P. Zhang, "An Efficient Model for Smartphone Forensics Using SMS Spam Filtering," 2020 3rd International Conference on Hot Information-Centric Networking (HotICN), Hefei, China, 2020, pp. 166-169, doi: 10.1109/HotICN50779.2020.9350843.

[16] N. Al-Ghamdi and T. Alsubait, "Digital Forensics and Machine Learning to Fraudulent Email Prediction," 2022 Fifth National Conference of Saudi Computers Colleges (NCCC), Makkah, Saudi Arabia, 2022, pp. 99-106, doi: 10.1109/NCCC57165.2022.10067685.

Authors

Vaibhav Pokhriyal, a fourth-year B-Tech CSE student at PES College set to graduate in 2024, boasts a rich educational journey. Starting at Army Public School, he continued his academic pursuit at Dayananda Sagar College for his diploma. Vaibhav has exhibited versatility through diverse projects spanning IoT, machine learning, digital forensics, and cybersecurity domains. His passion for innovation and technology is evident in these projects, showcasing a keen interest in cutting-edge fields. With a robust foundation from both

schooling and diploma, Vaibhav is poised to make significant contributions to the tech industry, driven by his curiosity and expertise in emerging technologies.

Vaishnavi K, a student hailing from Karnataka, distinguished herself by excelling in her (SSLC) examinations under the State Board. Her academic journey continued with a focus on Computer Science, as she embarked on a diploma program at MEI Polytechnic. Demonstrating a strong commitment to her field, Vaishnavi is currently furthering her education by pursuing a Bachelor of Technology (B.Tech) in Computer Science at the prestigious PES University.Her areas of interest include Machine Learning, Artificial Intelligence and Networking and Systems Administration

Aditi Mallya is currently a third-year BTech CSE student at PES College graduating in 2025, she completed her schooling in NPS. Passionate about technology, her interest lies in the Cybersecurity space and Machine Learning domain.She has done courses in neural networks, cryptography and image processing, also being a part of a project focused on enhancing night vision using advanced image processing techniques; and projects related to web development.She is interested in exploring many more domains related to forensics and blockchain and is keen on making impactful contributions in this field.

Dr. Sapna V M graduated with B.E. in Electronics Instrumentation from BVB, Hubli under VTU University, Belgaum, in 2004 and M.Tech in Computer Science and Engineering from BVB, Hubli under VTU University, Belgaum in 2006. She Completed her PhD in CSE with interdisciplinary research in the Department of BT from PES University in 2021. She has 15 plus years of experience in teaching and research. She Has published 10 research papers in National and International journals. She is a life time member of CSI Bangalore chapter. Her areas of interest includes Digital Forensics, Cybersecurity,Computational Biology, Bioinformatics, Computer Networks, Security, AI, Computer System Architecture and Software Engineering

Pavan A C, a Ph.D. scholar specializing in Computer Science and Engineering at Bangalore University. Currently working as an Assistant Professor at PES University in Bengaluru.My educational qualifications include a Master's degree in Computer Science and Engineering from P.E.S College of Engineering, Mandya, and a Bachelor's degree in Computer Science and Engineering from VidyaVikas Institute of Technology, Mysuru. I am well-prepared to contribute to the fields of wireless sensor networks, digital image processing, and digital forensics. I am driven by a passion for innovative solutions and advancements in these areas.