

Sankita J. Patel
Naveen Kumar Chaudhary
Bhavesh N. Gohil
S. S. Iyengar *Editors*

Information Security, Privacy and Digital Forensics

Select Proceedings of the International
Conference, ICISPD 2022



Lecture Notes in Electrical Engineering

Volume 1075

Series Editors

Leopoldo Angrisani, Department of Electrical and Information Technologies Engineering, University of Napoli Federico II, Napoli, Italy
Marco Arteaga, Departamento de Control y Robótica, Universidad Nacional Autónoma de México, Coyoacán, Mexico
Samarjit Chakraborty, Fakultät für Elektrotechnik und Informationstechnik, TU München, München, Germany
Jiming Chen, Zhejiang University, Hangzhou, Zhejiang, China
Shanben Chen, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
Tan Kay Chen, Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore
Rüdiger Dillmann, University of Karlsruhe (TH) IAIM, Karlsruhe, Baden-Württemberg, Germany
Haibin Duan, Beijing University of Aeronautics and Astronautics, Beijing, China
Gianluigi Ferrari, Dipartimento di Ingegneria dell'Informazione, Sede Scientifica Università degli Studi di Parma, Parma, Italy
Manuel Ferre, Centre for Automation and Robotics CAR (UPM-CSIC), Universidad Politécnica de Madrid, Madrid, Spain
Faryar Jabbari, Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA, USA
Limin Jia, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
Janusz Kacprzyk, Intelligent Systems Laboratory, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
Alaa Khamis, Department of Mechatronics Engineering, German University in Egypt El Tagamoa El Khames, New Cairo City, Egypt
Torsten Kroeger, Intrinsic Innovation, Mountain View, CA, USA
Yong Li, College of Electrical and Information Engineering, Hunan University, Changsha, Hunan, China
Qilian Liang, Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX, USA
Ferran Martín, Departament d'Enginyeria Electrònica, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain
Tan Cher Ming, College of Engineering, Nanyang Technological University, Singapore, Singapore
Wolfgang Minker, Institute of Information Technology, University of Ulm, Ulm, Germany
Pradeep Misra, Department of Electrical Engineering, Wright State University, Dayton, OH, USA
Subhas Mukhopadhyay, School of Engineering, Macquarie University, NSW, Australia
Cun-Zheng Ning, Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA
Toyoaki Nishida, Department of Intelligence Science and Technology, Kyoto University, Kyoto, Japan
Luca Oneto, Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genova, Genova, Genova, Italy
Bijaya Ketan Panigrahi, Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, Delhi, India
Federica Pascucci, Dipartimento di Ingegneria, Università degli Studi Roma Tre, Roma, Italy
Yong Qin, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China
Gan Woon Seng, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore
Joachim Speidel, Institute of Telecommunications, University of Stuttgart, Stuttgart, Germany
Germano Veiga, FEUP Campus, INESC Porto, Porto, Portugal
Haitao Wu, Academy of Opto-electronics, Chinese Academy of Sciences, Haidian District Beijing, China
Walter Zamboni, Department of Computer Engineering, Electrical Engineering and Applied Mathematics, DIEM—Università degli studi di Salerno, Fisciano, Salerno, Italy
Junjie James Zhang, Charlotte, NC, USA
Kay Chen Tan, Department of Computing, Hong Kong Polytechnic University, Kowloon Tong, Hong Kong

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering—quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Editorial Director (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

**** This series is indexed by EI Compendex and Scopus databases. ****

Sankita J. Patel · Naveen Kumar Chaudhary ·
Bhavesh N. Gohil · S. S. Iyengar
Editors

Information Security, Privacy and Digital Forensics

Select Proceedings of the International
Conference, ICISPD 2022

 Springer

Editors

Sankita J. Patel
Department of Computer Science
and Engineering
Sardar Vallabhbhai National Institute
of Technology
Surat, India

Bhavesh N. Gohil
Department of Computer Science
and Engineering
Sardar Vallabhbhai National Institute
of Technology
Surat, India

Naveen Kumar Chaudhary
School of Cyber Security and Digital
Forensics
National Forensic Sciences University
Gujarat, India

S. S. Iyengar
Florida International University
Miami, FL, USA

ISSN 1876-1100

ISSN 1876-1119 (electronic)

Lecture Notes in Electrical Engineering

ISBN 978-981-99-5090-4

ISBN 978-981-99-5091-1 (eBook)

<https://doi.org/10.1007/978-981-99-5091-1>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Paper in this product is recyclable.

Preface

The proceeding presented here comprises the peer-reviewed papers from the International Conference on Information Security, Privacy, and Digital Forensics, 2022. It encompasses an extensive spectrum of themes, incorporating digital forensics, cloud security, privacy, data intelligence, hardware security, network security, blockchain technology, distributed ledger, and others. The volume comprises original contributions and the most recent advancements made by experts in industry and academia, who are active in the domain of security, privacy, and digital forensics, with respect to technological and social aspects. This publication is anticipated to serve as an invaluable reference resource for researchers and academics worldwide.

Surat, India
Gujarat, India
Surat, India
Miami, USA

Sankita J. Patel
Naveen Kumar Chaudhary
Bhavesh N. Gohil
S. S. Iyengar

Contents

Cybersecurity Resiliency for Airports as a Critical Infrastructure	1
Shivendra Anand and Madhavi Dave	
Re-examining Laws Pertaining to Admissibility of Digital Evidence in Investigations	17
Kaushik Thinnaneri Ganesan	
Fog Forensics: A Comprehensive Review of Forensic Models for Fog Computing Environment	31
Konrad Śniatała, Yashas Hariprasad, K. J. Latesh Kumar, Naveen Kumar Chaudhary, and Michał Weissenberg	
Memory Forensics for Artefacts Recovery from Ether Transactions	43
Borase Bhushan Gulabrao, Digvijaysinh Rathod, and Aishwarya Tiwari	
A Lightweight Intrusion Detection and Electricity Theft Detection System for Smart Grid	55
Ayush Sinha, Ashutosh Kaushik, Ranjana Vyas, and O. P. Vyas	
Study and Analysis of Key-Predistribution Schemes Based on Hash Chain for WSN	69
Kanhaiya Kumar Yadav and Priyanka Ahlawat	
CERT-In New Directives for VPN: A Growing Focus on Mass Surveillance and Data Privacy	81
Neeraj Jayant, Naman Nanda, Sushila Madan, and Anamika Gupta	
Addressing DIO Suppression Attack in RPL based IoT Networks	91
Rajat Kumar, Jyoti Grover, Girish Sharma, and Abhishek Verma	
Modelling Identity-Based Authentication and Key Exchange Protocol Using the Tamarin Prover	107
Srijanee Mookherji, Vanga Odelu, Rajendra Prasath, Alavalapati Goutham Reddy, and Basker Palaniswamy	

Sensor Fusion and Pontryagin Duality	123
S. Jayakumar, S. S. Iyengar, and Naveen Kumar Chaudhary	
Lightweight Malicious Packet Classifier for IoT Networks	139
Seyedsina Nabavirazavi, S. S. Iyengar, and Naveen Kumar Chaudhary	
Cyber Security Issues and Challenges on Non-fungible Tokens	151
N. Kala	
The Rise of Public Wi-Fi and Threats	175
Prateek Bheevgade, Chirantan Saha, Rahul Nath, Siddharth Dabhade, Haresh Barot, and S. O. Junare	
Digital Forensic Investigation on Ponzi Schemes	191
Babu Madhavan and N. Kalabaskar	
Holistic Cyber Threat Hunting Using Network Traffic Intrusion Detection Analysis for Ransomware Attacks	199
Kanti Singh Sangher, Arti Noor, and V. K. Sharma	
Cyber Security Attack Detection Framework for DODAG Control Message Flooding in an IoT Network	213
Jerry Miller, Lawrence Egharevba, Yashas Hariprasad, Kumar K. J. Latesh, and Naveen Kumar Chaudhary	
Application of Digital Forensic Evidence in Hit and Run: A Comparative Study with Special Reference to § 304 Part II of IPC ...	231
Hiral Thakar and Manav Kothary	
A Forensic Video Upscaling Colorizing and Denoising Framework for Crime Scene Investigation	251
S. Prema and S. Anita	
A Review of Face Detection Anti Spoofing Techniques on Varied Data Sets	267
Pratiksha K. Patel and Jignesh B. Patel	
Ethereum Blockchain-Based Medicine Supply Chain	279
Jigna J. Hathaliya, Priyanka Sharma, and Sudeep Tanwar	
Machine Learning Algorithms for Attack and Anomaly Detection in IoT	291
Rahul Kushwah and Ritu Garg	
A Mini Review on—Physically Unclonable Functions: The Hardware Security Primitives	305
Harsh Panchal, Naveen Kumar Chaudhary, and Sandeep Munjal	
An Intelligent Analysis of Mobile Evidence Using Sentimental Analysis	317
G. Maria Jones, P. Santhiya, S. Godfrey Winster, and R. Sundar	

Forensics Analysis of TOR Browser 331
Adarsh Kumar, Kumar Sondarva, Bhavesh N. Gohil, Sankita J. Patel,
Ramya Shah, Sarang Rajvansh, and H. P. Sanghvi

**Phishing Classification Based on Text Content of an Email Body
Using Transformers** 343
M. Somesha and Alwyn R. Pais

**Vehicle Theft Detection and Tracking Using Surveillance Video
for the Modern Traffic Security Management System** 359
Charanarur Panem, Ashish Kamboj, Naveen Kumar Chaudhary,
and Lokesh Chouhan

**Resilient Risk-Based Adaptive Authentication and Authorization
(RAD-AA) Framework** 371
Jaimandeep Singh, Chintan Patel, and Naveen Kumar Chaudhary

Survey on Blockchain Scalability Addressing Techniques 387
B. S. Anupama and N. R. Sunitha

Anti-money Laundering Analytics on the Bitcoin Transactions 405
Rajendra Hegadi, Bhavya Tripathi, S. Namratha, Aqtar Parveez,
Animesh Chaturvedi, M. Hariprasad, and P. Priyanga

About the Editors

Sankita J. Patel received her Ph.D. from the Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat in 2015. Since 2009, she has been associated with the Department of Computer Engineering at SVNIT, Surat as an Assistant Professor. Her research interests focus on Applied Research in Information Security and Privacy in various domains like the Internet of Things, Cyber-Physical Systems, Online Social Networks, Data Analytics, Cloud Computing, Big Data, etc. She has worked as a joint investigator in research projects funded by the Ministry of Electronics and Information Technology (MeitY), the Government of India, and the Gujarat Council of Science and Technology, Government of Gujarat. She has co-authored many papers in refereed journals and conference proceedings.

Dr. Naveen Kumar Chaudhary has been Professor of Cyber Security and Dean, at National Forensic Sciences University, Gandhinagar, Gujarat, India since 2019. He is also a courtesy research Professor at Knight Foundation School of Computing and Information Science, Florida International University, Miami, USA. He holds Bachelor of Technology degree in Information Technology and Telecommunication Engineering and Master of Engineering degree in Digital Communication. He earned his Ph.D. in Engineering and advanced certifications in Cyber and Network security. He has an extensive experience of more than 25 years in engineering education, research, and Government. He has authored and co-authored many papers in refereed journals and conference proceedings. He is an editor-in-chief of NFSU Journal of Cyber Security and Digital Forensics and his research focus is in the domain of Cybersecurity, Digital Forensics, Drone-Forensics, Cyber-Physical Systems and Cyber-Forensic Investigations.

Bhavesh N. Gohil received his Ph.D. from the Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat in 2019. Since 2009, he has been associated with the Department of Computer Science and Engineering at SVNIT, Surat as an Assistant Professor. His research interests include performance and security issues in Distributed Computing like Cloud/Edge/Fog Computing, IoTs, etc. Prior to SVNIT, he worked with Amdocs as a Subject Matter Expert. He has been working as a

joint investigator in a research project funded by the Gujarat Council of Science and Technology, Government of Gujarat. He has co-authored many research papers in refereed journals and conference proceedings.

Dr. S. S. Iyengar is currently the Distinguished University Professor, Founding Director of the Discovery Lab and Director of the US Army funded Center of Excellence in Digital Forensics at Florida International University, Miami. He is also the Distinguished Chaired Professor (Hon.) at National Forensics Sciences University, Gandhinagar, India. He has been involved with research and education in high-performance intelligent systems, Data Science and Machine Learning Algorithms, Sensor Fusion, Data Mining, and Intelligent Systems. Since receiving his Ph.D. degree in 1974 from MSU, USA, he has directed over 65 Ph.D. students, 100 Master's students, and many undergraduate students who are now faculty at Major Universities worldwide or Scientists or Engineers at National Labs/Industries around the world. He has published more than 900 research papers, has authored/co-authored and edited 32 books. His books are published by MIT Press, John Wiley and Sons, CRC Press, Prentice Hall, Springer Verlag, IEEE Computer Society Press, etc. During the last thirty years More recently in Spring 2021, Dr. Iyengar was awarded a \$2.25 M funding for setting up a Digital Forensics Center of Excellence over a period of 5 years (2021–2026).

His path breaking discovery known as Brooks-Iyengar algorithm discovered in 1996 is a milestone in his career. This discovery has led to breakthrough in use of sensors in various applications across the globe. By adopting this algorithm, it was possible to use network of sensors which would give out precise outputs, though few of the sensors receive wrong inputs or faulty sensors. This algorithm is relevant even today and has received prestigious “Test of Time” award for its contribution over the decade by IEEE Congress in the year 2019.

Dr. Iyengar has received IEEE Fellow award, ACM Fellow, AAAS Fellow, Fellow of Artificial Intelligence AAIA, Fellow, National Academy of Inventors (NAI), Fellow of Institution of Engineers (India) among many awards he has received in his career. He has also received IEEE Technical Achievement Award in 1998. Dr. Iyengar is awarded the Lifetime achievement award at by International Society of Agile Manufacturing at IIT (BHU) in 2012. He has received Lifetime Achievement award from IEEE High Performance Computing in 2019. Dr Iyengar was also a Fulbright Distinguished Scholar and has received several honorary Ph.Ds. from around the world.

He has been awarded the Lifetime Achievement Award for his contribution to the field of Digital Forensics on November 8, 2022, during the 7th INTERPOL DIGITAL FORENSICS EXPERT GROUP (DFEG) MEETING at National Forensics Sciences University, Gandhinagar, Gujarat, India.

Cybersecurity Resiliency for Airports as a Critical Infrastructure



Shivendra Anand and Madhavi Dave

Abstract As the virtual world is expanding its span for the online applications, the need of cybersecurity is increasing. The cyberattacks and its mitigation techniques are competing with each other in many areas where the critical infrastructure is applicable. One such domain is aviation industry which needs utmost cybersecurity from various types of threats and attacks. This paper aims to demonstrate types of cyberattacks on IT infrastructure of airports and also describe mitigation techniques for each type of attack. The preventive techniques suggested in this paper are helpful for taking proactive steps so the IT attacks on critical infrastructure can be avoided in advance.

Keywords Cybersecurity · Critical infrastructure · Airport security · Cyberattacks · Mitigation techniques for cyberattacks

1 Introduction

Airports are the large-scale systems which are used by worldwide commuters on daily basis. The airports are itself inclusion of so many subsystems, people, technology, and interest groups. Some of them are closely connected subsystems, which means they are significantly reliant on one another, and a fault in one system can influence others [1]. The overall airport systems are safety-critical and extremely complicated due to inter-dependency. They are safety-critical because a malfunction or accident in these systems might result in the loss of human life as well as significant financial damages [2]. Every year, millions of passengers and cargo shipments are processed by airport authority. Customers' personally identifiable information (e.g., Aadhaar, PAN, Passport) and payment information, as well as staff data and biometrics, are all accessible to the system and its authority. Furthermore, because airports are part

S. Anand · M. Dave (✉)

School of Cyber Security and Digital Forensics, National Forensic Sciences University, Sector-9, Gandhinagar 382007, Gujarat, India

e-mail: madhavi.dave@nfsu.ac.in

of a country's critical infrastructure, security breaches can have far-fetching and impacting effects in addition to financial and reputations harm [3]. To make decisions, the information and data that travels through this system must be accurate, intelligent, rapid, and simple to comprehend. Security is essential for the information system to do the duties those are configured and assigned to the system. Airport systems must be thoroughly tested, with significant resources and time committed. To avoid an accident or incident, all components and subsystems must be coordinated and function together [3]. The attack surface of an airport is additionally increased by the different manned and unmanned assets. Internal and external cyberthreats can affect ticketing and POS terminals, luggage handling, e-boarding stations, parking systems, site management, and other types of workstations that control vital tasks. Considering that the pattern of cyberthreats in terms of their targets and frequency has been proven to be constant across various industries, it is feasible to assume that cyberattacks at airports will rise as well, based on prior developments [4].

2 Indian Airport: Documented Cyberattacks

There is dependency on cybersystems for performing the complicated task of aviation industry with ease. It also invites the risk at many levels for cyberattacks. The following are the documented cyberattacks on Indian airports of last decade [5].

- 2012—Officials from the National Technical Research Organization (NTRO) warned the Airports Authority of India (AAI) about major flaws in its cargo management system at the airports of Chennai, Coimbatore, Kolkata, Amritsar, Lucknow, and Guwahati. The biggest issues were weak passwords and old operating systems.
- 2013—According to CBI, a cyberattack caused technical issues at IGI Airport, resulting in the breakdown of the passenger processing system, which caused 50 aircraft to be delayed and their passengers to be manually checked.
- 2019—On May 15, 2019, a cyberattack caused the LAN at the Kolkata Airport to go down, blanking out airline check-in terminals, flight information display boards, and CCTV surveillance, delaying over 4,000 passengers and causing the CISF to deploy more officers. The problem was fixed, and the system was restarted after more than 9 hours of work by IT professionals.
- 2021—Two explosive devices were dropped from a suspected drone upon the Indian Air Force Station in Jammu. It was a first-of-its-kind attack where terrorists used drones to drop explosives.

3 Cyberattacks for IT Infrastructure of Airports

There could be many cyberthreats and cyberattacks from various sources and malicious intentions for damaging the IT infrastructure of airports. Following are the few with high occurrence according to the literature survey:

- **Insider Threat:** The threat that someone with insider information could compromise national security either knowingly or unknowingly by utilizing their position of power. This threat may involve harm to the nation caused by terrorism, espionage, illegal disclosure, or loss of departmental resources or capabilities [6, 7].
- **Critical Infrastructure Attack:** In this, the attacker creates malware that is then utilized to enter the airport's internal IT system without disrupting operations or triggering monitoring equipment. This type of attack contains a malware payload that includes one or more particular vulnerabilities for the airport ground support lighting system, which is required for safe airline landings and is linked to the internal network. It might employ a maintenance monitoring port or an out of band channel. This can result in the unauthorized external entity commanding those lights without being noticed [6–8].
- **Social Phishing Attack:** Employees with a lack of security knowledge and who fail to follow protocols might put the airport's cybersecurity in danger. Email is still the most common way for threat actors to penetrate a system, giving them complete access to the victims' accounts, identities, and authorization. Despite the fact that businesses implement filtering tools, phishing emails may still get through, tricking the victim into doing a harmful action without even realizing it, which can lead to the loss of personal data, identity theft, and remote execution of the malicious code [8].
- **Ransomware:** Ransomware disrupts an organization's operations by presenting a challenge to management. Ransomware is a type of malicious attack where attackers encrypt an organization's data and demand payment to restore access. Cyber-enabled technology is significantly used in civil aviation to increase the safety and effectiveness of air travel. However, ransomware might threaten the system's security by encrypting it and forcing the user to pay the attackers the ransom [7, 8].

DDOS (Distributed Denial of Service): A DDoS attack is a malicious attempt to obstruct a server, service, or network's regular traffic by overloading the target or its surrounding infrastructure with an excessive amount of Internet requests. Airports are subject to a number of cybersecurity threats, some of which are occasionally avoidable. Through the use of botnets or online networks of connected devices, several machines work together to overload a single website in an effort to crash it. If cybersecurity threats like DDOS are not properly addressed, they may have a severe impact on airports and airlines in many different ways [9].

- **Unpatched Vulnerabilities:** Unpatched vulnerabilities are flaws that let attackers use malicious code to take advantage of a known security problem that hasn't been



Fig. 1 ATP persistence process

fixed. When software providers become aware of these program vulnerabilities, they create “patches”—additions to the codes—to secure these flaws. An attacker may be able to perform actions they are not permitted to conduct or acquire the permissions of another user by taking advantage of a vulnerability, “bug,” or validation error in an operating system or application. Such attacks may facilitate the attacker in breaking into the system and exploiting the devices [9, 10]

- **Advance Persistent Threat:** Advance Persistent Threat is a general term for an intruder who has made a long-term, illegitimate presence on the network to harvest extremely sensitive data. Instead of attempting to enter and exit the targeted network as fast as possible, most APT attacks aim to gain and keep continuing access to it. Hackers frequently choose high-value targets, such as nation-states and major organizations, with the intention of collecting information over an extended period of time because APT attacks can require a lot of time and money [10] (Fig. 1).
- **Shadow IT:** Shadow IT describes the use of IT tools, equipment, programs, services, and software without the clear approval of the IT department. With the use of cloud-based services and apps, it has increased tremendously. Simply put, using shadow IT is to get things done more quickly is one of the main reasons. For instance, a coworker might find a more effective file-sharing program than the one that is officially allowed. Once they start utilizing it, other department members might follow the suit. Challenges which the organization can face due to shadow IT is not low but significant. With many predicting that passenger numbers will increase annually, infrastructural capacity issues are becoming more prevalent, new security risks are emerging, and passenger, consumer, and investor expectations are rising, there is a growing need for change. It’s unlikely that the physical security overlay we have today will completely vanish. It might be argued that this is a necessary fundamental layer for enhancing system resilience and contributing to the mitigation of unknown dangers. A combination of physical and virtual security procedures, some taking place at the airport and others beginning as soon as customers place an order to ship products or book a flight, may, however, become the standard over time [10]. This advancement in security measures must be seen in light of potential future challenges.

4 Mitigation Techniques Against Cyberthreat for IT Infrastructure at Airport

4.1 Dealing with Insider Threat

Organizations can design a proactive, prevention-focused mitigation program to find and identify risks, evaluate the risk, and manage that risk before an incident happens to battle the insider threat. We can minimize the risk of the insider threat by applying multiple solutions into the organization [11].

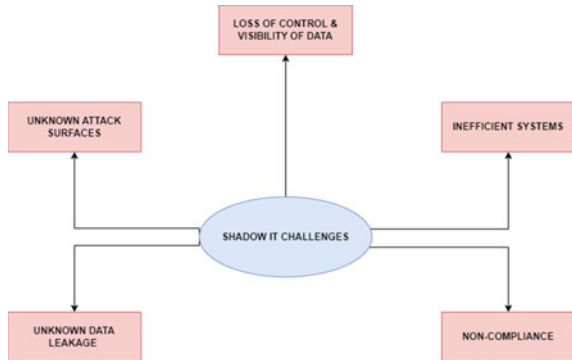
4.1.1 Technical Approaches to Identify Insider Threats

- **IDS (Intrusion Detection System):** The prevention component of a system’s security framework is covered by intrusion detection systems, which act as a second line of defense to strengthen the security measures implemented on the system. They seek to distinguish between events that compromise system security and those that do not by utilizing a variety of detection engines (Fig. 2).

Further processing is necessary to reduce the enormous amount of gathered data that is required for the detecting procedure. IDS’s primary drawback is that it never addresses internal threats, yet it is a huge advantage when it comes to external attacks on the Organization [12].

- **SIEM (Security Information and Event Management):** SIEM is a unique solution that gathers information received from logs and analyzes it on one management platform. It gathers data via secure network channels, including a range of security-related logs, workstation logs, and application logs, among others (e.g., of client workstations, servers, antivirus systems, network devices, honeypots, firewalls, IDS). SIEM correlates all this data after collection. A security administrator can try to detect potential insider activity before it causes system damage

Fig. 2 Threats shadow IT arises



based on this final correlated data. After an occurrence, he might perform forensic research.

DLP (Data Loss Prevention): Data Loss Prevention (DLP) is a technology that detects data exfiltration attempts by insiders early on. Three steps make up the process:

- (a) **System Discovery:** That includes scanning storage devices, observing the user's activity on the endpoint devices, and observing network data flow.
- (b) **Leaked sensitive data identification:** Using methods like keyword matching, regular expressions, or hashing fingerprinting, information found during the system discovery process may be determined to be secret information.
- (c) **Organization policy enforcement:** This phase stops any behavior that would compromise the security of the confidential information that was identified during the system discovery. The benefit of utilizing a data loss prevention technique is that, depending on the business requirements, we may use it to protect three different sorts, or sections, of data within an organization.
- (d) Following categories include data in use, data in motion, and data at rest.

- **Access Control System:** An ACS is essentially a physical operation used in high-security environments like data centers, military and government organizations, and similar institutions. Traditionally, an ACS maintains, watches over, and regulates human access to the facility or protected equipment. Most ACSs are made to accept a user-provided credential as input, authenticate/verify rights using the access control list (ACL), and then allow or deny access based on the results. An insider is a particular kind of user who employs access controls in a system.

Role-Based Access Control (RBAC), Mandatory Access Control (MAC), or Discretionary Access Control (DAC) is implemented for the protection of the data [11]. To prevent personnel from accessing data, they are not allowed to use it; access controls are implemented. This will make it easier to identify the insider immediately whenever any data is leaked.

- **Honey-Tokens:** Although there are various explanations of honey-tokens, they all involve the following three requirements:
 - (a) It must be a physical thing rather than a system, have no commercial value, and be used illegally as a result.
 - (b) Any resource that is kept and isn't regularly accessed for production purposes, such as a text file, an email, or a database record, falls under this definition.
 - (c) To prevent false positive warnings, honey-tokens must be distinct and extremely unlikely to occur in real traffic. They must also be challenging for an enemy to recognize as bait.

4.2 Critical Infrastructure Attack Prevention Measures

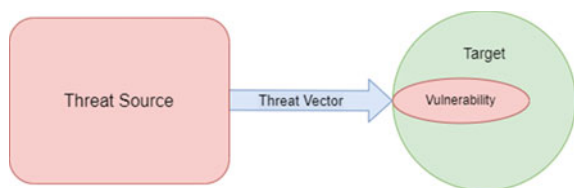
Threat modeling is an important part of dealing with critical infrastructure measures. The system's evaluation is critical for the protection of critical infrastructure. A threat event necessitates the inclusion of the following elements:

1. A threat source to carry out the event.
2. A threat vector to exploit the vulnerability.
3. A target with a vulnerability.

Physical security, network security, computer security, application security device security, policies, procedures, and awareness all perform multiple functions in safeguarding the critical infrastructure [15] (Fig. 3).

- **Physical Security:** Restricted locations, control rooms, high-security environments, electrical and network panels, server rooms, and other sensitive or restricted spaces fall under this category. The physical defensive layer includes suggestions like creating walls that are the right size, using door locks, setting up CCTV cameras, and establishing rules and regulations for dealing with visitors and guests.
- **Network Security:** Network security focuses on limiting access to logical areas of the ICS network, just like physical security does with regard to limiting permitted access to physical areas and assets of the ICS. The concept is to create security zones on the network by using an IDMZ, firewall rules, access control lists, and intrusion detection systems (IDS) to separate more sensitive areas of the network from less secure areas. Anomalies can be found and efficiently controlled by strictly controlling and monitoring traffic passing through the security zones.
- **Computer Security:** Computer security aims to prevent hackers from accessing the computer systems (workstations, servers, laptops, and so on). Patching methods, computer system hardening exercises, and the installation of security tools like antivirus, endpoint protection, and host intrusion detection/prevention (HIDS/HIPS) software are used to achieve this. Access to unused communication ports on computing devices can also be restricted or prevented using computer security controls. For example, physical port blockers can be used to prevent access to USB and FireWire ports, and endpoint protection software like Symantec Endpoint Protection (SCP) can also do this by applying a device policy. Computer security also includes updating and patching software to prevent vulnerabilities in computer systems.

Fig. 3 Threat attack

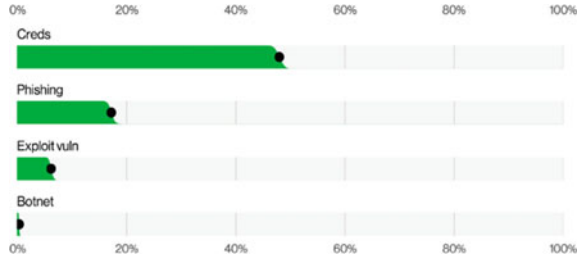


- **Application Security:** Application security focuses on preventing users from engaging in unauthorized interactions with the applications and services that are operating on the computer system, as opposed to computer security, which is all about keeping intruders out of a computer system. Implementing authentication, authorization, and auditing makes this possible. In this case, authentication confirms that the user is who they say they are, authorization limits their actions, and auditing records all of their interactions with the system. Application security also includes preventing vulnerabilities in apps through detection and patching.
- **Device Security:** The availability, integrity, and confidentiality (AIC) triad of ICS devices is the focus of device security, which includes actions and security policies relating to these devices. The term was purposefully reversed to reflect the ICS environment's priorities. In the context of an ICS, availability comes before the others since uptime (availability) is the top aim in production and has the most impact on profitability. For conventional IT systems and networks, the order of the security triangle is CIA or confidentiality, integrity, and availability.
- **Device security** includes patching, hardening, restricting physical and logical access, and establishing a device life cycle program which includes defining processes for purchasing, implementing, maintaining, managing configuration and modification, and disposing of devices.
- **Policies, Procedures, and Awareness:** Policies, procedures, and awareness are the last pieces that hold all the security controls together. Policies provide a high-level framework for ICS systems and devices' expected security posture. For instance, we shall encrypt all our databases. Procedures provide detailed instructions on how to carry out policy objectives, such as putting AES encryption on production recipe databases. Awareness (training) aids in drawing and maintaining attention to security-related elements of the ICS and its functioning. The most common form of awareness training is an annual security course that covers subjects including spam, insider threats, and tailgating customs (an intruder closely following a legitimate employee into a facility that is protected by physical access controls).
- **Phishing Prevention Mechanism:** According to Verizon's Data breach investigation report 2022 phishing is still at 18% of overall graph [16, 17].
- **Blocklist/Allow-list:** The usage of blocklists and allow-lists for URLs, IP addresses, and DNS queries is very crucial as it can mitigate the risk of phishing attack to some extent.

A URL blocklist comprises a list of URLs (domain names or specified pages within a domain) that have been classified as malicious websites. These lists may contain malicious advertising websites, phishing sites, and sites that host or propagate malware. The blocklist is compared to the URL that the end-user has clicked, and if the URL is present on the list, the end-user is discouraged against visiting that website. These lists can be found from a variety of sources, such as those created by PhishTank, PhishFindR, OpenPhish, the Anti-Phishing Working Group, and the National Cybersecurity Centre (Fig. 4).

- **Anti-phishing tools:** An anti-phishing toolbar, which typically appears as a plug-in extension for browsers, is a widely used anti-phishing tool. The toolbar may

Fig. 4 Keypath used for entering the organization (Verizon 2022)



automatically keep track of all the websites that users visited and display a warning notification when a user enters a suspicious URL.

- **Educating Clients and Employees:** Employee education is crucial to prevent phishing attacks. Because of their ignorance of phishing attacks, victims frequently fell victim to them. Users cannot distinguish between authentic and malicious emails or websites. Organizations must support efforts to spread awareness of phishing attacks through education. Demonstrations of common phishing attack methods and sharing of advice with users on how to avoid falling victim to one should be carried out. Employees will be less likely to fall for the traps that phishing attempts establish if they are informed about them.
- **Multifactor Authentication:** Two-factor authentication requires users to be aware of knowledge factors and possession factors to establish their identities, as opposed to just supplying passwords that users should always remember. Username and password are examples of information that a user will be aware of, whereas an access card, key, one-time password (OTP), or authentication code is an item that the user really owns.

4.3 Ransomware Prevention Techniques

- **Endpoint Hardening:** A Windows Firewall policy can be set up to limit the types of communications that are allowed between common endpoints in an environment during a ransomware event. Group Policy can be used to centrally or locally enforce this firewall policy. Between workstations, as well as between workstations and non-Domain Controllers and non-File Servers, at the very least, the common ports and protocols that should be blocked [17].
- **Disabling Hidden shares (Windows):** In order to attach to endpoints throughout an environment, certain ransomware versions will look for administrative or hidden network shares, even those that are not explicitly assigned to a drive letter. An organization might need to take immediate action to prevent endpoint access to hidden or default administrative shares. A service can be stopped, the registry can be changed, or the “Microsoft Security Guide” Group Policy template from the Microsoft Security Compliance Toolkit can be used (Fig. 5).

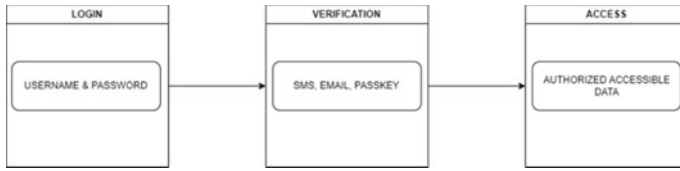
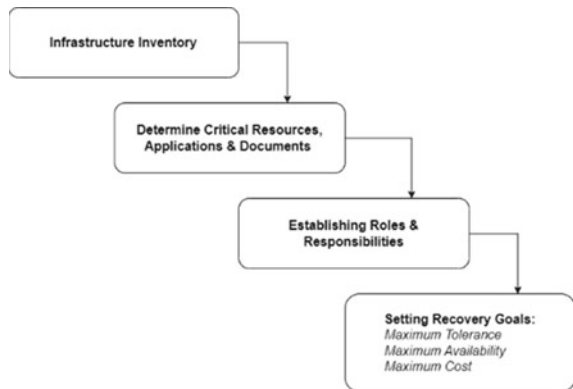


Fig. 5 Multifactor authentication flow

- **Data Backup (On Regular Intervals):** A strong backup and recovery plan is an additional defense against having to pay a ransom. Recovery from a backup might be a good option, depending on how rapidly the breach is discovered, how far it has extended, and the acceptable level of data loss. But this necessitates a more intelligent backup strategy that is in line with the importance of your data and the requirements of your company regarding recovery point objectives (RPO) and recovery time objectives (RTO).
- **Network Segmentation:** The majority of ransomware will attempt to go from the endpoint to the server or storage where all the data and confidential data and programs are kept. The spread can be controlled by segmenting the network and isolating crucial devices and software on a different network or on virtual LAN.
- **Sandboxing:** Before suspicious files may reach the network, they can be moved to quarantine for inspection using technologies like sandboxing. The documents are kept at the gateway until a decision is made. By putting defensive measures in place, such as policies that restrict related IP addresses or domains, or by deploying signatures to security appliances across the network, you can stop subsequent attacks if a file is discovered to be malicious. Multiple sandboxes can be used for such cases like Falcon Sandbox by Crowdstrike, Wildfire-Palo Alto Networks, Fortisandbox Cloud-Fortinet, etc. (Fig. 6).

Fig. 6 Backup recovery plan



4.4 Defending from Distributed Denial of Service

While it is difficult to prevent hackers from trying to launch a DDoS, cautious planning can minimize the probability and potential damage of an attack. When developing a new technology, it's critical to give DDoS mitigation top priority in order to reduce the danger of losing your consumers' trust and to preserve the integrity of the organization [18]. Critical machines can be protected from DDOS attacks using mitigation strategies based on detection, response, and tolerance. Nevertheless, despite new attack signatures and updates, DDoS attacks continue to pose a concern. As a result, there is a vast amount of research in the area of DDoS mitigation, which is the next stage of the defenses.

- Signature-Based Detection Approaches:
 - (a) IDS: It can be used for the signature-based detection of the DDOS attacks. It is a real-time network-based intrusion detection system that operates by watching an intruder's network activity. It has a lot of strong features, including a regular expression-based signature language, a reliable attack detection method, and options for expansion to accommodate additional events. However, it is vulnerable to DoS attacks that make use of advanced algorithms. Additionally, this needs to write scripts to handle the event manually in addition to the attack signature. Therefore, it is crucial to appoint a competent administrator in such circumstances.
 - (b) SNORT: A particularly well-liked network intrusion detection utility is SNORT. It is a simple rule-based tool for detecting a variety of attacks and scans. It has integrated anomaly-based detection with signature-based detection to broaden the scope of attacks it can identify. However, because SNORT relies on precise pattern matching, it may cause a bottleneck in the system's performance due to the high volume of traffic and Internet bandwidth.

The rule can only identify one IP address at a time and is not effective for identifying DDoS attacks. Another SNORT rule was employed to identify DDoS attacks. The rule is similar to the SNORT, which is in charge of directing traffic to its final destination. The source controls the traffic in the prior task (Fig. 7).

- Anomaly-Based Detection: Attacks with new signatures and freshly discovered attacks can both be handled by anomaly-based detection mechanisms. For these strategies, choosing a threshold value to distinguish between attack traffic and regular traffic is still a challenge.

Various techniques are there for the anomaly, some of them are

1. Statistical anomaly detection.
2. Detection based on machine learning.
3. Data mining-based anomaly detection.
4. AI-based detection.



Fig. 7 Defense against DDoS

- **Attack Source Identification:** In order to determine the attack source, we use the IDS’s input and look for malicious network traffic. This facilitates in the network’s genuine traffic filtering process and reduces the amount of traffic that receives false positive results. The act of looking for malicious traffic is crucial since it allows organization to blacklist harmful IP addresses and retrace the route of the traffic’s origin. The organization can prevent attacks in the future from zero days and other malicious traffic by having deep understanding about the traffic that is hostile. The data which can be used for identification against vulnerabilities which can be exploited by the attacker are as listed in Table 1.
- **Filtering, Rate Restrictions, and Responding based on capacity:** DDoS attacks target the victim by targeting primarily on certain vulnerabilities at the application or network level. In other words, the target of these cyberattacks may be a single host, a collection of hosts, or an entire network, and the attack method could be bandwidth exploitation of the target or protocol exploitation.
- **History-based IP filtering:** The edge router will drop the packets as soon as they reach the router by filtering the IPs based on a database that has been pre-fed with a list of malicious IPs.

Table 1 Type of data exploitation by attacker

Type of attack	Information exploitation
Affected IP addresses	IP addresses of the endpoints
Source and destination IP	Used for detecting the threat’s source IP and destination IP
Source and destination ports	Used for detecting the threat’s source and destination port
MAC address	Used for detecting the threat’s source and destination MAC
Protocols	It checks which protocols attacker used for entering the network
Direction	From which direction the threat has entered the network
Severity	The level of the severity of the threat

- **Route-based distributed packet filtering:** The routing information is utilized in this case to determine the source of the traffic. Typically, source IP addresses are restricted, so if someone tries to spoof them, the traffic can be stopped before it reaches the hosts.
- **Load Balancing:** In order to keep the crucial servers from going offline in the event of an attack, this strategy assists the network provider in increasing the bandwidth capacity on business-critical servers. The multiple server architecture can help localizing the data and send traffic to various servers so that the number of requests on each server can be distributed to accommodate the request load.

4.5 Fixing Unpatched Vulnerabilities

There are some vulnerabilities that simply cannot be fixed. For instance, if the device vendor goes out of business, no firmware updates will be provided for the acquired devices. With the current generation of devices, there are compatibility concerns with a few critical applications. For a limited time, we can use several workarounds to resolve the vulnerabilities. When creating patches, the severity of the unpatched vulnerabilities should also be taken into account. Typically, operating systems, networks, and web applications are where vulnerabilities are discovered. Since some operating system and software require outdated systems to function, they don't always receive updates. But for them to function flawlessly, they require that outdated configuration. Using the concepts of network segmentation and isolation, it is possible to isolate out-of-date software and hardware [19].

- The fundamental benefit of network isolation is the difficulty of finding and exploiting isolated devices. In addition, because the devices are in distinct segments, it is more difficult to attack other ones if the attacker discovers the device. Network segmentation, on the other hand, is also a smart choice because it allows for network virtualization and the use of multi-factor authentication as a security barrier between networks. The alternative is to use a new machine that has been completely patched and updated together with comprehensive security implementation in micro-segmentation alongside the vulnerable machine.

4.6 Advanced Persistent Threat Defenses

It is highly difficult to mitigate the ATP (advanced persistent threat), and it is impossible to foresee what vulnerabilities might be combined to launch an ATP attack. Therefore, the mitigation entails providing the organization with all techniques and resources necessary to defend against the attack [20]. Anomaly detection, black/whitelisting, IDS/IPS, SIEM, pattern recognition, risk assessment, and multi-layers of security are a few of the prevalent techniques.

- **Anomaly Detection:** Every network's traffic pattern has a baseline that is taken as the standard condition. It is therefore considered abnormal behavior if the pattern varies for a prolonged period of time. The baseline for expected system and network activity is provided by the anomaly detection system. The essential components that can be utilized to detect anomalies in a network environment are the traffic flow pattern and network performance data.
- **Blacklisting Whitelisting:** Depending on the trusted domains and network traffic of the application, IPs can be blacklisted or whitelisted. The assets of the organization should be accessible to the whitelisted IPs. The conventional way of blacklisting IPs, on the other hand, will assist in preventing harmful apps their activities from entering the organization.
- **IDS/IPS SIEM:** IDS will assist in identifying intrusions by monitoring protocols, IP addresses, and service port numbers. For the purpose of detecting threats, IDS will produce the warnings, when IPS detects the threat, it will take preventive action. If the risk level is high, it will drop the traffic or data packet; if not, it will quarantine the file or data that was acquired. The SIEM is the dashboard from which the SOC team can access all the logs and warnings generated by each endpoint of the organization and use them to take action against the threat that the IDS and IPS systems have identified.
- **Risk Assessment:** Risk assessment is monitoring an application's activity in a constrained environment to determine the dangers and attack potential that it poses. Different types of frameworks are present in the market for the same, such as NIST Cybersecurity Framework, ISO 27000, etc. Asset values are also taken into consideration for the same as this will tell us what the cost of prevention is if anything gets compromised.
- **Multi-Layer Security:** Multiple layers of security are used when communicating with the machines, enabling in safeguarding the organization's assets from ATP and other attacks. Role-based controls, access control lists, encryption, data leakage protection, data redundancy, and log consolidation are the techniques that can be implemented to ensure that the organization's security measures are up to date.

4.7 Managing Shadow IT Risks

The security of the critical infrastructures is being harmed by shadow IT. Employees who use applications that have not been given official approval can cause quite a stir within the company because such applications pose a number of risks to it. Unpatched vulnerabilities in unauthorized applications may enable an organization to be attacked successfully. Data loss is an issue since data is an organization's most valuable asset because it builds public trust in the company. The usage of unapproved software poses compliance concerns as well because non-compliance could result in financial losses for the company [21].

The tactics like deploying shadow IT discovery tools and keeping an eye on the endpoints' network traffic to reduce the risk of shadow IT can be used. Employees should be trained in the professional ethics of utilizing the software that is given by the organization itself, and apps that violate the organization's regulations should be forbidden. Some of the shadow IT discovery tools are present in the organization's cloud services also use them to get benefitted from them.

5 Conclusion

The types of cyberattacks which aim to disturb our critical infrastructure in any way can be identified and prevented in advance. The type of most popular attacks and its characteristics are mentioned in Sects. 2 and 3. Along with that the mitigation steps are depicted in detail, so the proactive step can be taken by the administrative control. This paper includes the categorical solution for each threat which can be implemented as the defense system for cyberattacks at airports as a critical infrastructure.

References

1. Gopalakrishnan K (2021) Cyber security for airports. *Int J Traffic Trans Eng* 3:365–376. [https://doi.org/10.7708/ijtete.2013.3\(4\).02](https://doi.org/10.7708/ijtete.2013.3(4).02)
2. Cyber-Security Challenges in Aviation Industry: A Review of Current and Future Trends, Fortinet (2012). Top 10 network security threats. Fortinet, Inc. Available from Internet: <http://www.fortinet.com/>
3. Lykou G, Anagnostopoulou A, Gritzalis D (2018) Implementing cyber-security measures in airports to improvecyber-resilience. In: 2018 global internet of things summit (GIoTS), pp 1–6. IEEE Publication
4. Lykou G, Anagnostopoulou A, Gritzalis D (2018) Implementing cyber-security measures in airports to improve cyber-resilience, 2018 global internet of things summit (GIoTS), June 2018, IEEE Xplore, ISBN:978-1-5386-6451-3. <https://doi.org/10.1109/GIOTS.2018.8534523>
5. International Air Transport Association (IATA) (2022) Online Magazine: <https://www.iata.org/contentassets/4c51b00fb25e4b60b38376a4935e278b/aviation-cyber-security-position.pdf>
6. Sinha AK, Kulshetra N, Singh BK (2018) Perceived cyber threats to aviation industry in India. *Res Rev Int J Multidisciplinary* 3(12)
7. Suciu G, Scheianu A, Petre I, Chiva L, Bosoc CS (2019) Cybersecurity threats analysis for airports. Springer. ISBN: 978-3-030-16184-2. https://doi.org/10.1007/978-3-16184-2_25
8. Aviation Cyber Security Strategy. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/726561/aviation-cyber-security-strategy.pdf. Accessed 1 Oct 2018
9. Suciu G, Scheianu A, Vulpe A, Petre I, Suciu V (2018) Cyber-attacks—the impact over airports security and prevention modalities. In: World conference on information systems and technologies, pp 154–162. Springer, Cham
10. Duchamp H, Bayram I, Korhani R (2016) Cyber-Security, a new challenge for the aviation and automotive industries. In: Seminar in information systems: applied cybersecurity strategy for managers, pp 1–4
11. Gheyas IA, Abdallah AE (2016) Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis. *Big Data Analyst J* 1, Article number 6

12. Kjaerland M (2006) A taxonomy and comparison of computer security incidents from the commercial and government sectors. *Comput Secur* 25(7):522–538
13. Rajasekaran RK, Frew E (2017) Cyber security challenges for networked aircraft. In: 2017 Integrated communications, navigation and surveillance conference (ICNS), pp 1–15. IEEE
14. Koch T, Moller DP, Deutschmann A, Milbredt O (2017) Model-based airport security analysis in case of blackouts or cyber-attacks. In: 2017 IEEE International conference on electro information technology (EIT), pp 143–148. IEEE
15. National Safe Skies Alliance (2018) <https://www.sskies.org/images/uploads/subpage/PARASO007.CybersecurityQuickGuide.FinalReport.pdf>
16. Rajapaksha A, Jayasuriya N (2020) Smart airport: a review on future of the airport operation, *global journal of management and business research: a administration and management* 20(3) Version 1.0 Year 2020 type: double blind peer reviewed international research journal publisher: Global J Online ISSN: 2249–4588 and Print ISSN: 0975-5853
17. Atapour-Abarghouei A, Bonner S, McGough AS (2019) Volenti non fit injuria: ransomware and its Victims. 2019 IEEE international conference on Big Data (Big Data), pp 4701–4707
18. Zargar ST, Joshi J, Tipper D (2013) A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. *IEEE Commun Surv Tut* 15(4):2046–2069
19. Arbaugh AA, Fithen WL, McHugh J (2000) Windows of vulnerability: a case study analysis. *IEEE Comput* 33(12):52–59
20. Huang L, Zhu Q (2018) Adaptive strategic cyber defense for advanced persistent threats in critical infrastructure networks. *ACM SIG-METRICS Perform Evaluat Rev* 46(2). ISSN: 0163-5999
21. <https://www.redscan.com/news/shadow-it-the-risks-and-how-to-mitigate-the-cyberattacks>

Re-examining Laws Pertaining to Admissibility of Digital Evidence in Investigations



Kaushik Thinnaneri Ganesan

Abstract Multiple laws exist today in India, which deal with the admissibility of electronic evidence. While the Indian Evidence Act, 1872 is the primary law in India dealing with all aspects of evidence, there are a few other specific laws which have also attempted to deal with the issue of electronic evidence. The primary law dealing with electronic evidence, namely, Section 65B of the Indian Evidence Act, 1872 has been the subject of multiple judicial battles and has run its course. There is an urgent need to relook at laws pertaining to electronic evidence and frame uniform laws on this issue which are easy to understand, simple to follow, and, most importantly, relevant to the times. This paper attempts to examine the historic background behind electronic evidence legislations and proposes a few solutions to issues being faced by investigating agencies.

Keywords Digital forensics · Electronic evidence · Section 65B · Indian evidence act · Section 138C · Customs act

1 Introduction

The role of electronic evidence in investigation, whether it be in relation to civil or criminal matters, has dramatically increased in the last few years. While investigating officers have come to realize the importance of electronic evidence increasingly, there appears to be significant gap in following procedures laid out in the law on handling electronic evidence. Part of this could be explained by the fact that the laws are based on enactments many years ago, in a manner which does not appear to hold relevance to the scenario today. Further, certain conditions mentioned in the law appear difficult, if not impossible, to follow, due to a multitude of reasons. In India, legal provisions pertaining to electronic evidence can be found in multiple

K. T. Ganesan (✉)

Indian Revenue Service, Central Board of Indirect Taxes and Customs, Ministry of Finance,
Government of India, New Delhi, India

e-mail: kaushik.tg@gov.in

laws that were enacted at different times. Unfortunately, there appears to be a vast divergence in the legal provisions contained in these various laws. Furthermore, the primary law dealing with evidence, namely, the Indian Evidence Act, 1872, includes a provision pertaining to admissibility of electronic evidence which is based on the United Kingdom's Civil Evidence Act, 1968. Ironically, the latter was repealed even before the same provision became law, vide its enactment in the Indian Evidence Act. Thus, it becomes amply clear that an overhaul of laws pertaining to electronic evidence is the need of the hour, by doing at least the following things—addressing recent developments in technology, clarifying the procedure to be followed, and ensuring uniformity of provisions pertaining to electronic evidence across different laws of the land. This paper attempts to address these issues and proposes solutions for these.

2 Role of Digital Forensics in Investigations

In an era where technology has become ubiquitous for everyday living, even offences are increasingly being committed making use of electronic devices at some time during the commission. It doesn't matter what kind of offence is committed—whether cyber-crime, murder, arson, tax frauds, smuggling, etc., but each of these offences involves usage of some form of electronic device. This is because communication is *sine qua non* in the commission of most offences, which, in today's world, makes it imperative to use digital devices.

Digital forensics helps investigators ascertain one or more of the following roles in the commission of any offence:

- Planning.
- Coordinating.
- Executing.
- Transporting.

The following broad types of information can be gleaned by Digital Forensics (Fig. 1).

Based on the above, the nature of evidence that can be gleaned from digital forensics is as shown below (Fig. 2).

Gone are the days when physical clues such as copies of invoices or other hard evidence are likely to be used by offenders which can be used by investigators as evidence. Increasingly, the role of electronic evidence is appearing crucial to investigations. The establishment of electronic evidence is invariably dependent on effective digital forensics performed on devices seized during the course of investigation.

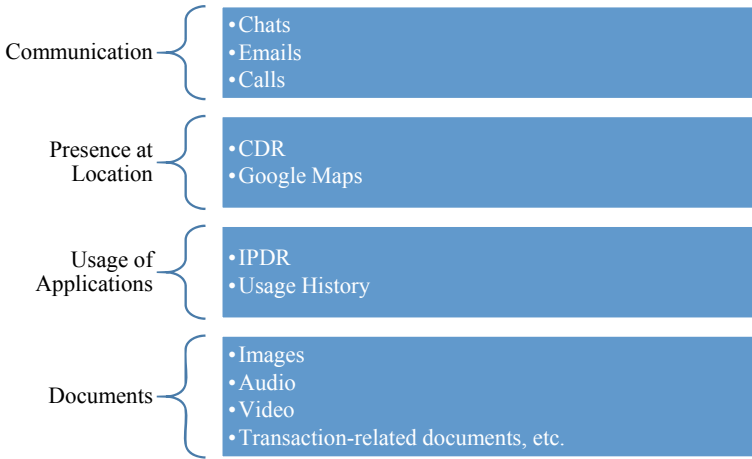


Fig. 1 Broad categories of information from digital forensics

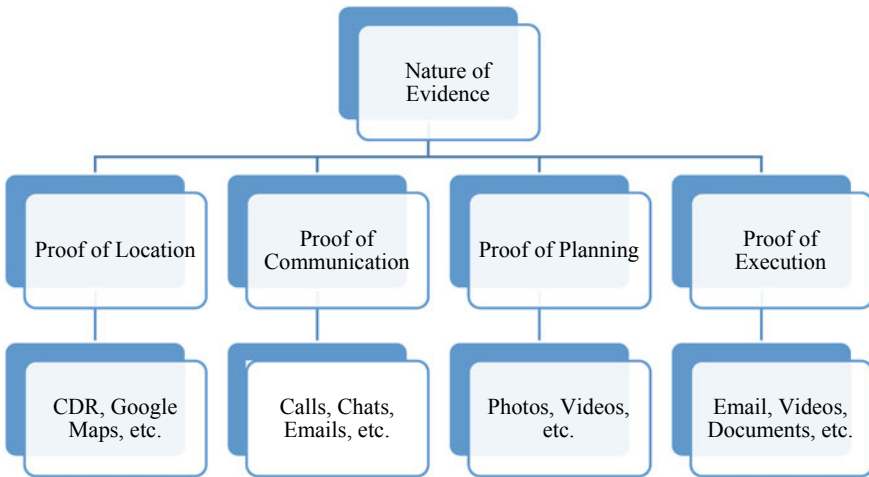


Fig. 2 Nature of evidence

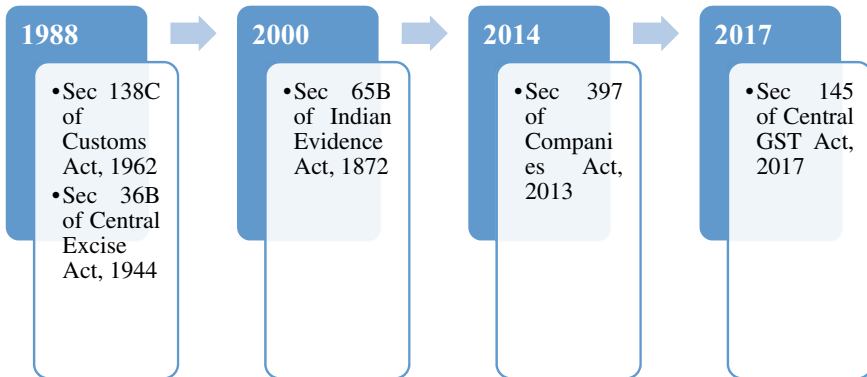
3 Legal Provisions Governing Digital Forensics

It is widely known that in India, Section 65B of the Indian Evidence Act, 1872 deals with the admissibility of electronic records. However, a lesser known fact is that the Indian Evidence Act, 1872 was not the first Indian legislation to address this issue. The Customs Act, 1962 contains a provision pertaining to admissibility of micro-films, facsimile copies of documents and computer print outs as documents and as evidence, which could be read as a precursor to the present-day Section 65B of the

Indian Evidence Act, 1872. A similar provision exists in the Central Excise Act, 1944. Most recently, the Central Goods and Services Act, 2017 has introduced a much-improved provision to address the issue of admissibility of electronic documents as evidence.

3.1 Existing Legal Provisions Pertaining to Electronic Evidence

The Indian Evidence Act, 1872 is the country's primary and general legislation addressing issues governing admissibility of evidence in Indian courts of law. Section 65B of the Indian Evidence Act, 1872 deals with the admissibility of electronic records. In addition to this, there are multiple other laws dealing with the issue of admissibility of electronic evidence. The list of all such laws is given below:



1. Section 138C of the Customs Act, 1962 [1].
2. Section 36B of the Central Excise Act, 1944.
3. Section 65B of the Indian Evidence Act, 1872 [2].
4. Section 397 of the Companies Act, 2013.
5. Section 145 of the Central Goods and Services Tax Act, 2017.

The complete sections of these laws are available in open source and may be referred to for further study.

A graphical timeline of the enactment of various laws pertaining to electronic evidence is given below:

- Both Section 138C of the Customs Act, 1962 and Section 36B of the Central Excise Act, 1944 were inserted vide Notification No. 18/88-C.E. (N.T.) dated 29.06.1988 by s.13 of the Customs and Central Excise Laws {Amendment}Act, 1988 (29 of 1988).

- Section 65B of the Indian Evidence Act, 1872 was inserted by Section 92 of the Information Technology Act, 2000 (Act 21 of 2000) and in the Second Schedule to the said Act.
- Section 397 of the Companies Act, 201 was inserted vide S.O. 902(E) dated March 26, 2014.
- Section 145 of the CGST Act, 2017 was part of the new legislation enacted along with the rollout of the Goods and Services Tax (GST) in 2017.

3.2 Origin of Legal Provisions Used in Section 138C of Customs Act and Section 65B of Indian Evidence Act

Both Section 138C of the Customs Act and Section 65B of the Indian Evidence Act have their origin in Sect. 5 of the Civil Evidence Act, 1968 of the United Kingdom (U.K.). The original act was enacted in the 1960s, which was the era of mainframe computers. Personal computers were rare, there was nothing called the Internet and smart phones, but figments of imagination in creative/scientific minds. The nature of offences that could be committed using such devices was vastly different compared to how electronic devices are used today, in the commission of offences.

An especially intriguing fact is that Sect. 5 of the U.K. Civil Evidence Act, 1968 was repealed by the Civil Evidence Act, 1995, thereby resulting in an ironic situation wherein Section 65B was incorporated in the Indian Evidence Act, by Act 21 of 2000, by copying Subsections (2) to (5) of Sect. 5 of the UK Civil Evidence Act, 1968, when Sect. 5 itself was not there in the U.K.

4 Issues with Section 138C and Section 65B

Of all legal provisions pertaining to electronic evidence, Section 138C of the Customs Act and Section 36B of the Central Excise Act are identical. Section 65B is similar to the above two except for the fact that it includes electronic documents also. The Companies Act is difficult and brief, but relies on the certificate. Finally, the GST Law is concise and simple.

4.1 Increasing Need for Issuance of a Certificate for Electronic Evidence

The key provision governing all laws pertaining to admissibility of electronic evidence is the issuance of a certificate under the relevant section(s) to authenticate/certify the evidence generated. The Hon'ble Supreme Court of India has ruled that this certificate (under Section 65B of the Indian Evidence Act or Section 138C

of the Customs Act) is required only when a secondary copy of evidence is relied upon and submitted as evidence in a court of law.

Going forward, submission of original electronic devices is going to be increasingly challenging because of the volatile nature of electronic data. Oftentimes, trial commences many months or years after initial detection of the offence. Relying on the original device till it is required to be exhibited as evidence poses a problem in two ways:

1. There is no guarantee that the device will continue to function properly after the passage of time. Electronic devices need to be regularly used to function properly. Prolonged non-usage is likely to render the device unusable and may wipe out the data stored in it. Original Electronic Devices are, after the initial seizure and subsequent forensic procedure, sealed back and kept in a secure location till it is required at the time of trial.
2. Some courts have, in recent times, based on applications filed by accused whose devices have been seized by agencies, ordered agencies to return back the original device after cloning the device. With the increased focus of courts on the personal rights of individuals, the retention of original devices with investigating agencies is going to be challenged and may even be discontinued after some time.

Given this scenario, it becomes untenable to rely on the original device as electronic evidence at the time of trial. Therefore, the role of a certificate to establish authenticity (whether under 65B or 138C or any other relevant section) is going to be critical to establish guilt/innocent of an accused.

However, the details pertaining to this certificate are ambiguous in the existing legal provisions, specifically the Indian Evidence Act and the Customs. The same are examined in detail below.

4.2 Issues with Section 138C of the Customs Act, 1962

There appear to be three major sets of issues with Section 138C of the Customs Act, 1962 as elaborated below:

1. Coverage of storage media
 - Section 138C of Customs Act covers only three categories of documents:
 - (1) a micro-film of a document or the reproduction of the image or images embodied in such micro-film (whether enlarged or not) or
 - (2) a facsimile copy of a document or
 - (3) a statement contained in a document and included in a printed material produced by a computer.

This means that information in any other electronic form is admissible, for instance, a pdf file contained in a computer seized or a WhatsApp chat found on a person's phone, without taking a printout of the same.

2. Difficulty in following provisions of Section 138C

Some amount of emphasis of Section 138C appears to have been given to computer printouts generated from a computer, considering an entire sub-section, i.e., 138(2) specifies the conditions in respect of such a computer printout for it to be admissible. The main challenge in Section 138(2) is clause (a)—which specifies when the computer printout should have been taken. According to 138C(2)(a), the computer printout containing the statement should have been

- produced by the computer during the period over which the computer was used regularly to store or process information for the purposes of any activities regularly carried on over that period;
- used by the person having lawful control over the use of the computer;
- used for the purposes of any activities regularly carried on over that period.

Unfortunately, in reality, printouts are seldom taken during the period which the computer was used regularly, as mentioned in the conditions. Oftentimes, printouts are taken in front of the owner of the device, so the document can be used as evidence. However, the following two factors ought to be kept in mind:

- The digital device is no longer used to “regularly store or process information” as specified.
- The digital device is no longer used for the “purposes of any activities regularly carried on over that period” as specified.

There are two major problems with this provision:

- (1) Emphasis on “used regularly”—It is not clear why emphasis has been given to regular usage of a computer to store/process information. There have been instances where a mobile phone or a tablet of an accused does not satisfy the regular usage condition specified, but has been specifically used to commit fraud. Moreover, the vagueness of what constitutes regular usage leaves it open to interpretation. Electronic devices are increasingly being used for short periods to enable commission of offences. To give an example, it has been noticed in multiple cases of gold smuggling that the mastermind or kingpin procures a new mobile device for every new smuggling attempt. Thus, the “regular use” criteria would fail more often than not.
- (2) Apparent rendering of printouts taken during the Forensic Analysis process as inadmissible.

This thus prevents the Department from effectively using even incriminating records contained in digital devices seized during search operations, as evidence in legal fora, since any printout of any information contained in the device does not satisfy at least one of the conditions specified in Sub-section (2).

3. Lack of Clarity about the Certificate to be issued under this section

Sub-section (4) talks about the certificate to be issued under this section. While the details of what the certificate should do have been specified, there appears to exist

some degree of confusion about who should issue the certificate. While the subsection says that the certificate should be “signed by a person occupying a responsible official position in relation to the operation of the relevant device or the management of the relevant activities”, there appears to be some ambiguity about whether the certificate is to be issued by somebody associated with the ownership/maintenance of the device during its normal course of activity (i.e., assessee or individual being investigated) or by an officer of the Department. It would not be practical to expect the owner of the device to provide this certificate. So there exists a certain degree of ambiguity about who would be the right person to give such a certificate.

4.3 Issues with Section 65B of the Indian Evidence Act, 1872

One significant difference between Section 138C of the Customs Act and Section 65B of the Indian Evidence Act is that the latter covers all kinds of electronic records (any information contained in an electronic record which is printed on a paper, stored, recorded, or copied in optical or magnetic media produced by a computer). However, the issues mentioned at points 2 and 3 above persist with Section 65B of the Indian Evidence Act as well.

5 Judicial Pronouncements Regarding Electronic Evidence

The issues with Section 65B have been dealt with in depth in *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal & Ors.*, wherein the Hon’ble Supreme Court of India has rightly held that there is an urgent need to have a relook at Section 65B, which has created huge judicial turmoil. While Section 138C may not have created a comparable degree of judicial turmoil, the complicated nature of the section combined with its outdated provisions has literally rendered the provision unusable, as is evidenced by a plethora of judgements pertaining to the Customs Act. Section 138C of the Customs Act, 1962 and Section 36B of the Central Excise Act, 1944 (which is *pari materia* with Section 138C of Customs Act, 1962) have been a sore point for the Department in various judicial fora and have led to defeats for the Department in multiple cases. A few illustrative cases were considered and studied in detail. The cases and the relevant portions pertaining to Section 138C of Customs Act, 1962 or Section 36B of the Central Excise Act, 1944 (as the case may be) are summarized below:

- (1) CESTAT, New Delhi—*M/s. S.N. Agrotech Vs. C.C., New Delhi* [2018 (361) ELT 761 (Trib. Delhi)]

In the present case, the main evidence on which Revenue has sought to establish the case of under-valuation and mis-declaration of the imported goods is in the form of the computer printouts taken out from the laptops and other electronic devices seized

from the residential premises of Shri Nikhil Asrani, Director in respect of which the requirement of Section 138C(2) has not been satisfied. On this ground, the impugned order suffers from incurable error and hence is liable to be set aside.

- (2) CESTAT, New Delhi—H S Chadha Vs. Commissioner of Customs, Preventive (Customs Appeal No. 51768 of 2016)

Emails and other electronic evidence cannot be relied upon to prove under-valuation in absence of compliance of provisions of Section 138C of the Act *ibid* as held by Anvar P. V and S.N.Agrotech.

- (3) CESTAT, New Delhi—Surya Boards Ltd. Vs. Commissioner of Central Excise, Rohtak [2014 (312) E.L.T. 282 (Tri.-Del.)]

..We find that even the reference in Section 36B is in respect of computer printouts produced by the computers, which was in regular use during the relevant period and such information was being regularly supplied to the computer in the ordinary course of the said activities. Admittedly in the present case, the laptop was purchased by Shri Jitendra Kejriwal only 4 months prior to the date of the seizure and, as such, cannot be said to be a computer which was being regularly used, in the ordinary course of the business, for the period in question.

In addition to the above, given below is a list of cases (not exhaustive) in which the Department's contentions were overruled for failure to comply with provisions of Section 138C of Customs Act, 1962 or Section 36B of the Central Excise Act, 1944:

- a. Tele Brands (India) Pvt. Ltd. Vs. Commissioner of Cus. (Import), Mumbai, [2016 (336) ELT 97 (Tri. Mumbai)].
- b. Commissioner of C.Ex & Customs Vs Ambica Organics, [2016 (333)ELT A-67(Guj)].
- c. Agarvanshi Aluminium Ltd. Vs Commissioner of Customs (I), Nhava Sheva, [2014 (299) E.L.T. 83 (T)].
- d. Harsinghar Gutka Pvt. Ltd. Vs Commissioner of C. Ex. Lucknow, [2008 (334) ELT 77 (Tri. Delhi)].
- e. Premier Instruments & Controls Ltd. vs. Commissioner of C.Ex., Coimbatore, [2005 (183) ELT 65 (Tri. Chennai)].

What appears to be common to most cases referred above is the absence of any certificate under Section 138C of Customs Act, 1962. The absence of any such practice of issuing certificates under Section 138C can be attributed to lack of clarity regarding the person for the issuance of the same. Moreover, considering the way the law has been framed, it appears that in the case of computer printouts, such a certificate can be issued only if such printouts are taken during the regular course of operation of the said computer—not during the course of forensic analysis. To satisfy the statutory requirement of “period over which the computer was used regularly to store or process information for the purposes of any activities regularly carried on over that period”, it seems essential that the print out should be taken during the course of search and also the certificate should be taken simultaneously. Therefore,

despite having legal provisions pertaining to electronic evidence, the Department's hands appear to be tied because of our inability to use the provisions effectively.

6 Other Challenges Faced by Agencies

In addition to the issues posed by the way the provisions pertaining to admissibility of electronic evidence have been framed in the various laws in India, there are certain practical challenges as well that are being faced by officers. These are elaborated in detail below:

a. Not sharing of password

One of the most commonly faced problems by investigators is non-cooperation of accused when it comes to sharing of password/code for unlocking an electronic device. This inability to access the contents of a locked electronic device is a major challenge to investigation since no amount of forensics can lead to recovery of certain data that is protected by the device lock. A legal solution has been provided by the Hon'ble High Court of Karnataka in its judgement in *Virendra Khanna v. The State of Karnataka*. A commonly held misconception is that forcing an individual to part with the password/code for unlocking a device is akin to forcing an individual to give evidence against himself/herself, i.e., self-incrimination. Not providing the password to enable an investigating Department to carry out investigation legally is more akin to withholding the keys to a safe in the custody of the accused rather than refusing self-incrimination. The Hon'ble High Court of Karnataka has ruled that in the event of the accused not providing the password, passcode, or Biometrics, the Investigating Officer can approach the competent Court seeking necessary directions to the accused to provide the same and/or carry out a search of the smartphone or any electronic equipment. Reliance has been placed on Section 100 of the Code Of Criminal Procedure, 1973 for this, i.e., "Persons in charge of closed place to allow search". It appears that not enough investigating officers are making use of this remedy to secure access to electronic devices of accused individuals, which becomes a major factor in conducting an effective investigation.

b. Recovery of Data from Cloud

Another challenge faced is recovery of data from cloud. In situations where the accused refuses to cooperate and data is located in the cloud, the investigating officers have no recourse but to write to the cloud service provider, which is invariably located outside Indian jurisdiction, for the data. Some major service providers provide some meta-data; however, invariably, due to lack of jurisdiction, investigating officers are forced to adopt extremely long-winded and time-consuming procedures such as Mutual Legal Assistance Treaties and/or Letters Rogatory to legally procure the said data. These procedures often take many years and require persistent follow-up with overseas authorities. Oftentimes, even if the data is received, by the time it is received, the domestic time limits have come into play, which means that a chargesheet or show

cause notice is already filed, thus, invalidating the data that has been obtained from abroad.

c. Coverage of deleted data

Today's forensic software oftentimes help in recovering data that has been deleted from electronic devices, specifically from instant communication applications used on smart phones. The law is silent on whether the deleted data that has been recovered is admissible in a court of law, since these are technological developments which could not be conceived even a few years ago.

7 Proposed Solutions

The challenges and issues discussed thus far in this paper are broadly of two kinds:

- a. Issues with existing laws pertaining to admissibility of electronic evidence.
- b. Issues not covered by existing laws that require legal solutions.

Possible solutions for both kinds of issues are explained in detail below:

- a. Issues with existing laws pertaining to admissibility of electronic evidence:

These are the issues covered in the section above pertaining to issues with both Section 138C of the Customs Act, 1962 and Section 65B of the Indian Evidence Act, 1872. The recently enacted Section 145 of the Central Goods and Services Tax Act, 2017 offers a solution worth considering. The section has been deftly drafted keeping it simple to follow and understand while, at the same time, ensuring that the essential principles of admissibility and reliability of the evidence are satisfied. This section also leaves it open as to who may issue the certificate, not leaving any ambiguity about whether the original owner of the device (from whom it was seized) is required to provide the certificate, thus making it more usable for law enforcement agencies. Section 145 of the CGST Act, 2017, thus, appears to be the most suited legal provision pertaining to admissibility of electronic evidence since it has done away with the legacy provisions retained from the erstwhile and highly outdated UK Civil Evidence Act, 1968.

While Section 145 of the CGST Act, 2017 has been elegantly drafted, it is worthwhile to also examine the Federal Rules of Evidence applicable in the federal judiciary in the United States of America. Reference is invited to Rule 902—Evidence That Is Self-Authenticating and specifically, to Sub-rules (11), (12), and (13) of Rule 902.

The framing of the above Rules of Evidence, while enumerating upon “how” electronic records are required to be certified to make them admissible in a court of law, don't specifically touch upon “what” kind of records are covered. This may be used for future amendments to the law, considering the open-ended nature of technological developments.

It is, however, proposed that Section 65B of the Indian Evidence Act, 1872 may also be amended on the lines of Section 145 of the CGST Act, 2017 to ensure that India's primary legislation dealing with the crucial issue of admissibility of electronic evidence is an updated, living, and usable legal provision that helps the noble purpose of law enforcement rather than placing unscientific and outdated conditions which complicate investigations and prevent law enforcement agencies from achieving convictions in courts of law.

b. Issues not covered by existing laws that require legal solutions

This section refers to the three issues mentioned in the section detailing other challenges faced by agencies. Possible solutions for these challenges are proposed point-wise below:

(1) Not sharing of password

The Hon'ble High Court of Karnataka has relied extensively on Section 100 of the Code of Criminal Procedure, 1973 dealing with Persons in charge of closed place to allow search. Be that as it may, it would still be more preferable to include a specific provision to deal with electronic records where passwords or biometric details are stored on the person, whereby the individuals are mandated, by law, to part with such information as may be necessary to gain access to documents that may be relevant to any investigation being carried out by an authorized agency of the Government. This would ensure that there is no ambiguity left in ensuring that individuals are required by law to cooperate with investigating agencies by sharing passwords/codes for electronic devices under their control.

(2) Recovery of Data from Cloud

Notwithstanding the passage of the proposed Data Protection Law which is expected to address the issue of storage of data pertaining to Indian legal entities in servers located outside Indian geography, it does need to be seriously considered whether the Indian Evidence Act needs to address this issue of electronic evidence located in overseas jurisdictions. The justification for this would be that though the service provider is located overseas and the data is stored in servers located overseas, nevertheless, since the data belongs to/has been generated by the client, being an individual accused of committing an offence/crime in India, the data is required to be shared with the investigating agency to aid its investigation.

(3) Coverage of deleted data

"Electronic record" is defined in Sect. 2(t) of the Information Technology Act, 2000 as meaning data, record or data generated, image or sound stored, received or sent in an electronic form or micro-film or computer-generated micro-fiche. A plain reading of this definition seems to imply that any deleted data is not likely to be covered under the ambit of electronic record. We may accordingly consider amending this definition so as to include deleted data that has been recovered by using any software used for performing digital forensic analysis on an electronic device.

8 Conclusion

Information Technology and Electronic Evidence laws are still evolving and leave considerable scope for interpretation in the various judicial fora. There is a long way to go before interpretation of electronic evidence-related aspects is uniform and issues are settled. Therefore, the onus lies with the Executive to draft laws and provisions which are easy to understand and use, primarily by officers of the Department involved in related activities. Further, the laws dealing with electronic evidence need to be reviewed regularly to ensure that they take into account technological developments of the day to ensure that justice is served correctly. At all times, it must be ensured that any law that is being drafted is within the grasp of the officer of the Department who is going to use it on a regular basis. Laws, no matter how eloquently they are drafted, serve no purpose if they contain ambiguous provisions or are beyond the grasp of the field officer who is the actual person implementing such laws.

References

1. The Customs Act, 1962
2. The Indian Evidence Act, 1872
3. The Civil Evidence Act, 1968 of the United Kingdom (U.K.)
4. The Companies Act, 2013
5. The Central Goods and Services Tax Act, 2017
6. The Information Technology Act, 2000
7. Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal & Ors.—Civil Appeal Nos. 20825–20826 of 2017, 2407 of 2018 and 3696 of 2018
8. Virendra Khanna v. The State of Karnataka—WP No.11759/2020
9. Federal Rules of Evidence of the United States of America
10. Dubey V, Admissibility of electronic evidence: an Indian perspective. *Forensic Res Criminol Int J*
11. Goodison SE, Davis RC, Jackson BA, Digital Evidence and the U.S. Criminal Justice System. RAND corporation, in partnership with the police executive research forum (PERF)
12. Romano LV, Electronic evidence and the federal rules. *Loyola Los Angeles Law Rev*

Fog Forensics: A Comprehensive Review of Forensic Models for Fog Computing Environment



Konrad Śniatała, Yashas Hariprasad, K. J. Latesh Kumar,
Naveen Kumar Chaudhary, and Michał Weissenberg

Abstract Numerous potential social advantages are offered by fog computing, including personalized healthcare, smart cities, agri technology, automated transportation, consumer IoT, and many more. Ambient computing at previously unfathomable scales is made possible by the extremely dynamic and complex nature of fog computing and its low latency communication networks connecting sensors, devices, and actuators. The need to look for digital forensic methods that may effectively be used to solve computer-related crimes utilizing IoT devices is being driven by the rise in IoT devices. Fog computing adds greater threats to privacy and security as it is becoming challenging given the increasing number of linked devices. The existing forensics models are not sufficient to handle data from the fog cloud. In this paper, we present a thorough review of the existing state-of-the-art forensic models that can be applied to fog cloud environment and this work can further be used to promote extensive research and development of fog forensic models.

Keywords Fog computing · Digital forensics · IoT · Privacy and security

K. Śniatała (✉) · M. Weissenberg
Poznan University of Technology, Poznan, Poland
e-mail: konrad.sniatala@doctorate.put.poznan.pl

M. Weissenberg
e-mail: michal.weissenberg@put.poznan.pl

Y. Hariprasad · K. J. Latesh Kumar
Florida International University, Miami, FL 33174, USA
e-mail: yhari001@fiu.edu

K. J. Latesh Kumar
e-mail: lkumarkj@fiu.edu

N. K. Chaudhary
National Forensics Sciences University, Gandhinagar, Gujarat, India
e-mail: naveen.chaudhary@nfsu.ac.in

1 Introduction

1.1 Fog Versus Cloud Computing

In recent years, an enormous increase in the amount of devices connected to the Internet has been observed. According to [1], currently, the amount of IoT appliances having access to the web in 2022 reached 14.4 billion and is predicted to increase by 2025 to 27 billion.

So far, the most popular used solution to process data is based on cloud computing. As defined by NIST (National Institute of Standards and Technology), cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [2]. Consequently, cloud computing can be understood as resources available online (usually as a service) by a third party. The most popular example of cloud computing is web file storage and management (e.g., Google Drive, Dropbox, and AWS). In this model, data collected on a device (e.g., smartphone, surveillance camera, temperature sensor, etc.), is sent via the Internet to a remote server located far away. In industrial applications, data in various formats, which are generated by IoT devices and sensors, are then transferred to remote cloud services (e.g., AWS or Microsoft Azure). Next, the information is processed and the results are sent back to the source device. Unfortunately, this solution could encounter many issues and limitations due to the unstable internet connection, security, bandwidth limit, and latency. However, not all data generated by the sensors needs to be sent immediately to the cloud. There are many cases where latency is critical for the system and the response is needed in real time (e.g., autonomous cars, emergency power cut sensors, etc.). In such cases, the processing needs to happen faster, then the time it takes to send data to a cloud server and receive an answer.

In order to overcome these limitations, fog computing has been introduced. The term “fog computing” has been used initially by Cisco, and is sometimes used interchangeably (incorrectly) with “edge computing”. Fog and edge are two different and separate layers. Fog is a layer placed between cloud and end (edge) devices. Fog computing devices receive data directly from the sensors. They process them, do the filtering, and return the result directly to the edge resource. Obviously, data can still be sent to the cloud, but this process does not have to be applied with each request. Information can be aggregated and sent to cloud less often for archiving or further analyses. Fog computing was introduced as an alternative to widely used “cloud computing”, but at the same time being complimentary. The main difference distinguishing these two approaches is data storage and processing location. In fog computing, a remote cloud is usually not used to store large amount of data. Instead, information is being kept in a more decentralized location, which is way closer to the source of the data (device which generates it).

Thanks to the reduced distance, all data transfer processes can be accomplished locally, which is much less complex. Fog computing layer (Fig. 1) can be compared to

Cloud

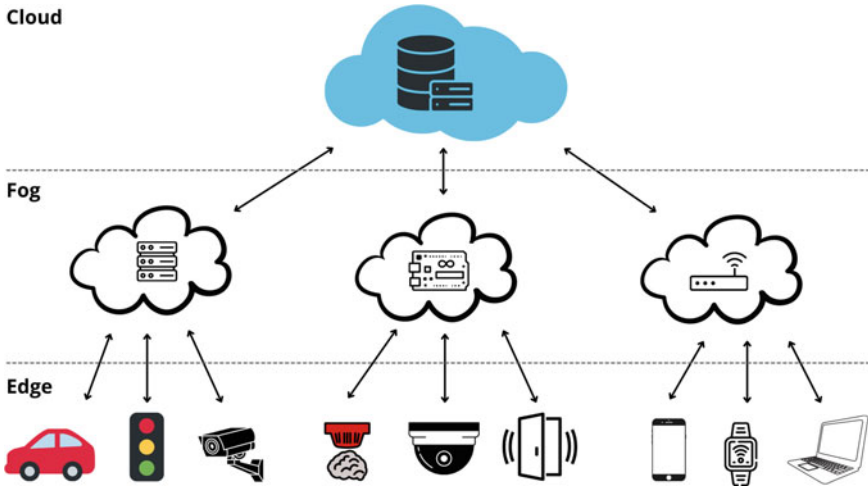


Fig. 1 Cloud, fog, and edge computing layers

a mediator between the edge device and the cloud. Such a solution makes transferring data using fog computing infrastructure quicker and much more efficient compared to the “traditional” cloud computing model.

Fog computing is widely used in a wide range of systems like road traffic control, air quality monitoring, waste management, and many others.

1.2 Digital IoT Forensics

These years due to the emerging amount of small portable devices, being able to connect to the Internet, cybersecurity and digital forensics are disciplines, which evolve extremely fast. Internet of Things (IoT) can be defined as an environment/system of interconnected and interrelated computing devices. IoT devices use technologies such as machine-to-machine communication, context-aware computing, or radio-frequency identification (RFID). Due to exchanging data with devices all around the world, IoT appliances have become a target for hackers, who try to expose the transmitted and stored information. According to [3], in 2021, there have been over 1 billion IoT attacks conducted, from which nearly 900 million were IoT-related phishing attacks. If an attack is successful and data is stolen, tampered, or encrypted, digital forensic specialists are in charge to trace the attacker. Digital forensics can be defined as the process of identification, preservation, analysis, documentation, and presentation (Fig. 2) of the results from digital evidence. This order of digital evidence processing has to be preserved to be officially accepted in a court. Following this scheme also reduces the opportunity for criminals to tamper with the evidence [4]. During many years, special tools have been developed to help and assist forensic

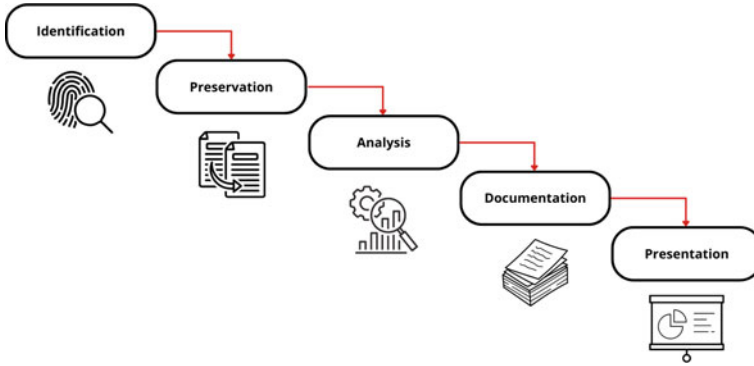


Fig. 2 Digital forensics process

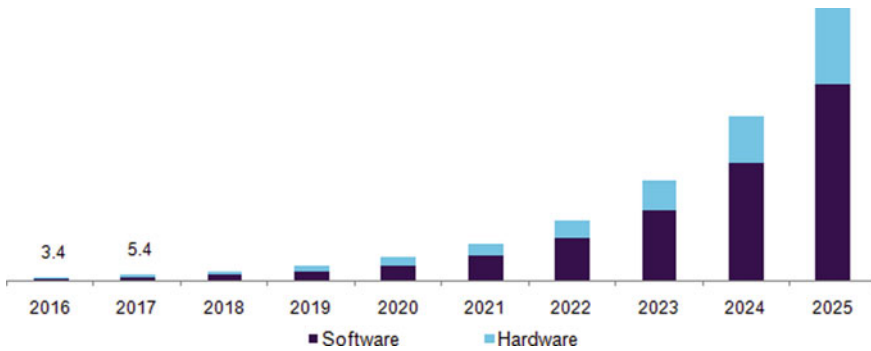


Fig. 3 US fog computing market

investigators with capturing, analyzing, and preserving the most valuable information.

IoT devices are most widely used in the fog computing environment. Due to this fact, the need to look for digital forensic methods that may effectively be used to solve computer-related crimes utilizing IoT devices is being researched.

Researchers are estimating that fog computing market is going to grow exponentially in the next 5 years as shown in Fig. 3. According to the authors, we can see the multi-fold growth only in the United States [5]. Considering the growth globally, we can estimate that the cyber attacks on fog computing would also grow at the same rate and we would need to have robust digital forensics methods and processes to handle the large number of forensics request that may be coming in to the law enforcement agencies.

In this paper, we present a thorough review of the existing state-of-the-art forensic models that can be applied to the fog cloud environment and this work can further be used to promote extensive research and development of fog forensic models.

2 Fog Forensics—Literature Survey

Nowadays, more and more cybercrime investigators, use digital forensics methods in order to solve cases. Big amount of data and the state-of-the-art technology used by criminals make conducting the analysis extremely challenging. The field of fog computing is a layer rich in data and information, which can be crucial to solving cases. Unfortunately, at the same time, diversity of available technology and various platforms, where fog computing can be applied, make it difficult to define universal forensic models for fog computing.

The authors in [6] investigate the methodological, technological, legal, and geopolitical challenges associated with digital forensic investigations in fog computing. The authors present a comprehensive review of the areas that require greater investigations. A framework to stimulate further consideration and discussion regarding the challenges associated with extracting digital evidence from fog computing systems is also presented in the article. This paper gives a clear understanding about the areas that need more concentration and development of intelligent technologies and digital forensics methods required in those areas to tackle the increasing number of cyber attacks.

Authors of [7] have reviewed various fog computing forensic methods. They have put most focus on Advanced Persistent Threat (APT) attacks. APT attacks are usually conducted by whole campaigns (group of hackers), who target military, government, or financial institutions in order to steal high sensitive data. According to [8], 2.8 billion malware attacks were recorded in the first half of 2022 and a sharp 77% rise in IoT malware and a 132% leap in encrypted threats sent via HTTPS. Taking into account the amount of cybersecurity attacks conducted so far in 2022 is highly possible that an organization or individual is going to be sooner or later targeted with a cyberattack and its data might be compromised. In order to gain knowledge and ways of conducting APT attacks aforementioned authors of [7], have proposed a method to implement an enhanced particle swarm optimization PSO algorithm. It is designed for detecting APT attack infections and studying their spread across the fog environment. As the authors describe, their aim is to develop a new approach in order to “Integrate Cyber Threat Intelligence with Digital Forensics analysis in a proactive security approach.” Such a solution will open possibilities to detect the APT attacks in the cloud and fog environment before they reach the destination environment. Afterward, data types affected by the APT attack and their prevalence behavior could be analyzed. Another APT attack was studied in [9]. Researchers analyzed Shamoon malware in fog computing using FPSO (Frequency Particles Swarm Optimization) based on the Travelling Salesman approach (TSP). The proposed system (experiment) consisted of the following steps [9]:

1. Fog nodes initialization (three data types: industrial, medical, and educational).
2. Creation of Shamoon attack—followed by distance matrix evaluation. Attack focuses on industrial data.
3. FPSO parameters initialization, along with particle function evaluation.
4. Finding the shortest path (nearest neighborhood).

5. Detecting local best and global best solutions.

Researchers evaluated the performance of the proposed system and observed attack distribution of Shamoon data. As a result, the authors [9] proposed a threat intelligence scheme for analysis and investigative behavior of Shamoon attacks spread (fog computing edges).

As fog computing is executed mostly on small processing units, many researchers covered in their works and papers IoT Fog Forensics. These days modern cars are equipped with multiple small computers and sensors. Fog computing is useful within vehicular processing. As presented in [10], such solution enhances the communication efficiency and overcomes multiple limitations such as latency or real-time response. The mentioned factors are relevant in terms of autonomous cars, where decisions based on multiple sensor data have to be made immediately. End fog processing nodes can be located at the end of vehicular networks. Thanks to such solution they can acquire, process, and store traffic and in-vehicle parameters in real time.

According to [11], a typical modern car has the power computing of 20 personal computers. The software contains over 100 million lines of code and with the use of over 100 sensors collects over 25GB of data per hour. Significantly, higher values appear when taking into consideration autonomous vehicles. As presented in [12], during autonomous test drives, a car can generate on average up to 20 TB of data a day. When a test drive is performed with a more advanced sensor set, this number can increase even to up to 100 TB/day. Such big amount of data is caused by multiple parameters measured by modern vehicles, e.g., location (GPS coordinates), performance (velocity, RPM), and physical parameters (G-force), usually several times per second [13]. It is worth mentioning that sensors, which provide data to the processing units are not always 100% reliable. If a sensor is faulty, the reliability is compromised. The authors of [14] proposed a hybrid algorithm which is designed to solve problem by making proper decision, even if some of the input data is faulty.

Such a big amount of data established various security and forensic challenges in vehicular fog computing. Unfortunately, the awareness of potential threats and mitigating risks in the field of vehicular fog forensics is at a very low level. According to [10], attacks directed in vehicular fog computing (VFC) systems can be categorized into passive and active. Attacks may be conducted by an external (without the knowledge of key computing components) or internal attacker, who is equipped with information originating from compromised fog nodes or other smart vehicles. Passive attacks aim at compromising private information stored in the car systems, whereas active attacks try to interrupt properly functioning VFC systems. As presented by the authors of [10], a secure VFC system implementation should have the following features: confidentiality, integrity, authentication, access control, non-repudiation, availability, reliability, and forensics. The purpose of forensics is to ensure fog nodes data collection, identification, and analysis in order to trace and compromise the source of attack. Most of the forementioned requirements can be fulfilled by applying various encryption techniques, but this only protects the VFC system from passive attacks—aimed to steal data. Unfortunately, in order to detect fog nodes compromises, more elaborate forensic techniques need to be applied. The

authors of [10] analyzed an attack to compromise fog nodes in a fog-assisted traffic control system. Afterward, in order to increase security, they proposed fog forensic models as countermeasures for attackers. The first and most important step is to identify the compromised fog nodes, without disturbing the performance of the functioning system. A solution, the use of evidence-based digital forensic approach combined with traffic-based analysis approach based on real-time and historical traffic data has been proposed in [10]. Evidence-based digital forensic approach focuses on smart vehicle data and (possibly) compromised fog node artifact analysis. In this approach, the authors prepared a traffic simulation, with smart vehicles having probabilities to properly identify compromised nodes or mistakenly badly mark proper nodes. Unfortunately, due to data noise generated by the smart vehicles, it is hard to detect compromised nodes. The second fog forensic analysis approach uses deep learning algorithms and big data analysis. Information from a compromised node, which usually differs from normal data, could be identified and downloaded from cloud servers containing historical and archive evidence. Relation between the fog nodes can be examined to identify the compromised nodes based on the real-time traffic changes [10]. Described solutions, combined with other approaches [15, 16] for detecting abnormalities in traffic network, show how challenging in terms of forensics this topic is. Authors of [17] have even proposed a dynamic traffic congestion management algorithm. It was based on social interactions between vehicles and commuters—Social Internet of Vehicles (SIoV) concept.

Fog computing processing is usually completed on IoT devices, which play a major role in three main domains: Society, Environment, and Industry. As presented in [18], the field of medicine and health care is a part of the society domain. Thanks to the high response time, low latency, and real-time data processing, fog processing takes healthcare applications to the next level. In the following paper [19], the authors have presented a systematic literature review of fog computing technologies in healthcare IoT systems. Researchers have reviewed a total of nearly 100 articles. Only papers on fog computing in healthcare applications have been included in the review. Results were divided into three major classes; frameworks and models, systems (implemented or architecture), and review and survey.

As presented in [19], patients' vital sign monitoring is one of the most important aspect in healthcare systems [20]. This is the reason many researchers focus on exploring and enhancing data collection solutions. An interesting example of a fog-based monitoring system was presented by the authors of [21]. They proposed a secure "Health Fog" framework of where fog computing was used as a layer between the cloud and the end users. In order to enhance privacy and security, additional cloud access security broker was implemented within the solution. As presented in [22], fog computing has the ability to handle a variety of devices and sensors in addition to provide local processing and storage. According to [19], fog computing is the most suitable technique for healthcare IoT systems, which due to the importance and highly sensitive data, require specific features. Mostly used solutions in healthcare IoT systems (based on cloud computing) cannot withstand excessive demands of healthcare systems like reliability, processing speed, or energy awareness.

In addition, a very important aspect concerning the analysis of systems with distributed data sources connected via fog computing is precision and accuracy, which is very difficult to maintain due to the distributed structure and the emerging noise in the system. Many sensor fusion algorithms can be found in the literature, which can help to determine the precision of the system. The paper [23] presents several approaches to information fusion such as Byzantine agreement, Marzullo's interval-based approach, and the Brooks–Iyengar fusion algorithm. Furthermore, the article [24] presents an information fusion approach for blockchain. Ensuring precision at an appropriate level is also crucial from the point of view of system security and the possibility of detecting potential attacks that affect precisely the precision.

As commonly known devices within fog computing environment might contain important data related to criminal activity, essential for forensic investigations. In the work [25], the author has conducted different experiments with fog networks and evaluated existing digital forensic frameworks for IoT fog devices. Research and testing were done in a specially prepared simulated fog environment. Such an approach gave the possibility to observe the way in which the dynamic service movement can affect the evidence location and nodes data storage possibilities. Conducted experiments were aimed to check the usability of digital forensic methods with IoT fog devices. Author of [25] prepared three scenarios of possible attacks on an IoT fog device:

1. Surveillance camera that captured footage of a criminal activity. This device was placed within a fog network.
2. IoT device located within fog network was infected with malware and further used as a part of a botnet attack.
3. IoT device located within a large-scale fog network, contained sensitive data. A criminal found and stole this data.

In the research paper [25], the author tested multiple frameworks implemented to help investigators with the forensic process. Scenario type and network scale significantly affected the applicability of the tested frameworks. The third case with the most large-scale and complex network, was the scenario where all of the tested frameworks were relevant give. In the first case, which differed from the two others, tested frameworks would not give any significant results. On the other hand, in this example due to the smaller network, it is easier to locate and identify the infected node. As stated by [25], usually specific frameworks focusing on fog IoT are aimed at detecting abnormalities in the network and stopping further incidents, which might not be applicable in all fog-based IoT networks.

The forensics methods that are currently employed by law enforcement agencies are not befitting the collection of evidence about an attack involving IoT and fog systems [26]. To bridge this gap, authors introduced “FoBI: fog-based IoT forensic framework” which is suitable for IoT systems which produce and handle large amount of data and when a large number of devices are deployed. The authors propose to filter the data that requires transmission and to obtain the evidence based on the interaction of the devices. Once the model detects an unusual activity, it alerts the

devices of potential threat. By doing this, an attack on all the other devices can be prevented and will not be propagated to the other connected devices.

The authors present two use cases to demonstrate the proposed model: Smart Refrigerator Use Case and Smart City Network of Sensors.

An interesting and comprehensive work, concerning fog computing privacy and security has been presented in [27]. Authors have marked two main fog computing privacy and security challenges: “Proximity of the fog nodes” and “Stringent requirements in fog computing”. A malicious attack, which later has to be analyzed using forensic methods, is a significant threat in the fog computing environment. According to the authors [27], these attacks can be categorized as follows:

- Attacks against the Network infrastructure (Denial of Service, Man-in-the-middle, and Rogue gateway).
- Attacks against the edge data center (Data leakage, Privilege escalation, Service manipulation, and Rogue data center).
- Attacks against the core infrastructure (Illegal data access, Service manipulation, and Rogue infrastructure).
- Attacks against virtualization infrastructure (Denial of Service, Misuse of resources, Data leakage, Privilege escalation, and VM manipulation).
- Attacks launched by user devices (Data injection and Service manipulation).
- Web-based attacks.
- Malware-based attacks.

In order to protect the devices and defend such attacks, it is required to apply certain countermeasures, e.g., secure the API and apply policy enforcement access mechanisms or intrusion detection systems. The authors [27] have also discussed the cross-border issues and fog forensics. It has been proven that fog forensic challenges are more difficult compared to cloud forensics. As an example, collecting logged data from the numerous fog network devices is way harder than getting data from, a cloud computing server (Fig. 4).

As stated in [27] fog forensics needs international legislation and jurisdictions in order to try to unify the forensic models, which could be used on fog computing devices.

Extracting digital evidence from fog architecture is a time-consuming and very challenging task. The most important difficulties are mentioned in the following points:

1. various manufacturers—fog layer devices are manufactured by different companies. Although there are some common communication protocols, many manufacturers equip their devices with unique systems. In such cases, forensic investigators need to initially analyze the operating systems in order to get access and download data.
2. data formats—some devices save data in specific unpopular formats or structures. This lack of standardization makes the information more secure and safe, but on the other hand, requires more complex and sophisticated forensic expert solutions and approaches to extract it.

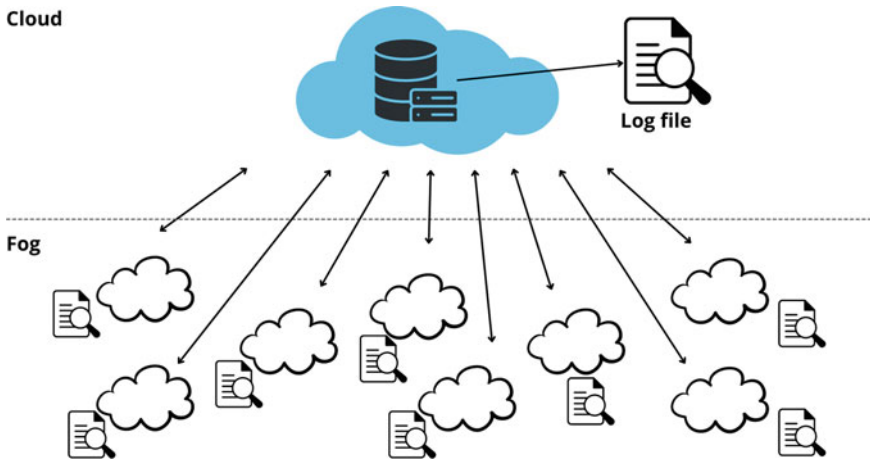


Fig. 4 Multiple fog nodes (devices) from which logs need to be collected, compared to one log file on cloud server. Based on [27]

3. data complexity and volume—currently, many fog devices collect as many data as possible (often send it further for cloud computing). In some cases, such a large amount of data is unstructured. It is a big challenge for the investigators to extract useful information from such big data portions. Usually, such process requires time-stamping and correlating obtained data with other information in order to make certain conclusions.
4. security mechanisms—forensic experts often need to decrypt data stored on fog layer devices, which may require additional computing resources.

Despite the technical aspects, fog layer devices may also collect sensitive data (e.g., medical and personal information), which are protected under different laws. Forensic experts in order to get access to such information need to obtain certain permissions, which may also slow down the data acquisition and investigation process.

3 Conclusions

There is a tremendous increase in the number of IoT devices in the last decade. IoT devices and sensors have become a requisite in most of the sectors including healthcare, transportation, agriculture, and so on. With the increase in such devices, there is also a huge explosion of data that is being generated every day. Computation of data over fog is revolutionary IoT. The enormous increase in data has led to increase in cyberattacks, hence the need of digital forensics for fog computing. Most of the digital forensics processes and principles that are being used by the law enforcement agencies are not suitable when fog computing and IoT is in picture. In this work, we have summarized the existing state-of-the-art forensic methods and principles that

can be applied and is compatible with fog computing and IoT devices. This work can potentially be used by researchers and technology experts around the world to develop new and advanced intelligent forensics methods that can be applied to IoT and fog computing.

Acknowledgements Research was sponsored by the Army Research Office and was accomplished under Grant Number W911NF-21-1-0264. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

References

1. Global IoT market forecast [in bilion connected IoT devices] <https://iot-analytics.com/number-connected-iot-devices/>. Accessed 12 Oct 2022
2. Mell P, Grance T (2011) The NIST definition of cloud computing, p 7
3. 2021 IoT security landscape, <https://securingssam.com/2021-iot-security-landscape/>. Accessed 10 Oct 2022
4. How well do you know digital forensics? <https://www.eccouncil.org/what-is-digital-forensics/>. Accessed 12 Oct 2022
5. Fog computing market size, share and trends analysis report. <https://www.grandviewresearch.com/industry-analysis/fog-computing-market/>. Accessed 13 Oct 2022
6. Hegarty R, Taylor M (2021) Digital evidence in fog computing systems. *Comput Law Secur Rev* 41:105576
7. Hwaitat AKA, Manaseer SS, Al-Sayyed RMH (2019) A survey of digital forensic methods under advanced persistent threat in fog computing environment
8. 2022 sonicwall cyber threat report. <https://www.sonicwall.com/2022-cyber-threat-report/>. Accessed 12 Oct 2022
9. Hwaitat AKA, Manaseer SS, Al-Sayyed RMH (2020) An investigation of digital forensics for shamoon attack behaviour in fog computing and threat intelligence for incident response
10. Huang C, Lu R, Choo K-KR (2017) Vehicular fog computing: architecture, use case, and security and forensic challenges. *IEEE Commun Mag* 55(11):105–111
11. What's driving the connected car. <https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/whats-driving-the-connected-car>. Accessed 12 Oct 2022
12. A driverless future depends on data. <https://blog.equinix.com/blog/2020/06/03/a-driverless-future-depends-on-data/>. Accessed 12 Oct 2022
13. Big data on wheels. <https://www.statista.com/chart/8018/connected-car-data-generation/>. Accessed 12 Oct 2022
14. Brooks R, Iyengar S (1996) Robust distributed computing and sensing algorithm. *Computer* 29(6):53–60
15. Lai C, Lu R, Zheng D, Li H, (Sherman) Shen X (2016) Glarm: group-based lightweight authentication scheme for resource-constrained machine to machine communications. *Comput Netw* 99:66–81. <https://www.sciencedirect.com/science/article/pii/S1389128616300238>
16. Hu H, Lu R, Huang C, Zhang Z (2016) Tripsense: a trust-based vehicular platoon crowdsensing scheme with privacy preservation in vanets. *Sensors* 16(6). <https://www.mdpi.com/1424-8220/16/6/803>
17. Roopa M, Ayesha Siddiq S, Buyya R, Venugopal K, Iyengar S, Patnaik L (2021) DTCMS: dynamic traffic congestion management in social internet of vehicles (SIoV). *Internet of Things* 16:100311. <https://www.sciencedirect.com/science/article/pii/S2542660520301426>

18. Pattar S, Buyya R, Venugopal KR, Iyengar SS, Patnaik LM (2018) Searching for the IoT resources: fundamentals, requirements, comprehensive review, and future directions. *IEEE Commun Surv Tutor* 20(3):2101–2132
19. Mutlag AA, Abd Ghani MK, Arunkumar N, Mohammed MA, Mohd O (2019) Enabling technologies for fog computing in healthcare iot systems. *Future Gener Comput Syst* 90:62–78. <https://www.sciencedirect.com/science/article/pii/S0167739X18314006>
20. Parimbelli E, Wilk S, Cornet R, Sniatała P, Sniatała K, Glaser S, Fraterman I, Boekhout A, Ottaviano M, Peleg M (2021) A review of AI and data science support for cancer management. *Artif Intell Med* 117:102111. <https://www.sciencedirect.com/science/article/pii/S0933365721001044>
21. Ahmad M, Amin M, Hussain S, Kang B, Cheong T, Lee S (2016) Health fog: a novel framework for health and wellness applications. *J Supercomput* 72:10
22. Moosavi SR, Gia TN, Nigussie E, Rahmani AM, Virtanen S, Tenhunen H, Isoaho J (2016) End-to-end security scheme for mobility enabled healthcare internet of things. *Future Gener Comput Syst* 64:108–124. <https://www.sciencedirect.com/science/article/pii/S0167739X16300334>
23. Ao B, Wang Y, Yu L, Brooks RR, Iyengar SS (2016) On precision bound of distributed fault-tolerant sensor fusion algorithms. *ACM Comput Surv* 49(1):5:1–5:23. <https://doi.org/10.1145/2898984>
24. Iyengar SS, Ramani SK, Ao B (2019) Fusion of the Brooks-iyengar algorithm and blockchain in decentralization of the data-source. *J Sensor Actuator Netw* 8(1):17. 1 Publisher: Multidisciplinary Digital Publishing Institute. <https://www.mdpi.com/2224-2708/8/1/17>
25. Gundersen JH (2022) Digital forensics on fog-based IoT devices. Master's thesis in Information Security
26. Al-Masri E, Bai Y, Li J (2018) A fog-based digital forensics investigation framework for IoT systems. In: *IEEE international conference on smart cloud (SmartCloud)*. IEEE, pp 196–201
27. Mukherjee M, Ferrag MA, Maglaras L, Derhab A, Aazam M (2019) Security and privacy issues and solutions for fog

Memory Forensics for Artefacts Recovery from Ether Transactions



Borase Bhushan Gulabrao, Digvijaysinh Rathod, and Aishwarya Tiwari

Abstract Use of cryptocurrencies in crimes like money laundering, ransomware, narcotics trade and terror funding has been on increase. There is a disturbing trend in the use of the cryptocurrencies even in conventional crimes like cheating, scams and financial frauds. The existing research in digital forensics of cryptocurrencies is dominated by Bitcoin and very less work has been done on digital artefacts identification in Ether transactions. There has been use of Ethereum in many criminal activities hence the knowledge of Ethereum forensics is very important for law enforcement agencies. According to Bankless Times, Ethereum is now used more than Bitcoin for illegal activities. The proportion of illicit transactions on the overall flow of Ethereum has risen to 0.33% versus 0.04% for bitcoin. In market capitalization, Ethereum is the second largest cryptocurrency. This paper is an endeavour to locate the digital artefacts related with Ether transactions in volatile memory. Cryptocurrency wallet “Exodus” has been used as it does not ask for KYC and such wallets are preferred by criminals. In all, 12 important digital artefacts were found in the volatile memory. As use of different cryptocurrencies other than Bitcoin is on rise in criminal activities, such research of digital artefacts identification with a given cryptocurrency will prove helpful for law enforcement agencies.

Keywords Bitcoin · Ethereum · Ether · Public key · Private key · Wallet · Digital artefacts · RAM dump · Block · Exodus

B. B. Gulabrao (✉) · D. Rathod
National Forensic Sciences University, Gandhinagar, India
e-mail: bhushan.phdcf21@nfsu.ac.in

A. Tiwari
SVPNPA, Hyderabad, India

1 Introduction

Digital forensics [1] is a relatively new branch of forensic sciences, though the origin can be traced to as early as in 1984, in the form of Computer Analysis and Response Team of FBI. Subsequently need for international collaborations led to formation of International Organization for computer evidence and Scientific Working Group on Digital Evidence [2, 3]. Exponential growth in Information and Communication Technology has led to existence of digital evidence in almost every crime. Cryptocurrencies have emerged as another form of money in just over a period of last one decade and has become preferred means of value exchange among criminals for committing crimes like ransomware, investment scams, frauds, money laundering, sextortion, narcotics trade and even terror funding. As on 29.09.2022, the market capitalization of all cryptocurrencies is about \$946 billion and top three dominant entities in this market are Bitcoin (Market capitalization—\$374 billion), Ether (Market capitalization—\$165 billion) and USD Tether (Market capitalization—\$67 billion) [4]. Use of cryptocurrencies in crime came to public knowledge through ill famous “Silk Road” case in which FBI arrested the criminals running this website for all sorts of illegal activities in period from 2011 to 2013. Transactions worth \$1.2 billion happened through this website on TOR [5]. As per report of company Chainalysis, the illicit addresses received cryptocurrency worth \$14 billion in 2021. It has become a major challenge for law enforcement agencies to identify the transaction flow and people behind use of cryptocurrencies in crimes [6]. In recent years, research work is being done in the field of cryptocurrency forensics but it is largely concentrated to bitcoin transaction analysis. Few other cryptocurrencies that have been studied are Monero, Litecoin, Verge and Dogecoin. There is very less research done as on date in the field of Ether transaction artefacts identification. This paper is an effort in that direction to fill that knowledge gap. Section 2 of the paper briefly describes about basics of Ether cryptocurrency and role in the cryptocurrency ecosystem. Section 3 covers about the instances of use of Ether in criminal activities and thus accentuating the need of more research in Ether transaction forensics. Section 4 mentions about the artefacts found in transaction analysis of other cryptocurrencies. Section 5 mentions about the forensic process and methodology adopted in the paper to complete the Ether transaction and artefacts identification in volatile memory. Section 6 discusses the result and analysis of the experiment conducted and Sect. 7 concludes the paper while mentioning the future scope of work.

2 Ethereum and Ether

Ethereum [7] is decentralized, open-source blockchain with smart contract functionality. Ether is the cryptocurrency of this platform. It was conceived in 2013 by Vitalik Buterin and it was an effort to improve upon the Bitcoin’s scripting language [8]. This blockchain allows to communicate, to transact, to hold assets and to build application

without the need of any central authority. The biggest improvement of Ethereum over Bitcoin is that it is programable. As on date more than 2970 projects have been built on Ethereum blockchain. It has more than 71 million wallets with Ether balance. More than 50 million smart contracts are functional on Ethereum blockchain and more than \$11 trillion value moved through Ethereum network in 2021. Smart contracts are the programs on Ethereum blockchain and they only execute when triggered by a transaction from user or another contract [9].

Externally Owned Accounts (EOAs) and contracts are the two types of accounts in Ethereum. Ownership of Ether by EOAs is established through digital private keys, Ethereum addresses and digital signatures. The private keys are at the heart of all user interaction with Ethereum. In fact, account addresses are derived directly from private keys: a private key uniquely determines a single Ethereum address, also known as an account.

In Ethereum address for an EOA is generated from the public key portion of a key pair. However, not all Ethereum addresses represent public–private key pairs; they can also represent contracts. There is no encryption as part of the Ethereum protocol—all messages that are sent as part of the operation of the Ethereum network can be read by everyone. As such, private keys are only used to create digital signatures for transaction authentication. Ethereum addresses are unique identifiers that are derived from public keys or contracts using the Keccak-256 one-way hash function. Ethereum addresses are hexadecimal numbers, identifiers derived from the last 20 bytes of the Keccak-256 hash of the public key. Unlike Bitcoin addresses, which are encoded in the user interface of all clients to include a built-in checksum to protect against mistyped addresses, Ethereum addresses are presented as raw hexadecimal without any checksum [10].

3 Use of Ethereum in Criminal Activities

In one of the research studies on Darkweb [11] marketplaces, 98% of the cryptocurrencies mentions are about Bitcoin, Monero and Ethereum [12]. Rug pull is a type of a scam in which developers build what appears to be a legitimate cryptocurrencies project, take investors' money and disappear. These Rug pull scams are prevalent in decentralized finance (DeFi) environments because with right technical know-how, it is easy and cheap to create crypto-tokens on Ethereum blockchain and get them listed on decentralized exchanges. In 2021, one of the major cryptocurrency thefts happened on Poly Network in DeFi environment and \$613 million were stolen. It involved three blockchains, viz., Ethereum, BSC and Polygon. Cryptojacking is an offence in which malware is installed on victim's computer and it uses the computing power of that computer to mine cryptocurrencies like Monero, Zcash and Ethereum. Non-fungible tokens are the products in crypto-ecology that are associated with Intellectual Property Rights like images, videos, audios, physical objects, membership, etc. and these NFTs are mainly stored on blockchains like Ethereum and Solana. In 2021, \$44 billion worth of cryptocurrency was sent to ERC-721 and ERC-1155

smart contracts that are mainly associated with NFT marketplaces. There is huge potential of NFT's being used as a means for money laundering.

4 Related Work in Other Cryptocurrencies

Memory forensics has been reported for cryptocurrencies like Monero, Verge and Bitcoin. For Monero cryptocurrency analysis, Ubuntu operating system was used and Monero GUI v0.11.1.0 was used. The memory forensics provided following digital artefacts—passphrase, mnemonic seed, public address of wallet, public address involved in transactions and transaction ID. In the Verge cryptocurrency analysis, Ubuntu operating system was used and the memory forensics of the same provided following digital artefacts—stealth address, passphrase, public address of wallet, public address involved in transactions, transaction amounts and labels [13]. For Bitcoin Cryptocurrency memory forensics, Trezor wallet was used on Microsoft Windows 7 operating system. Trezor One 1.8.3 wallet was used. Memory forensics gave following digital artefacts—transaction history, wallet address, extractable password, device metadata, extendable public keys and public keys. Trezor wallet client produced certain artefacts in memory which persisted for long even after its browser [3] tab was closed. Extended public keys remained in memory without being corrupted after the termination of the process. Extended public key allows for derivation of all past and future wallet addresses. Passphrase artefacts were overwritten immediately after the application was terminated [14]. Memory forensic examination of popular wallet Bitcoin core of Version v0.11.1 has been performed with operating system Microsoft Windows 7. It reveals following digital artefacts—all known private keys in binary format, all known public keys in binary format, known labels, transaction ID and backup wallet file. Passphrase used for wallet encryption was not found in process memory. A backup wallet file was saved in user-specified location. Researchers also located a digital string that preceded the private key. Wallet.dat file is the keystore containing the bitcoin keys and the user data. The debug.log file shows the transactions initiated by the wallet with date and time. It also shows the file location of any backup created by user. Registry keys used by the Bitcoin core client were identified using handles plugin. The presence of these keys indicates presence of an active instance of Bitcoin Core client [15]. In existing literature on Ethereum, most studies cover aspects about smart contracts and security vulnerabilities of smart contracts. Very less emphasis has been given on digital artefacts identification in Ether transaction. For law enforcement agencies as well as digital forensic investigators, it is very essential that they know about different type of digital artefacts associated with given cryptocurrency and about location where these artefacts could be located. This paper serves this important aspect in the digital forensics.

5 Forensic Process

Every forensic analysis comprises collection of data, examination of data, analysis of the data and reporting of the data. All these four steps have been used to conduct the extraction of digital evidence from the Ether transaction. In this paper, focus has been only on volatile memory forensics [16] of the transaction. Once the data collection is complete these values are compared with known values and then results are analysed and presented. The forensic workflow can be represented with the help of the diagram shown below.

As shown in Fig. 1, an experimental setup was established. The technical specifications of the system used are mentioned below.

Experiment setup

Environment—System with Windows OS—Windows 10 Pro Version- 21H1.

Processor-Intel® Core™ i7-2600 CPU @ 3.40 GHz 3.40 GHz.

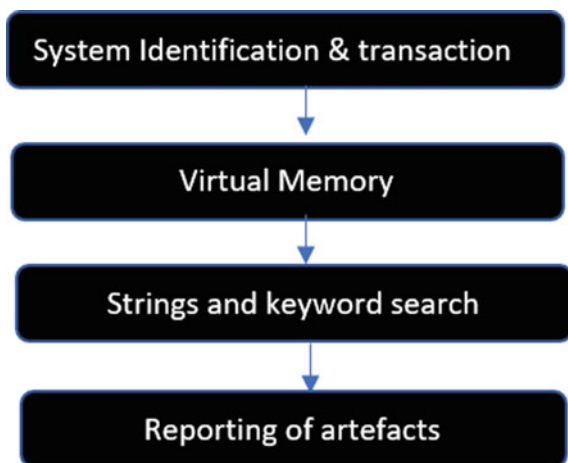
System type—64-bit operating system, × 64-based processor.

RAM—8.00 GB (7.89 GB Usable) has been used for conducting the experiment.

Ether wallet setup—Exodus wallet [17] was selected for this experiment as it has more anonymity features and most important one is that it does not require KYC. This wallet is with an incorporated decentralized cryptocurrency exchange and offers the ability to connect to a centralized one as well. It offers over 235 cryptocurrencies for storage and allows crypto-to-crypto and peer-to-peer swaps.

Exodus version—Exodus version 22.9.8 for Windows 64 bit is being downloaded into system. After completion of download process, the application is installed in the c drive of the system. Hash value of the exodus desktop wallet is as follows:

Fig. 1 Workflow of the experiment conducted



SHA256:

3e7119c59e495431668509a3d05caa54f12c8773bf97b81da012655864f
70b94 exodus-linux-x64-22.9.28.deb
39cc9136716ff544800b4e229a5e7d56127bb596157c88e914fa670789c
c5bd0 exodus-linux-x64-22.9.28.zip
682f91ceb60977d15d3eb6016741b24035231ac5cbb994f35369d544d97
dbdc9 exodus-macos-22.9.28.dmg
cdc350c9359c117bb31cfa3a95c51fb339944b41126fede722ae5df0e3bb
d22a exodus-macos-22.9.28.zip
25971733220ceb5bcd157fcbec2f434c57a9cdeef69eadd780910dceaa6
195a exodus-macos-arm64-22.9.28.dmg
1fa2ea4cce128509ea905f18a88f91d268db38135241aefae595a7d124f76
26f exodus-macos-arm64-22.9.28.zip
2bee8d3963039abd2afcbf38eafce68b46d774466222b244dcc7ab3ad57e
27bc exodus-windows-x64-22.9.28.exe
84db534f25c85445f2560e752afdc04407974486e6ce7d750816f4b4f620
39ec exodus-windows-x64-22.9.28.nupkg

-----BEGIN PGP SIGNATURE-----

wsFcBAEBCAAQBQJjNEkYCRCSbcWfu3WzKAAAJXUQAAsH37/
QqMJwmfCnUyGt4HjsiUkN5HaWhjcbRnfZrNbGEELgOV6FGoqJhz
S7OHeqf90mAfntz2rCRPDfcttZySi6nzDus1Hi1MR+Re5iRNlwZt+GF
jAUK9VMDqoqFGtApK9xhMBytvhAmfxyI6D/qAWRjJWOvrR/kTrs
k+c+cOaBy3jojGqT6LYTY6cPcx4KsiTuPFRDMfOYx0OtJfVGOOn/ZIj
jaHiR3m4x9P4WiviW55GtykFrSlteqT7RTQ7EPz0xiwijJcU/x3Sfoebk
C+GZcNzevuTAYzIYvNdPr832u2qCf4rTq94nTa0kBk2CNLve616Ax
eVKFXVbOEdX1TjKIVNkfTbSyJlJlxzjBz6ZlZkSTic1e5KTb+Bktxbz
YX+vbhuDM4f6OVYBI9D7JmOzpyUWUYXgoyVK+8Wn4jtiVlv0kq0tW
yM0AULnHmMCw+pnKFtl6ufTgAuExYxJXWRjj8Gr9ji16Bm8XzJI
W6N4iuE9EtFjy5W0aLL+TPMjeBsyklqwU66scR8INFixgJliU6kenpiz
/nN+4r/XfVQwokY9PhTu9BUw191rPvULcMZnnWc2707BaiLLD6kG
rvK0SA5uM/ZwvOvqdCjyJXkMiDEJ2plcYT3tJ+qZJ10hb6FKrGTB4
9g+cm7MewZwEzLcwA2kMbxK6Etk1S28rEZ=Sp89

-----END PGP SIGNATURE-----

Address Creation—Two addresses were created from two wallets.

Sender Address—0×8B3ebA6313c98698c892b00aA85ef2531227467c.

Receiver Address—0×90d6D60217fb73928A56c14C8602216887105D52.

Volatile memory image—To make it look like real-life scenario, no VM ware was used and instead the actual RAM dump was taken using FTK imager 3.4.0.1 on the windows system and the analysis of the RAM dump is been done using Bulk Extractor Version 1.5.0 and Win Hex version 15.5.

Four memory images were taken as mentioned below:

1. After turning the system on—image 1.
2. After installation of the wallet—image 2.
3. After conducting the transaction—Image 3.
4. After deleting the wallet—Image 4.

6 Results and Discussion

From the analysis of above-mentioned images, artefacts could be located in various images as mentioned below:

Image 1—A clean machine had no traces of any type of cryptocurrencies.

Image 2—Following artefacts were found:

In domain list, the information about downloading of the wallet.

In URL file, the URL of exodus has been found.

Image 3—By analysing the RAM dump, the following artefacts were found:

Process ID of the exodus wallet application.

Public key address of the sender and the receiver.

Pcap file is also been found which contains the IP address of the server that has been used by exodus.

Name of the wallet used in the transaction.

Balance in the wallet.

Block hash, block number, nonce value, gas price and type of Ether used.

Seed phrase and private key could not be located

Image 4—In domain list, the information of downloading the wallet existed. In URL file, the URL of exodus wallet used existed. Process ID of the exodus wallet application existed.

The results indicate that we can find the various artefacts from the RAM dump of the system in which Ether transaction has taken place. In any transaction, there is involvement of minimum two public key addresses—one of sender and one of receiver. In this experiment, during the analysis of RAM dump both of the addresses could be located as shown in Fig. 2.

In the investigation of any cryptocurrency, it's very crucial to identify the exchange or a wallet which is involved in the transaction. In this experiment, it was possible to retrieve the wallet, i.e. exodus which has been used to perform the transaction as well the process ID used by it is shown in Fig. 3. These details are available even after the wallet used to perform the transaction is deleted from the system.

Network traffic which was generated by the computer helped to identify the websites which have been accessed by the system. The pcap file was analysed using

15. Van Horst D, Luuc, Choo K-KR, Nhlen-An Le-Khac (2017) Process memory investigation of the bitcoin clients electrum and bitcoin core. *IEEE Access* 5:22385–22398
16. Walters A, Petroni NL (2007) Volatools: Integrating volatile memory into the digital investigation process. *Black Hat DC 2007*:1–18
17. Bamakan, Hosseini SM et al. (2021) Blockchain technology forecasting by patent analytics and text mining. *Blockchain: Res Appl* 2(2):100019

A Lightweight Intrusion Detection and Electricity Theft Detection System for Smart Grid



Ayush Sinha , Ashutosh Kaushik, Ranjana Vyas, and O. P. Vyas 

Abstract Smart grid systems have improved networking for power systems and many other industrial systems, but they still have many vulnerabilities, making them an easy target for cyber attacks. Recently, the number of attacks has also increased. The present work investigates the reliability and security of Smart Grid (SG). The reliability and security are investigated in two aspects that are electricity fraud detection followed by the intrusion detection system. This work presents the lightweight Intrusion detection system for SCADA and Modbus-based control systems that can detect intrusion with very high accuracy. The IDS developed is based on the ICS (industrial control system) dataset, which has 20 features (column) and 2,74,628 rows. The IDS dataset contains the Modbus packet's attributes and network and physical infrastructure attributes. The IDS work is followed by detecting electricity theft on a realistic electricity consumption dataset released by the State Grid Corporation of China. A total of 42,372 users' power usage data from 1,035 days is included in the data collection (from 1 January 2014 to 31 October 2016). Eight classifiers, as well as two basic neural networks (1DCNN and ANN), have been investigated on this dataset.

Keywords Intrusion detection · SCADA · Modbus · Electricity theft detection

Supported by C3iHub-IIT Kanpur, India.

A. Sinha (✉) · A. Kaushik · R. Vyas · O. P. Vyas
Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India
e-mail: pro.ayush@iiita.ac.in

A. Kaushik
e-mail: icm2017002@iiita.ac.in

R. Vyas
e-mail: rvyas@iiita.ac.in

O. P. Vyas
e-mail: opvyas@iiita.ac.in

1 Introduction

Two-way digital communication is the foundation of a smart grid, which uses digital technologies to provide power to users. Smart meters were used as part of the smart grid to help it overcome the shortcomings of traditional electrical networks. To combat climate change, improve disaster preparedness, and achieve energy independence, several governments throughout the globe are promoting the implementation of smart grids. So, two-way communication is being used to govern the usage of appliances in smart grid technology. However, the widespread availability of Internet connectivity has made the smart grid more viable to deploy. Users, operators, and automated systems can swiftly adapt to changes in smart grid conditions thanks to the efficient transmission of information through a wide range of smart grid devices.

SCADA systems are used in industrial smart grid infrastructure. Supervisory control and data acquisition (SCADA) is a group of software tools used to monitor, control, and collect data from industrial processes in real time from various distant locations. Data-driven choices about an organization's industrial operations are made possible by SCADA. Hardware and software components are both included in SCADA systems. Data is collected and transferred to field controller systems, which send it to other systems for processing and presenting to an HMI in real time. SCADA systems also keep track of and report on all process occurrences. Alarms are sounded in SCADA applications when dangerous situations arise. Mostly, SCADA uses the Modbus protocol for communication and managing SG. A serial communication protocol designed by Modicon for use with their programmable logic controllers, Modbus was released by Modicon in 1979. It is a way of sending data between serial-connected electrical equipment. It is termed a Modbus Master and a Modbus Slave when a device requests information from another. Every Slave in the 247-slave Modbus network has its Slave Address ranging from 1 to 247. It is also possible for the Master to transmit data to the Slaves. Intrusion detection is being done on the dataset, which consists of packets of Modbus.

2 Literature Review

Authors in their proposed approach in [4] used the temporal behavior of frequently occurring patterns in the SCADA protocols to identify assaults on SCADA systems using an Intrusion Detection System (IDS) specialized to SCADA. When it detects aberrant activity, the IDS sounds an alert. The IDS detected a significant number of assaults, but false alarms were kept to an absolute minimum. An operating system (OS) diversity-based intrusion detection system for SCADA systems is presented in this [5] research as a new and reliable intrusion detection method. SCADA communication over time is analyzed at the OS level, and the most suited OS is selected for intrusion detection based on reliability. According to experiments, OS diversity gives a wider range of intrusion detection options, increasing detection accuracy by

up to eight additional attack types. As a result of their idea, the system's accuracy can be improved by up to 8% on average when compared to a single OS method in the best situation. Anomaly detection systems (AbIDS) may be used to identify a stealthy cyber assault on the SCADA control system, which is being researched in this [6] work. Intending to choose a more effective IDS for SCADA security, we used the IDS tools Snort and Bro throughout the design phase and evaluated their detection rates and delay in alert packets. The timing-based rule is used to detect malicious packets based on the high temporal frequency of malicious packets in network traffic. They used the SCADA-based protection mechanism to shield the system from disruptions during the case study. The SCADA controller was hacked first, and then the data integrity of the system generator was compromised. Impact analysis and performance assessment of IDS tools are then carried out. A variety of network packet sizes were tested to see how quickly IDS solutions could detect cyber-attacks, and the findings showed that they were. Data from a gas pipeline system given by Mississippi State University is used in this [7] research to evaluate the effectiveness of Machine Learning (ML) in detecting intrusions in SCADA systems (MSU). This work makes two contributions: Two methods of data normalization were evaluated, one for accuracy and precision, and the other for recall and F1-score for intrusion detection, for a total of four methods of missing data estimates and normalization. There are two types of classifications distinguished here: binary and categorical. This research shows that RF has a high F1-score of 99% for detecting intrusions. Four distinct CPS datasets, this [8] research compares the performance of several machine learning techniques. To begin, the accuracy, precision, recall, F1-score, and AUC of machine learning algorithms are all measured and evaluated. It is also important to keep track of the amount of computing needed for training, prediction, and deployment. For critical infrastructure with diverse computing and communication limits, our extensive experimental findings will assist in choosing the appropriate machine model. According to the results of the experiments, a linear model is quicker and more suited for CPS bulk prediction. The decision tree is a suitable model for detection performance and model size.

This [9] research employs a SCADA dataset including DoS assaults and running the IEC 60870-5-104 protocol. The protocol will be wrapped into TCP/IP before being transferred so that the treatment in detecting DoS attacks in SCADA networks utilizing the IEC 104 protocol is not significantly different from a regular computer network. Intrusion detection systems (IDSs) are used to identify DoS attacks on the SCADA network using three machine learning approaches: Decision Tree, Support Vector Machine, and Gaussian Nave Bayes. 99.99 percent of the time, tests on the testing and training datasets reveal that the decision tree technique has the best performance detection. A deep learning-based intrusion detection system for SCADA networks is proposed in this [10] study to defend ICSs against conventional and SCADA-specialized network-based assaults. To define significant temporal patterns of SCADA data and identify periods when network assaults are occurring, we suggest using a convolutional neural network (CNN) rather than hand-crafted characteristics for individual network packets or flows. In addition, we devise a re-training method that allows SCADA system operators to augment our neural network models using

site-specific network attack traces. A deep learning-based solution to network intrusion detection in SCADA systems was shown to be effective in our tests utilizing actual SCADA traffic datasets, with high detection accuracy and the capacity to manage newly discovered threats. Using the autoencoder deep learning model (AE-IDS), we create an IDS for the SCADA system in this [11] study. The most often used SCADA communication protocol in the power substation is DNP3, which is the objective of the detection model. SCADA systems are particularly vulnerable to data injection and modification assaults, which fall under the broad category of “cyberattacks”. This research presents the training of an autoencoder network using 17 data characteristics collected from DNP3 transmission. We examine the accuracy and loss of detection of several supervised deep learning algorithms by measuring and comparing the results. Other deep learning IDS models perform better than the unsupervised AE-IDS model.

3 Problem Definition

1. The first primary objective of this work was to build a highly accurate intrusion detection system based on physical and network parameters for MODBUS-based systems while reducing the intrusion detection algorithm’s reliance on domain knowledge, i.e., the algorithm should not be pre-fed background information.
2. Another problem in smart grid infrastructure is electricity theft. The second objective of this study was to design the system to detect the same. Various ML techniques and classifiers are deployed and experimented with improving test accuracy.

4 Dataset and Proposed Methodology

In this section, we will discuss both datasets, dataset features, processing, and other details. We will also see the proposed methodology and what we plan to solve the problem at hand. In the next section, we will see the results of the methodology.

4.1 Intrusion Detection System

Dataset ICS(industrial control system: The system is simply a gas pipeline system that relays information back to SCADA about various system characteristics. Using the MODBUS packet, we can quickly determine the physical parameter’s value (for example, pressure). Now, the SCADA may give control instructions based on these data. SCADA, for example, receives a value of X-Y kPa from the field while the


```

% Mississippi State SCADA Lab
% Gas Pipeline Dataset
%
% Sources:
%   Author: Ian Turnipseed
%   Advisor: Dr.Morris
%   Date: Fri Dec 19 09:53:19 2014
%
@relation gas

@attribute 'address' real
@attribute 'function' real
@attribute 'length' real
@attribute 'setpoint' real
@attribute 'gain' real
@attribute 'reset rate' real
@attribute 'deadband' real
@attribute 'cycle time' real
@attribute 'rate' real
@attribute 'system mode' real
@attribute 'control scheme' real
@attribute 'pump' real
@attribute 'solenoid' real
@attribute 'pressure measurement' real
@attribute 'crc rate' real
@attribute 'command response' {0,1}
@attribute 'time' real
@attribute 'binary result' {'0','1'}
@attribute 'categorized result' {'0','1','2','3','4','5','6','7'}

```

Fig. 1 Attributes in dataset

pipeline pressure should be X kPa. SCADA then sends an order to raise the pressure by Y percent. The following is the data that the Modbus packet conveys (Fig. 1).

Proposed Methodology—IDS The objectives of algorithm selection are as follows:

1. If it can tell the difference between fault and assault, it is doing its job correctly. It should also be able to tell what kind of assault it is.
2. The algorithm must be lightweight: it should not interfere with the core function of the industrial computer.
3. No previous domain knowledge should be required: no networking or gas pipeline data should be provided in this case.

Because of reason number two, we decided against using a neural network. Logistic regression was our first thought when attempting to determine the likelihood of an assault. There was no noticeable difference in accuracy when the value of “C” was altered. Also, the next natural step was to explore SVM and try different improvisations.

4.2 Electricity Theft Detection

Dataset The State Grid Corporation of China made this data public (SGCC). A total of 42,372 power users were tracked for 1,034 days (1 Jan 2014–31 Oct 2016). One individual out of the first 3615 has been tagged as a fraudster. There are 40258 clients listed in the actual data.

Data Preprocessing Missing values are common in electricity usage statistics. The failure of smart meters, the inconsistent transfer of measurement data, the unannounced system maintenance, and storage concerns are all contributing factors. To fill in the blanks, we'll use the interpolation approach using the equation below:

$$f(x_i) = \begin{cases} \frac{x_{i-1} + x_{i+1}}{2} & x_i \in \text{NaN}, x_{i-1}, x_{i+1} \notin \text{NaN} \\ 0 & x_i \in \text{NaN}, x_{i-1} \text{ or } x_{i+1} \in \text{NaN} \\ x_i & x_i \notin \text{NaN}, \end{cases}$$

If x_i is a non-numeric character or a null value in the electrical consumption statistics throughout a period, we display it as NaN. (NaN is a set). Some of the values in the data are incorrect. Here's how we get back the original value:

$$f(x_i) = \begin{cases} \text{avg}(\mathbf{x}) + 2 \cdot \text{std}(\mathbf{x}) & \text{if } x_i > \text{avg}(\mathbf{x}) + 2 \cdot \text{std}(\mathbf{x}), \\ x_i & \text{otherwise,} \end{cases}$$

Avg(x), std(x): The average value of x and the standard deviation of x are shown in this equation. Because each user's power usage always exceeds zero, we only take into account the positive deviation in the preceding calculation. We must normalize the dataset since neural networks are sensitive to a wide range of data. As for scaling, we utilized MAX-MIN scaling using the equation below:

$$f(x_i) = \frac{x_i - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})}$$

Min(x) is the lowest value in x and max(x) is the highest value in x.

4.3 Data Visualization

After creating a new dataset that includes the first three rows of the original dataset and the final two rows of the original dataset (consumers without fraud), we can begin our Visualization.

We must plot customer data based on the following criterion: customers with fraud and customers without fraud. Dated customers are consumed. Consumption

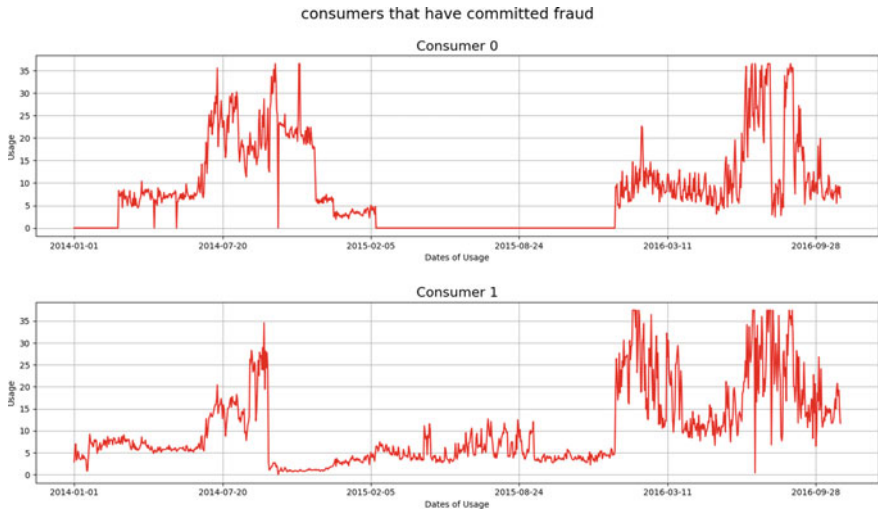


Fig. 2 Consumers who did not do fraud

data in the form of a histogram. The amount of power used in a certain area. Other data include the 50% value, maximum, and lowest value.

The first half of the dataset contains users who have committed the fraud, whereas the second half of the dataset contains the consumers who have committed the fraud. Figure 2 shows the electricity usage of the first two uses of the dataset over the whole time range. Figure 3 plots the usage of the last two users (Users 40255 and 40256) over

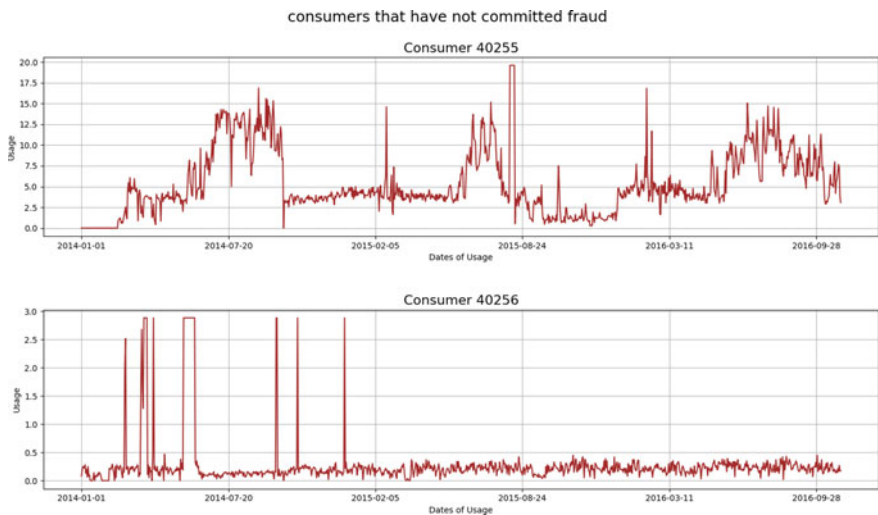


Fig. 3 Consumers who did fraud

Statistics for consumers that have not committed fraud

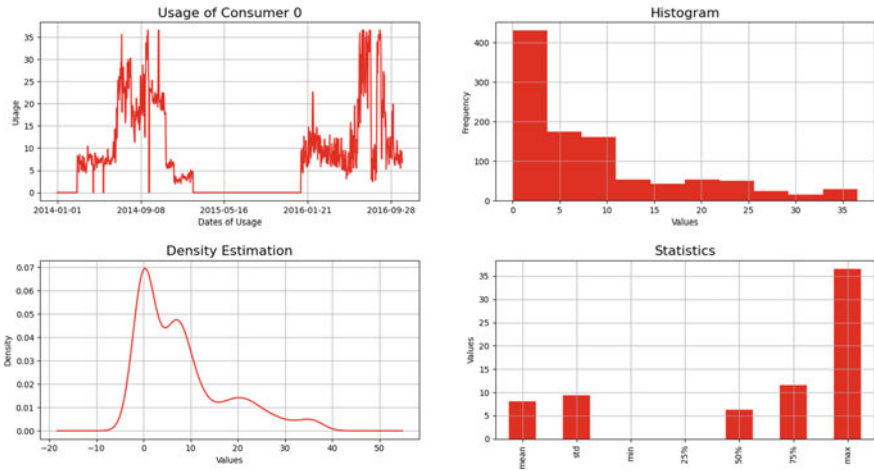


Fig. 4 Consumers who did not do fraud

Statistics for consumers that have committed fraud

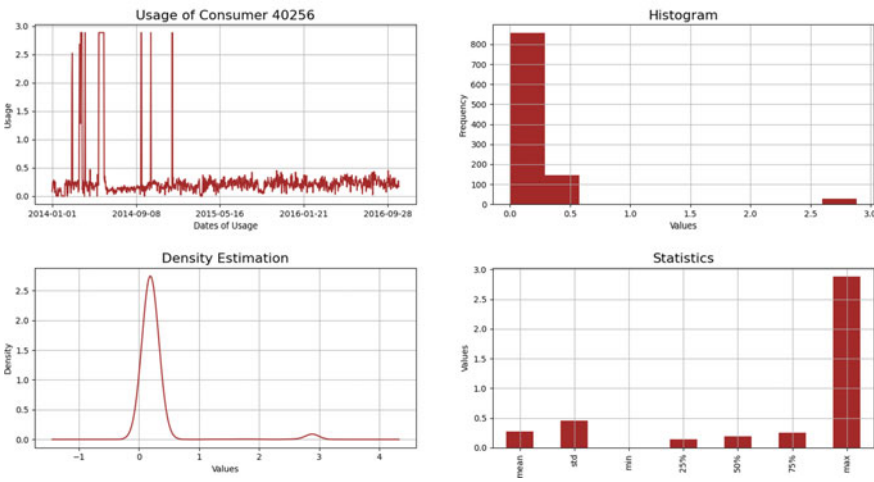


Fig. 5 Consumers who did fraud

the same period. Finally, Figs. 4 and 5 contain the shoes various statistical analyses of first and last users, and the histogram is of frequency of usage vs the amount of usage. Here, first user is a fraud, and the last user is not a fraud (Tables 1 and 2).

Proposed Methodology—ETD After processing the dataset, we now try different categories of algorithms and models and see if we can tune the parameters and get the improvement. In this existing work [14] (refer to Table 3), they already used deep

Table 1 Accuracy of basic models experimented

Model	Accuracy(%)
ANN	88.5678768157959
CNN1D	88.94066480273378
SVM	89.12705809257533
RF	89.12705809257533
DT	81.85771978875427
LR	85.425937556588354

Table 2 Accuracy of all of the classifiers experimented

Classifier	Accuracy(%)
XGB classifier	89.25132028580305
LGBM classifier	89.22025473749612
Gradient boosting classifier	89.22025473749612
CatBoost classifier	90.59304131717925

Table 3 Comparison of our results with the results of paper [14]

Classifier	Accuracy(%)	Accuracy achieved [14]
LR	89.25132028580305	0.8670
SVM	89.22025473749612	0.7536
RF	89.22025473749612	0.8864

neural networks and achieved excellent results. Still, we did try ANN and 1D-CNN with machine learning classifiers and wanted to improve performance in the classifier category, where we got little improvement. In the next section, let us discuss each experiment and model we tried in detail for both problems.

5 Results and Comparison

In this section, we will discuss the experiments we performed for the problem statement that we discussed and their result and compared them with the existing work.

5.1 Intrusion Detection System

Experiment 1: Logistic Regression We wanted to define the boundary of classification to be very precise and check whether having a loose or tight margin would

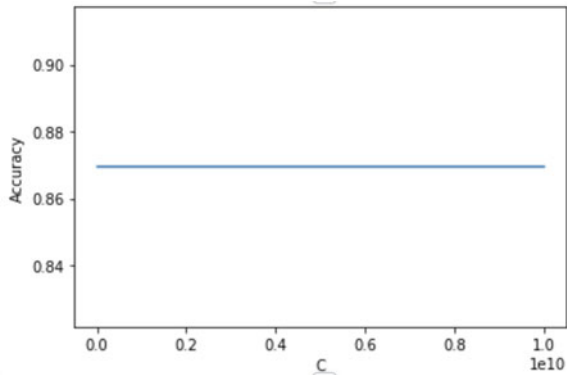


Fig. 6 Exp-1:Accuracy versus Control error V1

help us perform better, so we performed a loop on control error (C) to see whether LR’s accuracy improved was still 86.94 (Refer to Fig. 6).

Experiment 2: Division of Dataset Now, One thing to notice is dataset contains both command requests and command responses. Now allowing the algorithm to differentiate between this part does not fit our third aim, but if it significantly helps the algorithm, then it is just the knowledge of request and response. So we divided the dataset into one which contains “command_request” 1 and other which contains “command_response” 2. This ultimately helped the algorithm to distinguish between request and response. Now also note that the number of responses and requests were equal, so the total number of rows after response removal is $274628/2 = 137314$. This is the time it improved from 86.94 to 90.7, although when we tried to vary the value of control error (C) accuracy didn’t change much. It varies from 90.3 to 90.7 for a large range of C (Refer to Fig. 7).

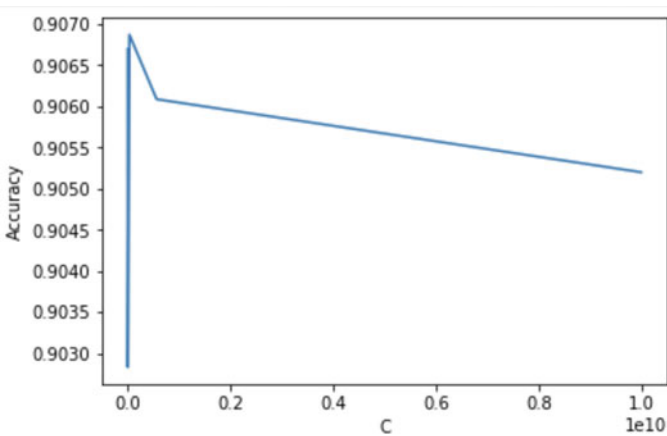


Fig. 7 Exp-2:Accuracy versus Control error V2

Experiments 3 and 4: SVM Now with SVM, we achieved an accuracy of around 94% (on the dataset without division). We tried SVM on the divided dataset and ran nested loop over C and gamma, and for $C = 5$ and $\text{gamma} = 0.09$, we got the accuracy of **99.4158%**. Another experiment with a different kernel in SVM (default is a radial basis function kernel (RBF)), i.e., polynomial kernel, resulted in reduced accuracy of 92.133. This paper [7] used SVM on the same dataset and got the best accuracy of 94.36%.

5.2 *Electricity Theft Detection*

Experiment 5 We have already discussed the preprocessing. Let us split the dataset into 80–20 for training and testing. Following is the table for results for every model we tested on. Now, we tried the gradient boosting method, which combines various methods which are weak and assigns weight to them. The classifier we tried vs accuracy is shown in the following Table 2.

We got 90.59 percent test accuracy with the CatBoost classifier, which is an improvement among the category of classifiers. Better results have been produced using neural networks and other combinations. However, in the case of just classifiers, this is better than the existing ones (published in the category of classifiers without using neural networks).

It is possible to increase the performance of a machine learning model based on gradient boosting and decision trees using the CatBoost Model. This may be used for categorical and continuous data values, making it even more versatile. CatBoost Classifier eases our burden of translating categorical data into the numeric form and begins creating the model, as well as we dive into the categorical values. The categorical characteristics or variables are enabled and handled automatically and treated as such. It has given us the best results as of now. Note that there is no work that compares existing methods or pre-built models (Gradient boosting versions) like we have used here.

6 Conclusion and Future Scope

In this research work, we have presented a lightweight IDS and electricity theft detection which can detect attacks with very high accuracy. We were able to get the improvement from 94.3%, which is existing work, to 99.4%. We used the ICS dataset published in 2014 and made the algorithm understand the difference between request and response, which lead to this huge spike in accuracy. We also provided it with information about the command request and response, which is the knowledge about the network packets. The second section of the work consists of the electricity theft detection on data released in 2017 by SGC of China. We tried basic methods and various pre-built versions of gradient boosting to improve the performance and

presented a comparison. A total of 10 different methods were experimented with. We established that though much recent work has already explored the neural network and other ways to optimize performance. However, for pre-existing classifiers, CatBoost is the recent one, and it gave better results than other previous classifiers. Further research can be done to improvise and not have the network knowledge while training.

7 Funding

The work is partially funded by the Department of Science and Technology(DST), and C3i-Hub (Indian Institute of Technology Kanpur), India, for the Risk Averse Resilience Framework for Critical Infrastructure Security(RARCIS) project.

References

1. <http://www.ece.uah.edu/thm0009/icsdatasets/IanArffDataset.arff>
2. Zheng Z, Yang Y, Niu X, Dai H-N, Zhou Y (2018) Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Trans Ind Inf* 14(4):1606–1615. <https://doi.org/10.1109/TII.2017.2785963>
3. Cat Boost classifier: <https://github.com/catboost/catboost>
4. Sayegh N, Elhadj IH, Kayssi A, Chehab A (2014) SCADA intrusion detection system based on temporal behavior of frequent patterns. In: MELECON 2014—17th IEEE mediterranean electrotechnical conference, pp 432–438. <https://doi.org/10.1109/MELCON.2014.6820573>
5. Bulle BB, Santin AO, Viegas EK, dos Santos RR (2020) A host-based intrusion detection model based on OS diversity for SCADA. In: IECON 2020 the 46th annual conference of the IEEE industrial electronics society, pp 691–696. <https://doi.org/10.1109/IECON43393.2020.9255062>
6. Singh VK, Ebrahim H, Govindarasu M (2018) Security evaluation of two intrusion detection systems in smart grid SCADA environment. In: North American power symposium (NAPS), pp 1–6. <https://doi.org/10.1109/NAPS.2018.8600548>
7. Lopez Perez R, Adamsky F, Soua R, Engel T (2018) Machine learning for reliable network attack detection in SCADA systems. In: 2018 17th IEEE International conference on trust, security and privacy in computing and communications/12th IEEE international conference on big data science and engineering (TrustCom/BigDataSE), pp 633–638. <https://doi.org/10.1109/TrustCom/BigDataSE.2018.00094>
8. Kumar A, Choi BJ (2022) Benchmarking machine learning based detection of cyber attacks for critical infrastructure. In: International conference on information networking (ICOIN), pp 24–29. <https://doi.org/10.1109/ICOIN53446.2022.9687293>
9. This research employs a SCADA dataset
10. Yang H, Cheng L, Chuah MC (2019) Deep-learning-based network intrusion detection for SCADA systems. In: IEEE conference on communications and network security (CNS), pp 1–7. <https://doi.org/10.1109/CNS.2019.8802785>
11. Altaha M, Lee JM, Aslam M, Hong S (2021) An autoencoder-based network intrusion detection system for the SCADA system
12. Zheng Z, Yang Y, Niu X, Dai H-N, Zhou Y (2018) Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Trans Ind Inf* 14(4):1606–1615. <https://doi.org/10.1109/TII.2017.2785963>

13. Abdulaal MJ et al (2022) Real-time detection of false readings in smart grid AMI using deep and ensemble learning. *IEEE Access* 10:47541–47556. <https://doi.org/10.1109/ACCESS.2022.3171262>
14. Lepolesa LJ, Achari S, Cheng L (2022) Electricity theft detection in smart grids based on deep neural network. *IEEE Access* 10:39638–39655. <https://doi.org/10.1109/ACCESS.2022.3166146>
15. Alkuwari AN, Al-Kuwari S, Qaraqe M (2022) Anomaly detection in smart grids: a survey from cybersecurity perspective. In: 3rd International conference on smart grid and renewable energy (SGRE), pp 1–7. <https://doi.org/10.1109/SGRE53517.2022.9774221>
16. Lee J, Sun YG, Sim I, Kim SH, Kim DI, Kim JY (2022) Non-technical loss detection using deep reinforcement learning for feature cost efficiency and imbalanced dataset. *IEEE Access* 10:27084–27095. <https://doi.org/10.1109/ACCESS.2022.3156948>
17. Ullah A, Javaid N, Asif M, Javed MU, Yahaya AS (2022) AlexNet, adaboost and artificial bee colony based hybrid model for electricity theft detection in smart grids. *IEEE Access* 10:18681–18694. <https://doi.org/10.1109/ACCESS.2022.3150016>
18. Xia X, Xiao Y, Liang W, Cui J (2022) Detection methods in smart meters for electricity thefts: a survey. *Proc IEEE* 110(2):273–319. <https://doi.org/10.1109/JPROC.2021.3139754>
19. Zhao Q, Chang Z, Min G (2022) Anomaly detection and classification of household electricity data: a time window and multilayer hierarchical network approach. *IEEE Internet Things J* 9(5):3704–3716. <https://doi.org/10.1109/JIOT.2021.3098735>
20. Althobaiti A, Jindal A, Marnerides AK, Roedig U (2021) Energy theft in smart grids: a survey on data-driven attack strategies and detection methods. *IEEE Access* 9:159291–159312. <https://doi.org/10.1109/ACCESS.2021.3131220>
21. <https://www.sciencedirect.com/science/article/pii/S2090447920301064>
22. <https://accelconf.web.cern.ch/ica99/papers/mc1i01.pdf>
23. Reynders D, Mackay S, Wright E (2004) Modbus overview. Edwin PY-2004/12/31, SP-132, EP-141, SN-9780750663953, T1. <https://doi.org/10.1016/B978-075066395-3/50012-7>
24. Ghosh S, Dasgupta A, Swetapadma A (2019) A study on support vector machine based linear and non-linear pattern classification. In: International conference on intelligent sustainable systems (ICISS), pp 24–28. <https://doi.org/10.1109/ISS1.2019.8908018>
25. Huang M (2020) Theory and implementation of linear regression. In: 2020 International conference on computer vision, image and deep learning (CVIDL), pp 210–217. <https://doi.org/10.1109/CVIDL51233.2020.00-99>
26. Ho TK (1995) Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition, vol 1, pp 278–282. <https://doi.org/10.1109/ICDAR.1995.598994>
27. Navada A, Ansari AN, Patil S, Sonkamble BA (2011) Overview of use of decision tree algorithms in machine learning. *IEEE Control and System Graduate Research Colloquium* 2011:37–42. <https://doi.org/10.1109/ICSGRC.2011.5991826>
28. Uhrig RE (1995) Introduction to artificial neural networks. In: Proceedings of IECON '95—21st annual conference on IEEE industrial electronics, vol 1, pp 33–37. <https://doi.org/10.1109/IECON.1995.483329>
29. Upadhyay D, Manero J, Zaman M, Sampalli S (2021) Gradient boosting feature selection with machine learning classifiers for intrusion detection on power grids. *IEEE Trans Netw Serv Manag* 18(1):1104–1116. <https://doi.org/10.1109/TNSM.2020.3032618>

Study and Analysis of Key-Predistribution Schemes Based on Hash Chain for WSN



Kanhaiya Kumar Yadav and Priyanka Ahlawat

Abstract In recent years, Wireless Sensor Network (WSN) has become a leading area of research. WSN is ad hoc, without infrastructure wireless network that connects a huge number of wireless sensors. WSN consists of multiple sensor nodes which have limited processing speed, storage capacity, communication bandwidth and base stations. They can be thought of as the network's "sensing cells" and "brain", respectively. They are emerging as a better technology in the future due to their large range of applications in surveillance and people-related domains. This paper proposes a study and analysis of hash-based key predistribution for a better understanding of the uses of key predistribution or random key predistribution over the wireless sensor network to prevent the makers and also provide better connectivity, lower storage overhead, lower communication overhead and less computation complexities.

Keywords Wireless sensor networks · Key-Predistribution management scheme · Hash chain-based key predistribution · Probability of resilience · Hashing · Security · Node capture attack

1 Introduction

A distributed wireless sensor ad hoc network is known as wireless sensor network (WSN). It can be defined as a wireless network with no self-configuration and no infrastructure for monitoring physical or environmental conditions, i.e. sound, temperature, vibration, environmental pollution and various monitoring areas where information can be seen and analysed. The base-station act likes a medium between users and the network. The base station servers as a connection point between users

K. K. Yadav (✉) · P. Ahlawat
Department of Computer Science and Engineering, NIT Kurukshetra, Kurukshetra, India
e-mail: kanhaiya.yadav10@gmail.com

P. Ahlawat
e-mail: priyankaahlawat@nitkkr.ac.in

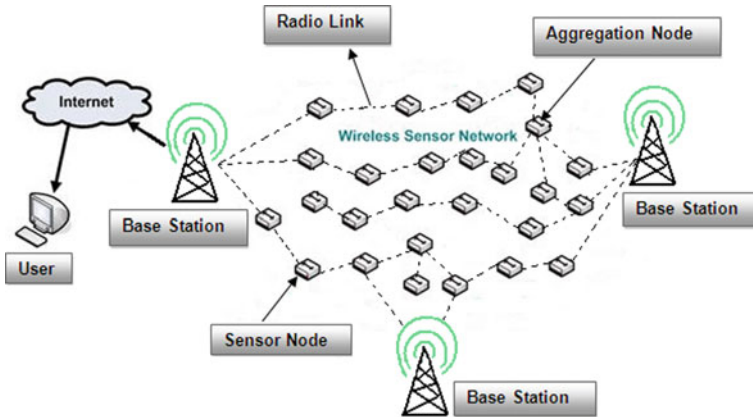


Fig. 1 Wireless sensor network

and the network. Specific queries can be used to get critical information from the network, and the results can be gathered from the base or sink [1].

As a new technology, WSN is being widely used in various domains and fields, including ecological monitoring, smart home and industrial fields. However, there are some important challenges too with the working environment of WSN. Some of these important challenges are network coverage, deployment optimisation of nodes, network energy consumption, hardware constraint, fault tolerance, scalability and security (Fig. 1).

From the list of challenges, one of the big issues in WSN is security, if nodes are not secure then the whole network will damage. Sensor networks are highly vulnerable for node capture threats. The intruder physically seizes the node, allowing him to perform a variety of network activities and easily compromise the entire network. Intruders can access confidential data from the memory of the node through this attack. Though node capture attack seems to be devastating, this attack has extremely low attack efficiency, and the cost and time to attack as required number of nodes to compromise the full network are high. WSN nodes use the shared key to exchange messages that a potential attacker can use the old key to execute a replay attack after new key is updated and transmitted to each node in the WSN [2].

Therefore, to enhance the security of the node we study and analyse and come to a good conclusion, rather than apply direct key to a node identifier, apply hashed key on the node, and hashing is directly proportional to the node identifier. It overcomes these limitations. So we can combine the multiple objectives like large node contribution, least resource and shared key contribution so as to discover an optimal solution for the problem.

1.1 Features of Wireless Sensor Network

The WSN features are as follows:

- Power limit consumption of nodes with batteries.
- Ability to deal with node failures.
- Some degree of node mobility and node non-uniformity.
- Scalability to large distributed scales.
- The capacity to ensure adverse environmental circumstances.
- Easy to use.
- The design is cross layered.

1.2 Application of Wireless Sensor Networks

Applications of sensor networks are in a very large scale as follows:

- Applications in the military: In radio or wireless communications, computing, intelligence, war zone surveillance, detection purpose and targeting enemy systems are anticipated to all use wireless sensor networks as a key component.
- Transportation: WSNs are gathering real-time traffic data to subsequently utilise transportation-model and notify drivers of the vehicle for congestion and traffic report.
- Health-related applications: Tele-monitoring of human physiological data, an environment for the disabled, integrated monitoring of patients, and drugs management and doctors or patients management inside a hospital.
- Environmental sensors: The name “Environmental Sensor Networks” has evolved to encompass a wide range of WSN applications in earth scientific research. Sensing of forests, seas, glaciers, volcanoes, traffic-lights, habitat, guards or security, industrial sensing, etc. is included in this. Below is a list of further important areas:
 - Air pollution monitoring.
 - Forest fires detection.
 - Greenhouse monitoring.
 - Landslide detection.
- Structural monitoring: using the WSN, structure of the bridge, rooms, and caves are able to be captured using movement of particles. The best example is in robotic.

2 Background Knowledge

2.1 Motivation

New sensors are designed, and information technologies and wireless systems are the results of recent advances in engineering, communication and networking. These sophisticated sensor nodes can act as connection edge between the physical and digital environment.

Sensors are used in lot of devices, industries, mechanical machinery and the environment to keep on eye over infrastructure failures, accidents emergencies, natural resource conservation, wildlife preservation, increased production and security. We can build more powerful processors that are smaller in size than previous generation and good with the help of this era semiconductor technologies in terms of process speed. Small, low-power and low-cost sensors, controllers, and actuators have resulted from the miniaturisation of advance computation and process technologies.

Thus, a huge feature of wireless sensor network makes its security a main concern for being used. The challenges may impact the design of wireless sensor networks. So WSN has a mind blowing impact on our life. To overcome this constraint, we simulate various node-capturing attack techniques.

A number of security services, as secrecy, password protection and so on, require key management as most important thing.

The effective key management services in traditional networks using public key-based solutions are worthless for WSNs owing to resource limits. As a result, the best paradigm for securing (WSN) wireless sensor network is a symmetric-key setup.

Literature Review. A distributed wireless sensor ad hoc network is called a wireless sensor network. Specific queries may be used to get critical information from the network, and the results can be retrieved from the base or sink. At the moment, these sorts of networks are just in the process of being implemented in real time. However, in the next years, this technology may be applied in a variety of fields all over the world [3]. From its multiple concerns, WSN's largest challenge is security. It is vulnerable to many types of physical assaults by intruders because to its openness to the outside environment. Node capture attack is a significant assault in wireless sensor networks that allows an attacker to do numerous operations on the networks and quickly compromise the entire network. Node capture attack is one of the hazardous attacks in WSNs. Many security services, for secrecy, authentication and so on, require key management as one of the required building elements [4]. The key pool is further divided into three sub-key pools which are non-overlapping sub-key pools, each hashing key based on cell and node IDs to improve host takeover resistance and communication overhead. It's also shown that, when compared to other systems, the proposed system outperforms them in terms of connection probability and processing overhead [5].

The security of EG scheme is enhanced and a scheme is suggested known as the q-composite scheme. In this scheme, at least q number of keys are common between

the adjacent nodes for the setup of a communication link [2]. If shared keys are more than q -keys between the adjacent nodes, then the setup will be like take the hash of every shared key. This is efficiently good for small-range attacks, but for the larger attacks, it fails. Bechkit et al. for the improvement of resilience against node capture of the key pre distribution scheme [2]. In this scheme for the improvement of resilience against node capture hashing the key to their node identifiers. Lin et al. [6] proposed a scheme that is built on the irreparable of the hash chain and deployment-based knowledge which is known as a pair-wise key distribution scheme. Here key pool is of two levels using random key predistribution which was introduced by Mohaisen et al. [6], key pool is divided into various non-overlapping sub-key pools. It concludes that its computation overhead, storage overhead, and communication overhead are improved ones. Based on the work of Blundo et al. [7], an extension of the basic random key predistribution (RKP) is proposed. Liu and Ning [8] proposed a polynomial-based key predistribution scheme for wireless sensor networks (WSNs). In this, nodes are pre-loaded by polynomial at the place of keys and to initiate a connecting link among two nodes a along with b , node “ a ” evaluates its polynomial at the tag of the neighbour node “ b ” and vice versa. So the secret key between nodes a and b is $K_{a,b} = P_a(b) = P_b(a)$, whereas node identifiers are unique. For the multiphase wireless sensor networks (WSNs), a hash graph-based key predistribution scheme is suggested by Sarimurat and Levi [9]. This is a generation-based scheme, in this, every generation of nodes utilises a dissimilar key pool which is produced by the support of the previous generation key pool and hence makes the different generations of nodes communicate securely. This increases the connectivity of the nodes with different generations but decreases the resilience.

Based on a variety of essential variables, including network resilience in case of node capture; secure connectivity for coverage; the requirement of storage, overhead for communication, and computational complexity, we evaluate and compare our schemes with already proposed schemes. As a result, when building security protocols, we must consider the possibility of an adversary intercepting traffic and compromising some sensor nodes. Security protocols for WSN communication systems rely on key management. Pair-wise key distribution (PKD) is the initial type of key management in a WSN, and it involves creating a pairwise key between two sensor nodes [1, 2]. Group key distribution (GKD) is the second kind, which allows all sensor nodes in a group communication to share a single conference key.

In this paper, we analyse all schemes and give an overview about all the schemes.

Key Predistribution. One method for key predistribution in sensor networks is to produce a huge pool of keys, select a small percentage at random (referred to as the keyring) and store it in each sensor [1].

This is a random KPS in which two sensors can establish a direct key by having similar keys in their key rings [10]. Those who don't have common keys may be able to create an indirect key by following a key path that includes sensors with direct keys. This strategy is the first real solution to the critical predistribution problems in sensor networks.

Types of Key Predistribution. Pair-wise key distribution (PKD) is the first form of the key management system, which is creation of a pair-wise key between two sensor nodes [1]. Group key distribution (GKD) is the second kind, which allows every sensor node in a group communication to share a single conference key [2].

2.2 Key Management Schemes and Its Type

- A key management system (KMS) is a collection of processes and systems that ensure secure communication between authorised parties in accordance with a predetermined security policy [2].
- The probabilistic and deterministic schemes are two types of KMSs.
- In comparison to deterministic methods, probabilistic schemes have a higher storage overhead.
- Deterministic schemes that assure entire protected connectivity coverage and probabilistic schemes that are not guaranteed and are dependent on the presence of shared keys.
- As a result, the best paradigm for securing WSN transactions is symmetric-key configuration. We normally don't have a trust in third party who can assign a pair-wise secret key to surround nodes in WSNs because of the lack of infrastructure, and therefore key predistribution is the best solution.

Based on the following five parameters, key management conspires in WSN, we can evaluate performance

- **Network resiliency against node capture.** Sensor nodes in WSN are typically not tamper resistant due to resource constraints. An adversary gets access to all secret messages stored in a node's memory if it is penetrated.
- **Secure connectivity coverage.** The likelihood of a given pair of neighbouring nodes establishing a secure link.
- **Computation complexity.** To create direct secure links, the magnitude of data sent between pair of nodes is measured by the computation overhead.
- **Overhead in communication.** To create direct secure links, the quantity of information sent between a node pair is measured by the communication overhead.
- **Overhead in storage.** Sensor nodes are limited in terms of memory resources due to their small size.

2.3 Key-Predistribution Schemes

To increase the resiliency of the network, a hash-based approach may be used to current pool-based key-predistribution systems. To accomplish it, we are presenting new approach, one-way hash chain that applies hash function in such a way like if

attacker gets a key of a node then he can't move backward so backward nodes are safe.

EG-scheme. The very first scheme is the EG scheme [1], introduced by Eschenauer and Gligor in 2002, which is based on a probabilistic scheme. This is a basic approach, i.e. random key predistribution scheme (RKP). This scheme is suitable for distributed networks and is based on the probability theory. It works in three stages—A. The Key Predistribution over the nodes keeps the keys with key identifiers in the sensor nodes. B. Finding the common key between the nodes, in this every node broadcasts its identifiers, and neighbouring nodes have to find a common key. C. Establishment of the path keys, in this stage on condition that nodes are not able to discover a common key they will set up a path key with a node that is common in both end nodes. Each node exchanges a list of key identifiers with neighbouring nodes. A further enhancement is kept going in terms of security, computation overhead, connectivity and storage overhead.

q-Composite scheme. In [8], Chan et al. devised a protocol dubbed the q-composite scheme that improves RKP's robustness. Only if two neighbouring nodes share at least q-keys can they build a safe link in this solution. The hash of every shared key concatenated to each other yields the pairwise session key, and this strategy reduces the network's secure connectivity coverage.

HC (q-composite) scheme. [11], The suggested protocol, dubbed HC (q-composite), is more detailed and resistant to node capture.

Before the WSN is implemented, a big pool $|P|$ of keys and their identities are produced offline. Every node has a keyring and with M number of keys randomly selection is to be from the key pool—P. Prior to the deploying, we applied a hash function H ; $I \bmod L$ times to the pre-loaded keys of every node, here I is the node identity of each node and L is the class parameter for each node, enabling this to minimise the number of hash operations as mentioned earlier.

2DHC scheme. Ehdai et al. [12] utilise the commutative 2DHC to expand 1DHC-KPD system of Bechkit et al. to the 2DHC-KPD (two-dimensional approach) (b). Sensor node (a, b) , where $0 \leq a, b \leq (l-1)$, holds the key $x_{a,b} = h_a h'_b(\times 0,0)$.

The following is a concise statement of the approach for constructing a secure connection or link between two sensor nodes A and B using the keys $k_{aA}, b_B = h_a h'_b(k_0,0)$ and $k_{aB}, b_B = h_a h'_b(k_0,0)$, where $0 \leq a_A, b_B \leq (l-1)$. In this approach, hashing is two dimensional, i.e. vertical and horizontal.

DLHC scheme: Despite the fact that the 1DHC-KPD and 2DHC-KPD regimes improve resilience relative to the HC-KPD regime, they continue to suffer from ANA [13]. By using the diagonal layer distribution of sensor nodes from the 2-D keymap in the 2DHC-KPD conspire, compared to the 1DHC-KPD scheme, we can improve the resilience of node capture attack while simultaneously repelling the ANA.

Location-based key management schemes—A location pointed key generation and distribution system was put out by Younies and Eltoweissy [14].

This technique generates keys using the location and exclusion-based system (EBS). The generated keys are random and strict to each node based on the location. It is based on SHEEL.

Key generation and distribution methods depending on point or place were provided by Choi et al. [14]. In this method, network keys are created using grid-based coordinates. Nine data coordinates and eight neighbour coordinates are utilised. These coordinated paired keys are established during the network’s first and second stages. The sequence number of every packet that a node sends is also used. This technique offers protection against numerous internal and outer threats.

Blom proposed scheme—In this scheme, a generator matrix of the maximum distance separable (MDS) code and a it uses symmetric matrix is to generate key spaces for nodes [15].

It uses symmetric matrix small amount of data which is generated,so dependencies among the keys will exist, and hence users in the system might be able to degrade their uncertainties about keys. The MDS code is mainly defined by the condition that the minimum required range of the code is $n-k + 1$.

2.4 Comparison of Various Key Management Schemes

Key management schemes	Structure type	Advantages	Limitations
EG-2002	Probabilistic key management, pair-wise	Encryption is end to end Implementation is simple Scalability is excellent	Huge memory is required Increasing the number of capture nodes degrades the performance
Chan et al. 2003	Pair-wise, q-composite scheme	Increase the resilience against EG scheme Only q-keys are common so less computation	Lower key connectivity than EG scheme
Du et al. 2004	Based on deployment scheme, pair-wise and network-wise	Key connectivity and resilience improved than EG and q-composite scheme Simple and reasonable energy and cost	Deployment information is required earlier More complex and not fit for group-based network
Ling et al. 2008	Polynomial based	t collusion resistant and secure	Overhead is very high in terms of memory and computation
Du et al. 2007	Non-balanced predistribution	Good for low-end sensors and energy efficient	It is tough to use for homogeneous WSN

(continued)

(continued)

Key management schemes	Structure type	Advantages	Limitations
Bechkit et al. 2013	Hash chain based	Resilience against node capture is increased	Storage and computation overhead in terms of hash function
Mohaisen et al. 2010	Key pool are applied in two levels	Communication overhead is reduced	It becomes basic scheme and results in high overhead in terms of storage, communication and computation when all the sub-key pool identifiers are matched
Gandino et al.	Symmetric and pair-wise	Energy efficient and scalable	Storage overhead Take more time in starting up
Zhu et al. [16]	Group based, pair-wise	Complexity and scalability is reasonable	Storage overhead Weak initial security
Younies et.al	Pair-wise and random	Location-based system, efficient and lightweight system	High complex and storage overhead
Choi et al.	Pair-wise	Based on location and grid data, self-adjustable and unique	Not robust and not scalable
Qin et al.	Based on public key	Energy efficient, computation time, cost-effective in storage and communication	Complexity is high
Yao et al.	Based on clustering, tree type structure	Secure against node capture offense. It supports rekeying and refreshing	Cost is high for computation
Swaminathan et al.	Based on clustering, distributed spanning Tree structure	Cost is low	Computation and storage is high
Lu et al.	Asymmetric predistribution, keys are pre-configured	Supports heterogeneous model and good in terms of connectivity, reliability and resilience	Cryptographic overhead is high

(continued)

(continued)

Key management schemes	Structure type	Advantages	Limitations
Wang et al.	Group-wise key distribution	One-way hash-based key chain, key distribution is group-wise. Strong collision attack impedance capability	Secure group management setup is difficult
Suganthi et al.	Group-wise and pair-wise key, ID shared	Overhead is low in computation, communication and storage	Issue in large-scale key distribution of mobile nodes
Zhan et al.	Symmetric matrix, polynomial	Supports multicast and key regeneration	Overhead is high for cryptographic

The above table shows a comparative analysis between various schemes. We have observed that many of them target to produce security by using concept of location [17], generation, q-keys overlaying and hash functions.

3 Results and Observation

3.1 Comparison Table in Terms of Resilience

KPD schemes	Probability of link compromise	Resistance
q-composite scheme	$p(LC NCx) = \sum_{j=q}^m \left(1 - \left(1 - \frac{m}{ S }\right)^X\right)^j \frac{p(j)}{p}$	No
HC(q-composite scheme)	$p(LC NCx) = \sum_{j=q}^m \left(1 - \left(1 - R \cdot \frac{m}{ S }\right)^X\right)^j \frac{p(j)}{p}$	No
1DHC scheme	$p(1D(n)) = \sum_{j=q}^m \left(1 - \left(1 - a \cdot \frac{m}{ S }\right)^X\right)^j \frac{p(j)}{p}$	No
2DHC scheme	$p(2D(n)) = \sum_{j=q}^m \left(1 - \left(1 - a' \cdot \frac{m}{ S }\right)^X\right)^j \frac{p(j)}{p}$	No

(continued)

(continued)

KPD schemes	Probability of link compromise	Resistance
DLHC scheme	$P(DL(n)) = \sum_{k=q}^m \left(1 - \left(1 - a'' \frac{m}{ S } \right)^X \right)^k \frac{p^{(k)}}{P}$	Yes (More Resilience)

where

$$R = \frac{L + 1}{2L}$$

$$a = \left(1 - \frac{(l-1)(2l-1)}{6l^2} \right)$$

$$a' = \left(1 - \frac{(l'-1)(2l'-1)}{6l'^2} \right)^2 = \left(1 - \frac{(\sqrt{l}-1)(2\sqrt{l}-1)}{6l} \right)^2$$

$$a'' = \left(1 - \frac{(l''-1)(2l''-1)}{3l''^2} \right) = \left(1 - \frac{(l-1)(2l-1)}{3l^2} \right)$$

The chances of the compromised are less during an advanced node capture attack by the attacker with more number of times hashing applied.

That is, increasing the number of hashing increases the resilience (total number of nodes in the network divided by number of captured nodes). So the security will be enhanced and also reduces the probability of node captured. That is why resilience is an important factor to analyse schemes. Here Diagonal Layer hash chain-based key-predistribution scheme is more resilience and it is more suitable.

$$P(N) > p[Q(N)] > p[HC - Q(N)] > p[1D(N)] > p[2D(N)] > p[DL(N)]$$

4 Conclusion

The study includes many aspects of the existing key management schemes and their key predistribution techniques, which is shown in Sect. 2.4. That helps in selection of proper scheme according to the use and application of wireless sensor networks. Whereas faith is another main angle that impacts the security in WSN. Moreover increasing the hashing factor will increase the security which further may lead to computation overhead based on network architecture, but security is guaranteed. DLHC scheme is more suitable in terms of security and overhead. The hashing at multiple level reduces the probability of compromised key and communication overhead.

5 Future Works

The robustness and security of optimisation techniques are further enhanced in the future. Despite the positive results, there is definitely room for improvement in the future. To begin with, because the time of capturing a network is very long, we can focus on the time complexity element of enhanced key-predistribution strategies over nodes.

References

1. Yang C-N, Kao C-L, Wang C-J (2021) Two-dimensional diagonal layer hash chain based key pre-distribution scheme. *J Inf Security Appl* 63:103038
2. Bechkit W, Challal Y, Bouabdallah A (2013) A new class of Hash-Chain based key pre-distribution schemes for WSN. *Comput Commun* 36(3):243–255
3. Ahlawat P, Dave M (2018) Deployment based attack resistant key distribution with non-overlapping key pools in WSN. *Wirel Pers Commun* 99(4):1541–1568
4. Castelluccia C, Spognardi A (2007) A robust key pre-distribution protocol for multiphase wireless sensor networks. In: *IEEE Securecom*, 351–360
5. Liu D, Ning P (2003) Establishing pairwise keys in distributed sensor networks. In: *ACM CCS*, pp 52–61
6. Chan H, Perrig A, Song D (2003) Random key predistribution schemes for sensor networks. In: *Proceedings of 2003 IEEE symposium on security and privacy*, pp 197–213. California, USA
7. Ruj S, Nayak A, Stojmenovic I (2011) Key predistribution in wireless sensor networks when sensors are within communication range. In: Nikolettseas S, Rolim J (Eds), *Theoretical aspects of distributed computing in sensor networks. Monographs in theoretical computer science. An EATCS series.* Springer, Berlin
8. Simonova K, Ling AC, Wang XS (2006) Location-aware key predistribution scheme for wide area wireless sensor networks. In: *Proceedings of the fourth ACM workshop on Security of ad hoc and sensor networks*, pp 157–168. ACM
9. Yu CM, Li CC, Lu CS, Kuo SY (2011) An application-driven attack probability-based deterministic pair wise key pre-distribution scheme for non-uniformly deployed sensor networks. *Int J Sensor Netw* 9(2):89–106
10. Chan H, Perrig A, Song D (2003) Random key predistribution schemes for sensor networks. In: *IEEE SP '03*
11. Du W, Deng J, Han Y, Chen S, Varshney P (2004) A key management scheme for wireless sensor networks using deployment knowledge. In: *IEEE INFOCOM*, pp 586–597
12. McWilliams FJ, Sloane NJA (1977) *The Theory of error correcting codes.* North-Holland, New York
13. Eschenauer L, Gligor VD (2022) A key-management scheme for distributed sensor networks. In: *ACM CCS '02*, pp 41–47
14. Gautam AK, Kumar R (2021) A comprehensive study on key management, authentication and trust management techniques in wireless sensor networks. *SN Appl Sci* 3(1):1–27
15. Blom R (1984) An optimal class of symmetric key generation systems. Report LiTH-ISY-I-0641, Linkoping University
16. Zhu S, Setia S, Jajodia S (2003) Leap: efficient security mechanisms for large-scale distributed sensor networks. In: *ACM CCS*, pp 62–72
17. Liu D, Ning P (2003) Location-based pair wise key establishments for static sensor networks. In: *Proceedings of the 1st ACM workshop on security of ad hoc and sensor networks*, pp 72–82. ACM

CERT-In New Directives for VPN: A Growing Focus on Mass Surveillance and Data Privacy



Neeraj Jayant, Naman Nanda, Sushila Madan, and Anamika Gupta

Abstract Digitalization efforts are rewarding as Information Technology is bringing changes in almost every sector. Virtual Private Network (VPN) was expected to be a safeguard for sensitive and personal information for individuals. The focus of India's cybersecurity watchdog, Indian Computer Emergency Response Team (CERT-In), focuses on safeguarding or prevention with feasible effort. It is difficult to maintain data privacy without hampering user identity. CERT-In directives try to enhance cybersecurity by bridging the gap in cyberincidence analysis. VPN is ever growing with Bring Your Own Device (BYOD), Work From Home (WFH) in place. A VPN allows users to browse the Internet while masking their device's IP address, encrypting data, and routing through secure networks in other states or countries with no logs. The new CERT-In directives emphasize obligatory data collection, retention, and integration for Virtual Private Server (VPS) providers, VPN services, and Cloud providers for a minimum of 5 years. There is an urgent need to increase the security of the country's digital infrastructure in the best feasible ways, but some new directives may not be privacy-friendly hampering user identity and data protection framework. It has major market implications and an increase in operational costs. Thus, making an Un-CERT-In time for VPN providers in India. This directive does not only defeat the purpose of VPNs but is also possibly aimed at state-sponsored surveillance. We have proposed a few solutions to go through this new rule for the end users.

Keywords CERT-In directives · Data privacy · Information security · VPN · Surveillance · OpenVPN · Data protection framework

N. Jayant (✉) · N. Nanda · S. Madan · A. Gupta
Delhi University, New Delhi 110089, India
e-mail: neeraj21713@sscbs.du.ac.in

N. Nanda
e-mail: naman21712@sscbs.du.ac.in

A. Gupta
e-mail: anamikargupta@sscbsdu.ac.in

1 Introduction

CERT-In aims at strengthening the country's cybersecurity infrastructure and is taking adequate steps for the same. The new directives that were released by CERT-In on April 28, 2022 aim at logging of the user data and presenting it to the government as and when required. The services that need to adhere to this directive are the cloud service providers, data centers, and VPN companies. The directives come out to be the complete opposite of what these service providers claim to provide, that is, "anonymity" [1]. There are two sides to the coin one being the cybersecurity concerns pertaining to the country and the second being the breach of privacy and mass surveillance by the government. We aim to emphasize and pinpoint the important aspects of the directives and provide a viable solution for the same.

VPN is developing exceptionally quickly. The impact of recent innovations and COVID-19 has led to major changes in the industry with Bring Your Own Device (BYOD) and Work From Home (WFH) increasing the need for security [2]. VPN was expected to be a safeguard for sensitive and personal information for the individuals. Our examination demonstrated that almost 50% of users use VPNs for general security reasons, such as staying away from identity theft. While 40 percent involved VPNs for general protection from hackers and their snooping on public networks [3]. The other 10% used it for more uncommon reasons which were bypassing school, office, school, or government restrictions set by firewalls [4]. VPN provides a sense of security as it ensures the traffic is encrypted and passes through a secure tunnel preventing any leaks. Thus making it a preferred choice over proxy servers. India is one of the countries with the most noteworthy VPN use. This might be somewhat in light of the fact that there isn't a lot of web opportunity in India [5]; occupants have limited admittance to virtual entertainment and "negative substance" which can incorporate the accompanying pornography, psychological oppression, extortion, hate speech, misleading data, defamation, etc.

- The other major use cases of VPN are as follows:
- Data privacy from your ISP.
- Data privacy from the Government.
- Safety on a public Wi-Fi.
- Blocking malware when accessing the Internet.
- Secure access to networks when accessed remotely.
- Access to geolocation-specific content.

A user installs a VPN client which helps them encrypt the data that is being sent over to the VPN server using VPN protocols. The ISP is unable to identify the encrypted data and simply forwards the requests to the VPN server. The VPN server does the rest by hiding the IP and geolocation of the user. The user is able to surf the Internet anonymously while connected to the VPN (Fig. 1).

Secure Sockets Layer (SSL) and Transport Layer Security (TLS) together create a VPN connection where the host is a web browser and the user has restricted access to the application. Online shopping portals commonly use these protocols. Web

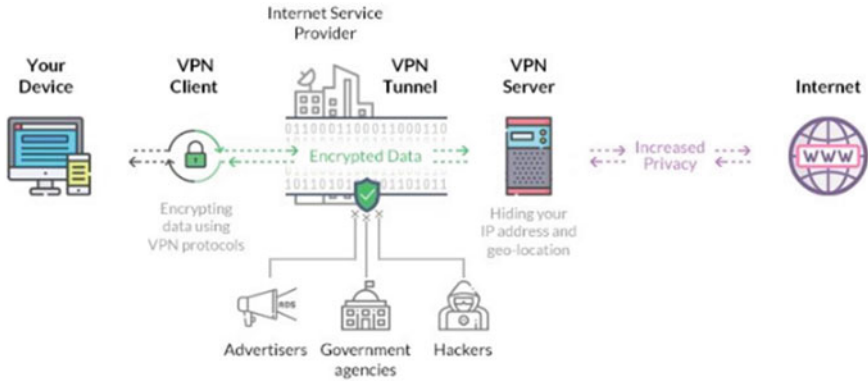


Fig. 1 A figure caption is always placed below the illustration. Short captions are centered, while long ones are justified. The macro-button chooses the correct format automatically [6]

browsers can easily switch to SSL as they are embedded with both SSL and TLS. “HTTPS” in the first part of the URL denotes a SSL connection instead of “HTTP”. The ISP, however, sees the following information when you use a VPN:

- The IP address of the VPN server.
- The timestamp of when you connected.
- The port your VPN protocol is using.
- The amount of data you’re sending or receiving.
- Encrypted and unreadable data traveling between you and the VPN server.

1.1 VPN Protocols

PPTP—The point-to-point tunneling protocol developed by Microsoft is one of the oldest protocols in existence [7]. The protocol was essentially used for dial-up connections.

L2TP/IPSec—PPTP was replaced by the Layer 2 Tunneling Protocol. The protocol itself is not very secure and needs to be used in unison with a security protocol. It is usually clubbed with the IPSec protocol [8].

SSTP—Secure Socket Tunneling Protocol is a very popular protocol that has been used by Microsoft ever since Windows Vista was introduced [9]. This protocol is popular because of the 2048 SSL/TLS for encryption and 256-bit SSL keys.

OpenVPN—This is an open-source protocol with the strongest encryption. Being open source, the underlying code of the protocol can be accessed and modified according to the requirements of the developer [10]. This protocol is more popular because of its open source and strong encryption capabilities.

There have been a lot of research papers on the technical aspects of VPN, the security they use, and the technologies that they use for encryption or data transmission.

A case study based on a real-world scenario has never been published or discussed. There are works based on GDPR and its impacts on privacy and advertisement-providing companies (Tobias Urban, Dennis Tatang, Martin Degeling; June 2020) [11]. Our research focuses mainly on the average Indian user and the repercussions of the new directives by the governing body. The ways to deal with the scenario change that will be brought forth with it. We aim to emphasize and pinpoint the important aspects of the directives and provide a viable solution to manage the data privacy from CERT-In new directives and VPN mass surveillance, impacting the Indian companies.

2 Discussion

2.1 *New CERT-In Guidelines and Its Repercussions*

The Indian Computer Emergency Response Team (CERT-In) released a new privacy rule in April 2022 asking all the VPN companies, cloud companies, and data centers functioning in India to hold customer data and Information and Communication Technology (ICT) transactions for a period of 5 years. They released this in an effort to counter cyberattacks and strengthen the online security of the country. The new guidelines would require VPN administrations and other cloud specialist co-ops to log all client information and ICT exchanges over a period of time. The VPN business has reproved the new mandates, saying that such tough regulations conflict with the essential reason and strategy of Virtual Private Networks. Some of the top VPN providers have already eliminated their actual Indian servers.

The detailed analysis of the rule mentioned in Directive 5 is that the Data Centers, VPS providers, Cloud Service providers, and VPN companies need to hold the user's information for 5 years or longer and hand it over to the government as and when required by them. The data should be held even after the user has canceled their subscription [12]. This meant that any VPN company residing in India with physical servers need to store the following information:

- Full name and address.
- Phone number.
- Email address.
- Actual IP address.
- New IP address (issued by the VPN).
- Timestamp of registration.
- Ownership pattern of customers.
- Purpose of using the VPN.

The study demonstrates a correlation between the many VPN-providing companies that have already pulled off their business from India by taking down physical servers [13]. They will however continue to operate in India by providing virtual

servers. A virtual server is a software representation of a physical server. It surely provides the functionality of a physical server but also lacks the underlying machinery and power. The issues with virtual servers are that they are very high resource hogging and perform low. A physical server that hosts many virtual servers creates this issue and users face the issues of lower bandwidth and slower load times [14]. Thus, the biggest advantage of virtual servers is that they can redirect the data and don't force the user to abide by the new Section 70B directives for VPS services. Moreover, these guidelines also issue a statement regarding the cloud service providers and crypto- or virtual asset-providing companies. According to the directives, virtual asset exchange providers and custodian wallet providers as defined by the Ministry of Finance from time to time shall mandatorily maintain all Know Your Customer (KYC) details [12] and records of financial transactions for a period of 5 years. These again hint toward the breach of the identity of an individual.

According to reports, India is pushing for international action to stop unauthorized access to technologies like virtual private networks (VPNs), end-to-end encrypted messaging services, and blockchain-based products like cryptocurrencies. Indian officials made recommendations to members of a United Nations Ad Hoc committee that was debating a comprehensive international convention on combating the use of information and communication technologies for criminal purposes saying that "the anonymity, scale, speed, and scope offered to (terrorists) and the increasing possibility that they will remain untraceable to law enforcement agencies" by using these technologies continues to be one of the major challenges the world faces.

This came as a shock to the VPN companies as it is the exact opposite of what they advertise. The VPN companies offer complete anonymity and follow a strict no-log policy. This means that they do not hold or retain any of the customer data. They function on volatile RAM-based servers [15] and as soon as they are powered off the data is lost. The new laws also have created a backlash regarding the privacy of users in India. The Internet Freedom Foundation (IFF) [16] has actively been asking questions regarding the new laws and appealing the annulment of this decision. They claim that the collection of data will prove to be a bigger threat to the cybersecurity of the country and also will result in a breach of privacy of the individuals. It would also mean that the costs of these services will increase as the companies will have more data centers to hold such volumes of data.

Currently, the directives placed in order to curb cybercrime in India are seeking integration challenges to the existing system. There is no law in the country like the General Data Protection Regulation (GDPR). A bill that is pending, known as the PDP bill (personal data protection), is an Indian adaptation of the GDPR of the EU [17]. We have outlined some of the most crucial factors that must be taken into account in the PDP bill in order to address the security implications of the new directives. Once passed as an act, there would be more clarity on the data retention policies and privacy of the user's data held by the VPN, cloud, and data center companies.

2.2 Proposed Solution

This proposed method as per the compliance aspect—The new directives by CERT-In create a lot of ambiguity for the VPN-providing companies as currently there is no law in India governing data privacy and data protection. A bill that is pending, known as the PDP bill 2019 (personal data protection), is an Indian adaptation to the General Data Protection Regulation (GDPR) of the EU. Once passed as an act, there would be more clarity on the data retention policies and privacy of the user's data held by the VPN, cloud, and data center companies.

Currently, the IT Act 2000 is the governing umbrella under which some provisions that safeguard the privacy of the user exist. For example, Section 43 deals with the loss or damage to the personal computer and the compensation that the victim is entitled to by the attacker. Section 43-A [18] specifically deals with the Compensation for Failure to Protect Data. But many organizations or service-providing companies find loopholes and get past them. A PDP bill [19] if enforced will protect the interests of the consumer further and more efficiently. Some of the important aspects of the PDP bill are mentioned below:

1. **Right to Access Data Handling**—This gives the individuals a right to request information about their data from the data fiduciary. They can also request to ask if their data has been processed by the data fiduciary or not. Organizations engaged in profiling or observing the behavior of Indian citizens will be subject to additional requirements. The individuals can request a summary of how their data has been processed. The data fiduciary needs to submit the information to the individual in a clear and readable format.
2. **Right to Control Personal Data**—The individuals have a right to correct misleading or inaccurate personal data. They have the right to complete/update the incomplete personal data. They also have the right to erase personal data that is no longer required by the data fiduciary for the purpose for which it was processed. Organizations must adopt security procedures that enable them to keep track of information and activity and safeguard information by creating contracts in writing with vendors.
3. **Right to Data portability**—The individuals have the right to obtain their personal data in a structured and machine-readable format from the data fiduciary. This will enable them to transfer, copy, move, and reuse their personal data across different IT environments and services.
4. **Right to Data Ownership**—Individuals have the right to restrict or prevent the disclosure of their personal data when the data has served the purpose for which it was being collected by the data fiduciary.
5. **Right to Data Breach Notification**: Data breaches must be reported by organizations to the appropriate authorities and, in some cases, the impacted individuals.

The PDP bill is based on the guidelines of the GDPR of the EU and is directed toward the data privacy and protection of individuals. Once this bill is passed as an



Fig. 2 GDPR and data logs

act there would be less ambiguity and confusion related to any of the new directives that may be released by CERT-In in the future pertaining to data and its privacy and protection.

This other proposed method of creating and implementing most secured, open-source, and self-reliant VPN—currently, VPN companies are not bound to audits as there are no trails left behind. After these laws come into action, there would be ease of auditing and the service-providing companies would be kept in check (Fig. 2).

If the PDP bill is not passed by the government and things continue to function as they are, the users could create a self-hosted VPN [20] and continue surfing the Internet without the interference of the new directives. Self-hosted VPNs are nothing but a user creating a VPN of his own by purchasing a VPS of his choice. Self-hosted VPNs are simply VPNs that a user creates on his own by selecting and paying for a VPS. Users with an IP from that particular country are able to browse the Internet owing to the VPS of that nation. Data is encrypted over a tunnel while the user's original IP address is hidden, ensuring their anonymity. Numerous websites offer the option for customers to buy a VPS and browse incognito. These virtual private servers are governed by local laws in that area. This is due to the fact that the physical servers needed to route user traffic are situated there. CERT-recommendations In's advise keeping logs on people who use the services.

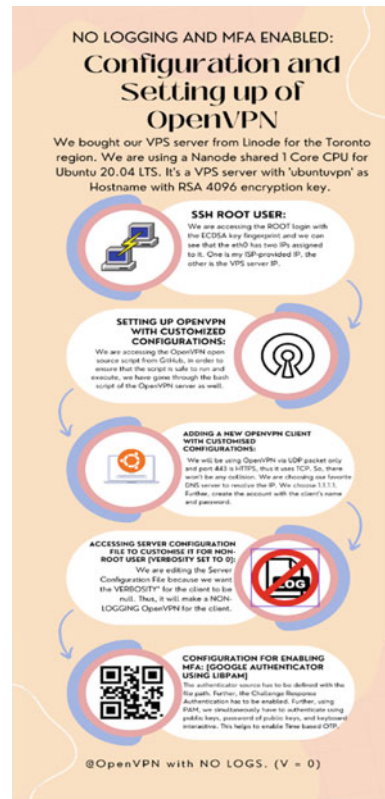
The whole purpose of this solution is to have a secure and reliable VPN connection to be able to access the Internet. The new laws according to CERT-In would not be applicable to self-hosted VPNs. The main emphasis is on security and which is why we would be using the most secure protocol currently available which is OpenVPN. This protocol being virtually unbreakable and open source can be modified according to the requirements and also makes use of SSL/TLS tunneling making it highly

secure in setting up a connection. It uses AES-256-bit encryption and 2048-bit RSA authentication and a 160-bit SHA1 [21] hashing algorithm.

The users can set up a VPS using Linux, purchase a VPS, and use the OpenVPN protocol as the underlying mechanism for creating a VPN. The users can then connect to the Internet using this VPN connection to safeguard their data. We will also add another layer of security which would be Time-Based One Time Password (TOTP) for authentication [22]. This would assure that the users are authenticated well before using the VPN services (Fig. 3).

If the users wish to connect to the Internet to access specific geolocated services or content, they can purchase a VPS (multiple available online) and set up a VPN to access that content.

Fig. 3 Flowchart for configuring customized VPN



3 Conclusion

There are many operational and market implications that hinder the idea of implementing new directives laid down by CERT-In. Reporting within 6 hours will impair the efficiency of flow management [23]. An adequately structured laid-down risk-based approach should be followed to improve the approach to collection and management of data logs for VPS, VPN, and cloud providers keeping the operational costs and risk appetite of the business in regard. Therefore, achieving a cyber-secured nation is a constant effort, it is essential to being a positive game plan by balancing cybersecurity with the right to privacy, market implications, and security concerns. The proposed solution aims to alter the PDP bill with GDPR directives that give the right to the end user that can be edited/deleted/stored/manipulated [24]. The alternate approach is to configure a VPS to set up your own VPN server that can be configured with no data logs, data privacy, and multi-factor authentication.

4 Future Aspects

The reporting of cyberincidents from the security perspective is complex, and the new directives are vague in terms of data privacy stated by IFF (Internet Freedom Foundation). The new rules are exempted from central agencies and rigorous clauses around storing data logs within the permitted area of the country could lead to some major VPN and cloud providers diverting advantageous future investments in India. On August 3, 2022, the center withdrew the Personal Data Protection Bill 2019 after the Joint Committee of Parliament recommended 81 changes in the proposed law. A clearer picture of the Personal Data Protection (PDP) bill is required considering the Personal Identifiable Information (PII) [25] and Protected Health Information (PHI) for ease in portability, and other important information. For laying down an important bill like this, the cyberwatchdog, CERT-In, should refer to the industry standards and requirements for stating important rules and regulations for data processing.

References

1. Vyas R (2022) New VPN rules and how it alters your privacy. <https://economictimes.indiatimes.com/tech/technology/ettech-explainer-what-indias-new-vpn-rules-mean-for-your-privacy/>
2. Ferguson P, Hutson G (1998) Everything about VPN. Research paper revised in Cisco systems
3. Vojinovic I (2022) VPN Statistics for 2022—"keeping your browsing habits private" <https://dataprot.net/statistics/vpn-statistics/>
4. Anand A (2022) The increasing rate of cybercrime in India. <https://www.cnbctv18.com/india/cyber-crime-are-on-a-rise-in-india-amit-shah-cyber-security-ncrb-data-13913912.htm>
5. Shim T (2022) The many use-cases of VPN: how a VPN can be useful <https://www.webhostinggsecretrevealed.net/blog/security/how-a-vpn-can-be-useful/>

6. Tim Mocan (2018) How does a VPN server work. <https://www.cactusvpn.com/beginners-guide-to-vpn/what-is-a-vpn-server-how-does-a-vpn-server-work/>
7. Harry S (2020) All about PPTP, via GeeksForGeeks <https://www.geeksforgeeks.org/pptp-full-form/>
8. Vojinovic I (2022) What Is the L2TP VPN Protocol?": An outdated protocol we still use out of convenience. <https://dataprot.net/guides/what-is-l2tp/>
9. Dahan M (2021) How to use a VPN with secure socket tunneling protocol (SSTP). <https://www.comparitech.com/blog/vpn-privacy/vpn-sstp/>
10. Josh (2020) What is OpenVPN? how it works & when to use It in 2022. <https://www.allthingssecured.com/vpn/faq/what-is-openvpn/>
11. Urban T, Tatang D, Degeling M, Holz T (2019) Study on subject data access in online advertising after GDPR. International workshop on data privacy management, luxembourg
12. Ministry of Electronics and Information Technology (MeitY), Government of India (2022) CERT-In_Directions_70B. https://cert-in.org.in/PDF/CERT-In_Directions_70B_28.04.2022.pdf
13. Sharwood S (2022) ExpressVPN moves servers out of India to escape customer data retention law. https://www.theregister.com/2022/06/02/expressvpnserver_out_of_india/
14. Max S (2021) VPN server location: physical servers vs digital servers. <https://www.cloudwards.net/virtual-server-vs-physical-server/>
15. Vishwanath A (2021) The laws for surveillance in India, and concerns over privacy. <https://indianexpress.com/article/explained/project-pegasus-the-laws-for-surveillance-in-india-and-the-concerns-over-privacy-7417714/>
16. Internet Freedom Federation (2022) CERT-in directives are vague <https://twitter.com/internetfreedom/status/1521797466496004097>
17. Sen P (2021) EU GDPR and Indian data protection bill: a comparative study. Indian Institute of Technology (IIT), Kharagpur. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3834112
18. Pal Dalmia V (2011) Data protection laws In India. <https://www.mondaq.com/india/it-and-internet/133160/data-protection-laws-in-india#:~:text=Under%20Section%2043A%20of%20the,any%20person%2C%20then%20such%20body>
19. Veera Vanamali K (2022) Why does India doesn't have the PDP bill yet. https://www.business-standard.com/podcast/economy-policy/why-does-india-not-have-a-data-protection-bill-yet-122080500071_1.html
20. Perkins S (2022) Self Hosted VPNs. <https://www.androidpolice.com/how-to-make-personal-vpn-30-minutes/>
21. OpenVPN community (2022) OpenVPN cryptographic layer. <https://openvpn.net/community-resources/openvpn-cryptographic-layer/>
22. OpenVPN community (2022) TOTP multi factor authentication. <https://openvpn.net/vpn-server-resources/google-authenticator-multi-factor-authentication>
23. Thathoo C (2022) Cyber security incidents within 6 hours To CERT-In. <https://inc42.com/buzz/report-cyber-security-incidents-within-6-hours-to-cert-in-govt/#:~:text=%E2%80%9CAny%20service%20provider%2C%20intermediary%2C,%2C%E2%80%9D%20CERT%2DIn%20said>
24. Baig A (2108) What GDPR has for the VPN users. <https://www.globalsign.com/en/blog/what-gdpr-means-for-vpn-providers-and-users>
25. Burchfiel A (2022) India's personal data protection bill impact businesses. <https://www.tokenex.com/blog/ab-how-will-indias-personal-data-protection-bill-impact-businesses>

Addressing DIO Suppression Attack in RPL based IoT Networks



Rajat Kumar, Jyoti Grover, Girish Sharma, and Abhishek Verma

Abstract The Internet of Things (IoT) has brought a revolution in technology in the last decade. IoT is susceptible to numerous internal routing attacks because of the characteristics of the sensors used in IoT networks and the insecure nature of the Internet. The majority of the IoT ecosystem's problems come during the routing phase. While routing, the attacking node causes a number of challenges with the packet transmission mechanism. Routing Protocol for Low-Power and Lossy Networks (RPL) is susceptible to numerous types of attacks. The effects could be disruptive to network performance and resource availability. In this paper, we investigate the impact of a novel attack known as the DIO suppression attack and propose a mitigation mechanism for this attack on RPL-based network. This attack disrupts the topology of a network, and as a result, certain number of nodes are disconnected. Attacker nodes exploit the trickle algorithm to execute this attack. The impact of DIO suppression attack in different topologies and scenarios is studied in this research. We have also proposed a lightweight mitigation technique to defend the networks from this attack. This technique leverages the trickling timer's DIO Redundancy Constant k for each node to identify the attacking node in the network.

Keywords IoT · RPL · DIO suppression attack · Security · Routing attack

R. Kumar · J. Grover (✉) · G. Sharma
Malaviya National Institute of Technology Jaipur, Jaipur 302017, India
e-mail: jgrover.cse@mnit.ac.in

R. Kumar
e-mail: 2020rcp9012@mnit.ac.in

G. Sharma
Manipal University Jaipur, Jaipur 303007, India

A. Verma
Babasaheb Bhimrao Ambedkar University, Lucknow 226025, Uttar Pradesh, India
e-mail: abhiverma@iiitdmj.ac.in

1 Introduction

The term Internet of Things (IoT) refers to a network of interconnected devices that are built with sensors, software, and other technologies to transmit and receive data to and from other devices. IoT is used in a variety of industries, each with its own set of security concerns, including health care, smart homes, and autonomous cars. IoT devices are susceptible to various security attacks because of their resource restrictions if they are connected to one another via lossy communication networks.

The majority of Internet of Things applications are made possible by the widespread use of IPv6 over Low-Power Wireless Personal Area Networks (6LoWPAN) [15], a form of Low-Power and Lossy Networks (LLNs). RPL was created by the IETF [12] Routing Over Low-power and Lossy Networks working group (ROLL) that provides routing functionality in LLN. Traditional routing protocols are inappropriate for LLN due to its features [9]. A network may be vulnerable to different routing problems from both internal and external attackers due to insufficient security. RPL is vulnerable to various security attacks that can affect the security and privacy of its users because of its self-organization, self-healing, openness, and resource-constrained nature. Majority of security solutions concentrate on applying cryptographic techniques to secure the RPL control messages. However, if the encryption keys have already been compromised, cryptographic methods cannot defend the network from inside attackers [5]. By utilizing the hacked nodes, internal attackers have the ability to forcefully reduce the network performance and control communication.

In this paper, the DIO suppression attack and its effects on RPL-based networks are examined, and a mitigating method is proposed. This attack disrupts the topology of the network by exploiting trickling algorithm. Therefore, certain number of sensor nodes get disconnected in the network. DIO suppression attacks have the potential to drastically reduce the Average End-to-End Delay (AE2ED), Average Power Consumption, and Packet Delivery Ratio (PDR) of RPL-based networks.

In a DIO Suppression Attack, a malicious node broadcasts DIO messages to legitimate nodes. If the attacker node sends same DIO packet consistently [7], legitimate receiver nodes start suppressing their own DIO transmission which is governed by trickle algorithm [4]. Because DIO packets are used to identify neighbors and network topology, their suppression may result in network partition and some routes may remain undiscovered. The contributions of this paper are listed below:

- On RPL-based IoT networks, a comprehensive analysis on the impact of the DIO suppression attack in various topologies and circumstances is conducted.
- A lightweight mitigation technique to address DIO suppression attack is presented. This method leverages the trickling timer's DIO redundancy constant k for each node to identify the attacking node in the network.

The remaining paper is organized as follows. Section 2 presents the working of RPL protocol. In Sect. 3, we describe related work. The DIO suppression attack is presented in Sect. 4. A detailed discussion of the experimental evaluation of the DIO suppression attack is presented in Sect. 6. A lightweight solution to address DIO suppression attack is discussed in Sect. 7. The paper is concluded in Sect. 8.

2 Background

This section presents the overview of RPL protocol and DODAG construction.

2.1 Introduction to RPL Protocol

RPL protocol can provide routing capability in LLNs, where devices are severely resource constrained. Network devices are arranged into Directed Acyclic Graphs (DAGs) by distance-vector routing technique. A network that has no round-trip routes between any two nodes is referred to as a DAG. Then, traffic is directed toward one or even more DODAG root. The DODAGs, which are directed acyclic networks only with single root node and sink all data, are contained within the DAG. One or more DODAGs may be present in each of the numerous RPL instances that coexist inside a DAG, enabling several applications to run concurrently and independently over the network.

Internet control management protocol version 6 (ICMPv6) is the foundation for RPL control messages [2]. Following control messages are used by RPL in DODAG construction, DODAG Information Object (DIO), DODAG Information Solicitation (DIS), Destination Advertisement Object (DAO), Destination Advertisement Object Acknowledgment (DAO-ACK), and Consistency Check (CC).

2.2 DODAG Construction and Working

Exchanges of DIO messages are used to construct DODAGs, and this process always begins at the root node. The majority of the DIO base fields, including DODAG version, DODAG ID, RPL instance ID, and RPL mode of operation, is set by the root node. When a DIO message is received, each node determines its rank with the help of specified Objective Function. Figure 1 represents different steps of DODAG construction.

When parents are selected based on rank, routing loops are avoided. DIO messages are always exchanged frequently to maintain the routing topology, and nodes may choose to discard a new DIO message if it does not cause any changes in the current DODAG scenario at the receiving node (such as a change in the DODAG version number) or a change in the node's preferred parent. RPL employs the Trickle algorithm to limit the quantity of DIO messages in order to preserve the limited resources of nodes. Each node maintains a DIO counter and a timer with a threshold value. When the trickling game concludes or when a DIO message that updates the RPL configuration is received, a DIO message is dispatched.

The DIO packet count will now be raised each time a DIO packet is transmitted and ignored. The count and trickle timer are both reset and the trickling time is

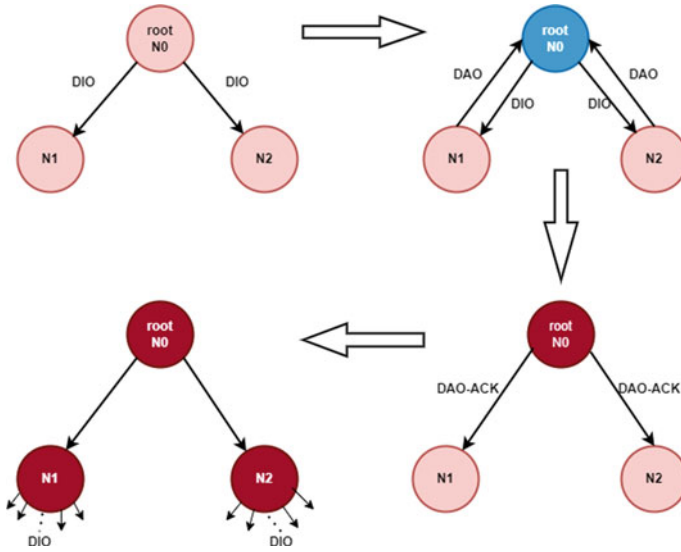


Fig. 1 DODAG construction

doubled if the counter hits the threshold number. Additionally, when a modification is made as a result of a DIO message received, the DIO count and trickle timer will reset to their initial values. Whenever the network is stable, this approach enables fewer DIO message broadcasts and enables rapid topology updates when there are changes.

3 Related Work

The authors [7] suggested two mitigation mechanisms for the DIO suppression attack. The first is inserting the message integrity code into the messages specified by the RPL specification. The second is to implement MAC-layer encryption. The latter approach uses more computing power and increases network traffic overhead.

Yavuz et al. [16] presented a deep learning-based IDS against version number and hello flood attacks on RPL-based IoT networks. The authors suggested five hidden layers in a neural network. They used the Cooja simulator to simulate networks with 10 to 1,000 nodes. Their experimental findings for version number attack and hello flood attack revealed the precision and recall of 94% and 97%, respectively. The published study did not, however, include the false-positive rates.

Mayzaud et al. [6] described the hybrid placement IDS for version attack. Numerous “monitoring sensors” are strategically placed throughout the network to monitor DIO messages. The IDS underwent assessment by the writers. Their trial’s findings showed high detection rates. Additionally, it was shown that false-positive detection might be decreased.

Sedjelmaci et al. [10] describe the distributed placement IDS. Their theory is that signature-detection can detect frequent attacks while anomaly-detection is only done when malicious traffic is identified. The authors combined their methodology with a scheme to limit the number of false positives. The IDS was subject to a sinkhole attack test. The evaluation produced comparable results to SVELTE while consuming less energy.

The Parent Failover and Rank Authentication techniques, covered in [14], protect against the Sinkhole attack. The first method uses a one-way hash that is created and added to DIO messages to enable genuine nodes to determine whether another node on the route to the sink is inadvertently reporting a rank. In the latter, a sink node tells a child node that it is not delivering enough traffic (based on a predetermined baseline).

The authors [1] examined several variables, along with the periodicity of DIO packets, packet delivery ratio, and packet loss, to examine the impact of black-hole attacks. Additionally, they proposed a protection system based on a per-node scheme based on the forwarding habits of network neighbors.

In [3], the authors offered a variety of wormhole attack detection methods. Giving the nodes and, by extension, the neighborhood geographic information is one strategy. Another choice is to use various link layer secret keys for every network segment, which prevents communication between two nodes in different parts. It is more challenging to use a Merkel tree authentication schema to build the topology.

In [13], it has been proposed to counteract the selective forwarding attack by establishing alternate paths inside the RPL topologies that are dynamically selected by nodes.

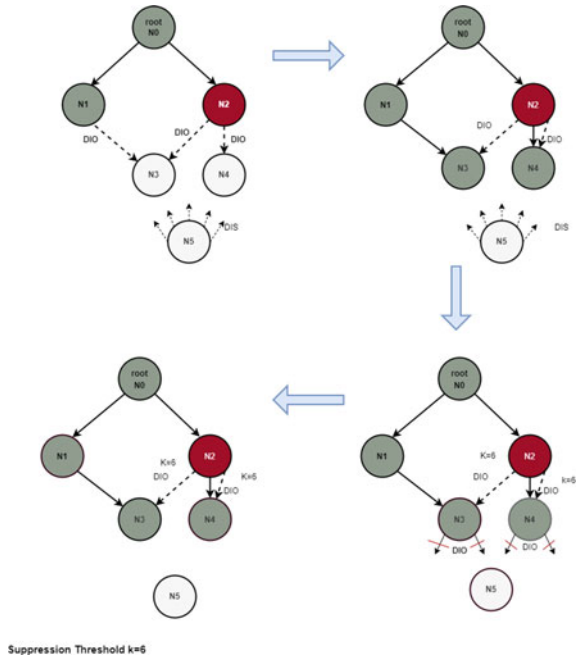
Reference [8] claims that almost no special defence against the Hello Flood attack has been developed. The simulation shows how the RPL Global and Local repairing mechanisms will quickly deal with the attacker.

4 DIO Suppression Attack Overview

A malicious node broadcasts DIO messages to legitimate nodes in a DIO suppression attack. If the attacker node sends the same DIO packet repeatedly, the recipient nodes consider it consistent[7]. If they receive consistent DIOs, nodes will suppress their own DIO transmission, which is governed by the trickle algorithm[4]. Because DIO packets are used to identify neighbors and network architecture, their suppression may result in some nodes remaining hidden and some routes remaining undiscovered. Attacks on DIO suppression harm the performance of IoT network protocols like the RPL protocol.

The transmitter in RPL broadcasts the DIO while DODAG is being created. Once the receiver has received the DIO from the transmitter, it adjusts its sibling list and parent list rank and transmits a DAO packet with route information. A malicious node will repeatedly send DIO messages to legitimate nodes after receiving it. Honest nodes will stop transmitting DIOs when they get a DIO packet from a malicious node.

Fig. 2 Working of DIO suppression attack during DODAG formation



As a result of continuous suppression, some nodes might continue to be hidden, and some routes might continue to be undiscovered.

In Fig. 2, N0(root), N1, N2, N3, N4, and N5 nodes are available to construct a DODAG to transmit the data between the nodes.

N0 initiates the construction of the DODAG by broadcasting the DIO message to the nearest nodes. N1 and N2 receive the DIO messages from N0. N1 and N2 acknowledge the DIO message with the DAO control messages, and N0 sends back another DAO-ACK message as an acknowledgement. Now, N1 and N2 are connected to node N0. N1 and N2 transmit the DIO messages to join the other nodes in the network. N2 is a malicious node in this network. N2 then sends the DIO message to the nodes that want to join the network. But this attacking node is programmed to send the same DIO message every time. N2 sends a DIO message to N3 and N4. We have set the DIO redundancy constant (threshold) to 6. So N3 and N4 will get the same DIO message. If N3 and N4 receive the six consistent DIO messages then these nodes will not transmit the DIO message in future.

5 Experimental Setup

This section discusses the impact analysis of a DIO suppression attack based on simulation. Using the NetSim simulator, a number of sets of experiments were conducted to examine the impact of a DIO suppression attack on an RPL-based network [11],

Table 1 Parameters for simulation model

Parameter	Value
Simulator	NetSim
Topology	Grid, Random
Number of nodes	5, 10, 15, 20, 25, 30
Number of nodes in grid	16
Number of malicious nodes	10%, 20%, 30% of legitimate nodes
Routing protocol	RPL protocol
Area	500 m * 500 m
Simulation time	100 s
Transmission range	50 m
Interference range	100 m
Data packet size	127 Bytes
Mobility	Random mobility

which is the most reliable and widely used network simulator. Table 1 presents simulation parameters considered in various experiments.

Two network topologies are used to simulate this attack: (1) grid topology and (2) random topology. For grid topology, we took 16 nodes and compared them with the 16-node random topology. In other scenarios, we took 5, 10, 15, 20, 25, 30 nodes and varying numbers of malicious nodes, i.e., 10%, 20%, 30% malicious nodes. All these simulations are done for static and mobile nodes and compared with each other, which is discussed in the Results and Analysis section of this paper.

6 Results and Analysis

The findings and analysis of the simulation is presented in this section. The attack's impact is evaluated using three parameters: throughput, average battery consumption, and delay.

Throughput—The amount of data moved successfully from one place to another in a given time period, and it is measured in kilo bits per second (kbps).

Delay—It represents the typical time required for all packets to travel from the source application to the target application layer.

Average Battery Consumption—It displays the average battery consumption over the whole network of connected nodes.

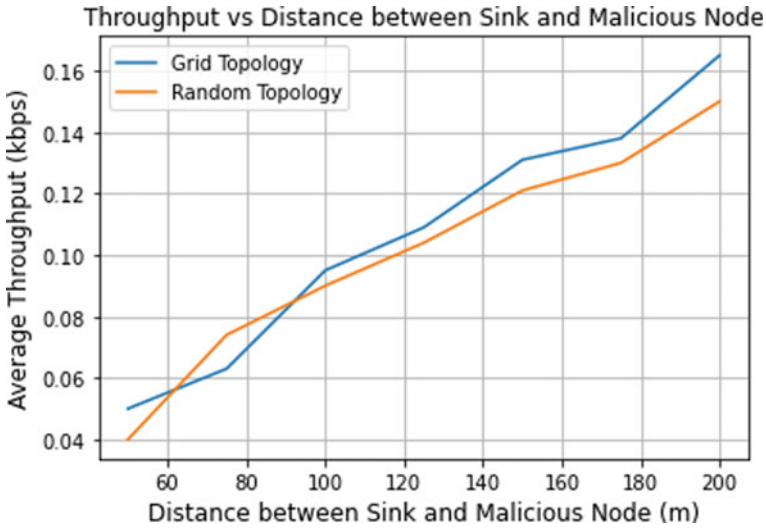


Fig. 3 Comparison of throughput and distance between sink node and malicious node in 16-node grid and random topology

6.1 Impact of Distance Between Malicious and Sink Node on Throughput

In the DIO suppression attack, some nodes remain disconnected because of the suppression of DIO control messages, which are responsible for the construction of DODAG in routing. Figure 3 shows that throughput decreases if the distance between the sink and malicious node decreases, i.e., if the attacker node is near the sink node, then this attack is more fatal.

Throughput drops in the random topology. Random topology increases the probability of disconnection caused by an attacker node, which causes more packet losses in the network and a reduction in performance. Figure 4 shows that throughput is decreased exponentially if nodes are mobile.

6.2 Impact of Varied Malicious Nodes in Different Topologies on Throughput

Our analysis demonstrate that the throughput decreases as the malicious nodes in the topology increase. Figure 5a illustrates the effects of malicious nodes in a static scenario with percentages of 10%, 20%, and 30% in networks of 5, 10, 15, 20, 25, and 30 nodes.

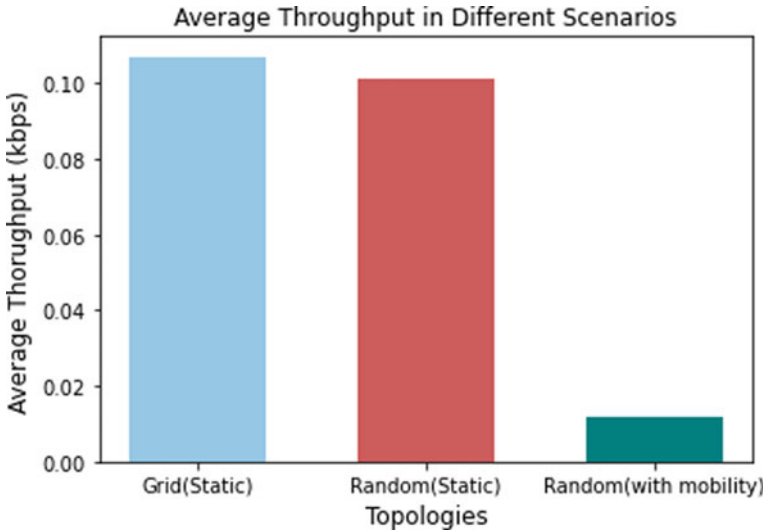
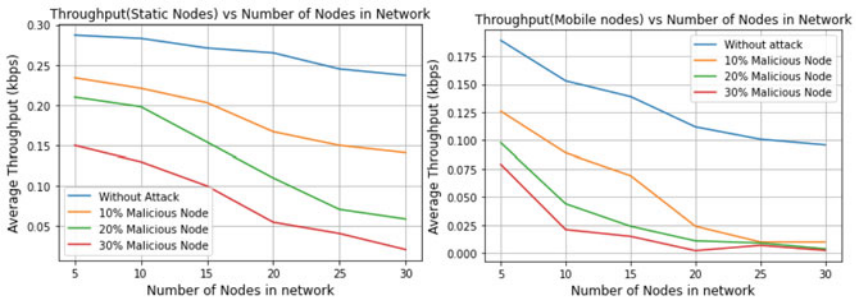


Fig. 4 Average throughput in different scenarios and topology of 16 nodes



(a) Throughput in Static Scenario vs Number of Nodes in Network

(b) Throughput in mobile scenario vs Number of Nodes in Network

Fig. 5 a Throughput in static scenario versus number of nodes in network, b Throughput in mobile scenario versus number of nodes in network

If all nodes are mobile, the DIO suppression attack would become more severe. If all nodes were mobile with an increased number of malicious nodes, throughput would further be significantly reduced as can be seen in Fig. 5b.

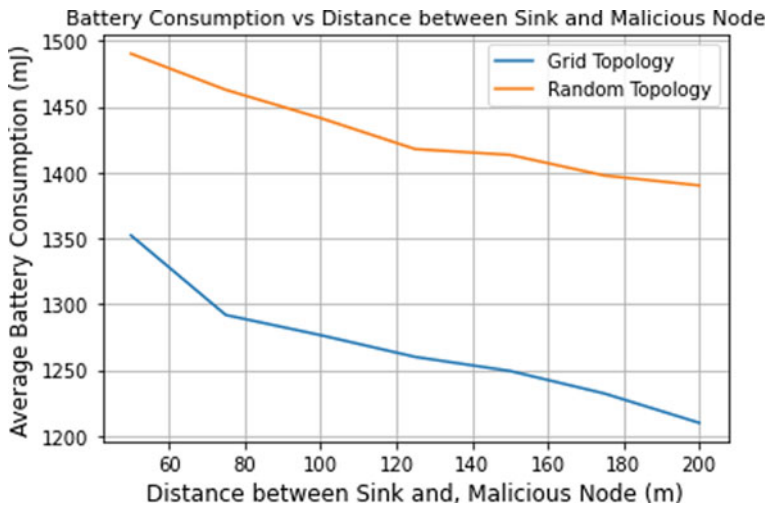


Fig. 6 Battery consumption versus distance between the malicious node and sink node graph

6.3 Impact of Distance Between Malicious and Sink Node on Battery Consumption

This section analyzes the battery usage of different topologies and scenarios. As malicious node moves away from the sink node, less average battery power is used as can be seen in Fig. 6. The average battery consumption is higher if the malicious node is close to the sink node. Figure 7 shows the battery consumption between grid and random topology. The average battery consumption in random topology is greater than in grid topology.

If nodes are mobile, then the battery consumption is highest because DODAG construction is more frequent in a mobile scenario that needs more processing power, so battery consumption is increased.

6.4 Impact of Varied Malicious Nodes in Different Topologies on Battery Consumption

Battery consumption increases if the number of total and malicious nodes increases in the network. From Fig. 8a, we can analyze the attack's impact on battery consumption. The battery consumption will increase if the number of nodes in the network increases. If we change the malicious nodes from 10% to 20%, battery consumption increases.

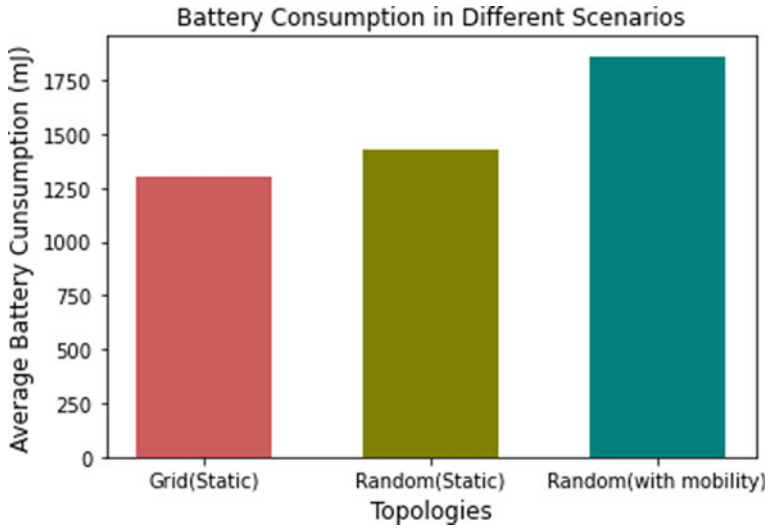
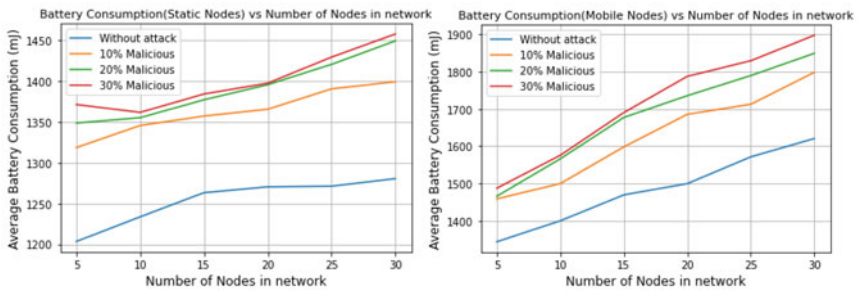


Fig. 7 Average battery consumption in different scenarios and topologies



(a) Battery Consumption in different topologies with varied malicious nodes

(b) Battery Consumption in different topologies with varied malicious nodes

Fig. 8 a, b Battery consumption in different topologies with varied malicious nodes

6.5 Impact of DIO Suppression Attack on Delay in RPL-Based Networks

In this section, we will compare the delay for different scenarios and topologies in the network. Delay will increase if the number of attacking nodes increases in the network. DIO suppression attack disconnects the nodes which result in the increase of delay. Figure 9a shows that in the static scenario for 10%, 20%, 30% malicious nodes in the network of 5, 10, 15, 20, 25, 30 nodes, the delay is increasing significantly. Increasing delays affect the communication between the nodes during the route, which gives less throughput.

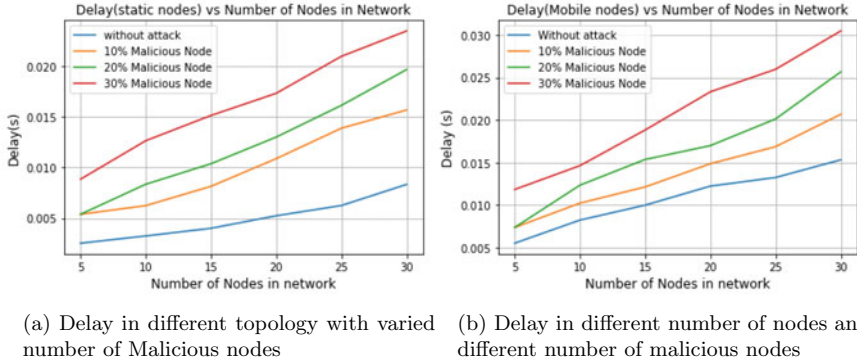


Fig. 9 **a** Delay in different topologies with varied number of malicious nodes, **b** Delay in different number of nodes and different number of malicious nodes

If nodes are mobile, then this attack shows more fatal results. If nodes are moving, then a malicious node can get in touch with the greater number of nodes in the network. It affects the topology of the network, and more nodes remain disconnected, which results in higher delays in the network. As we can see in Fig. 9b for the mobile scenario, in the network of 5, 10, 15, 20, and 30 nodes with 10%, 20%, 30% of malicious nodes, the delay increases significantly. The delay also increases if the number of nodes increases.

7 Mitigation Mechanism for DIO Suppression Attack

As we have seen in the result and analysis section, this attack is becoming more fatal if the number of malicious nodes increases or if malicious nodes get closer to the sink node. If nodes are mobile, then the negative impact of this attack will increase exponentially. In this section, we will propose a mitigation mechanism for this attack. This mitigation mechanism is a frequency-based solution to mitigate and detect this attack.

This solution works on every node during the transmission of the data packets. In the first step of this mechanism, we set a DIO_MAX, which is less than the threshold value to suppress the DIO message for any node. The second step is started when a node transmits a DIO message. There is a trickle timer which is working for every node. The trickle algorithm divides the time into variable trickle intervals. If a node receives consistent messages equal to the threshold value, then the trickle algorithm suppresses the transmission of the DIO messages from that node.

Algorithm 1 Mitigation Algorithm for DIO Suppression Attack

```

SET DIO_MAX ((STEP1)
DIO_MAX=DIORedundancyConstant(k)-1

DIO TRANSMIT ((STEP2)
for each trickle interval do
    DIO.Counter=0

DIO RECEIVE (STEP3)
if DIO.Counter < DIO_MAX then
    process the DIO Message
    if Consistent DIO Message then
        DIO.Counter++
    else
        DIO.Counter=0
else
    Discard the DIO

```

In the second step, for each trickle timer, we will set our DIO_Counter for every trickle interval; for every trickle interval, it will count the number of consistent messages. In the third and essential step, it will check if the DIO_Counter is less than the DIO_MAX then it will process the DIO packet. After processing, if that packet is consistent, it will increment the DIO_Counter; otherwise, it will make DIO_Counter zero because of the received inconsistent packet. If DIO_Counter becomes equal to the DIO_MAX, it will discard the next receiving packet because the next packet may be a consistent message which can suppress the DIO of the child node. Figure 10 shows the working of the mitigation mechanism for the proposed algorithm for RPL-based IoT networks.

8 Conclusion and Future Work

In this paper, we have presented the analysis of DIO suppression attack in different scenarios of a RPL-based IOT network. This attack is more severe if the attacker is present near the sink node. Also, if the number of malicious nodes increases in the network, this attack becomes more fatal. In this attack, the victim node suppresses its DIO messages because it gets consistent messages from the malicious node that equals the threshold value. We have also analyzed this attack on mobile networks, and the results are more fatal on mobile networks. We have also proposed a counter-measure for this attack which may reduce the risk of the legitimate node becoming a victim node of this attack.

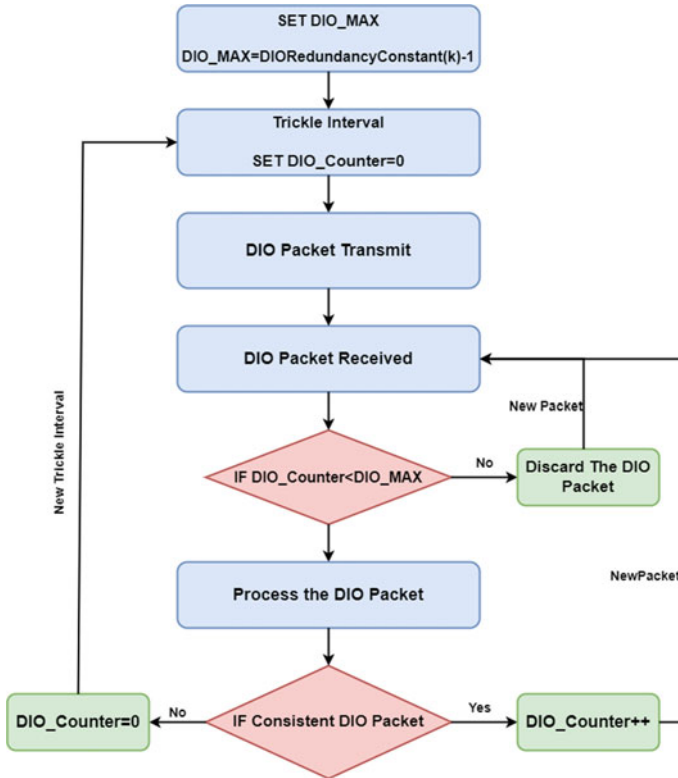


Fig. 10 Working of mitigation mechanism for DIO suppression attack

References

1. Airehrour D, Gutierrez J, Ray SK (2016) Securing RPL routing protocol from blackhole attacks using a trust-based mechanism. In: 2016 26th International telecommunication networks and applications conference (ITNAC). IEEE, pp 115–120
2. Cao Y, Muqing W (2018) A novel RPL algorithm based on chaotic genetic algorithm. Sensors 18 (10):3647
3. Khan FI, Shon T, Lee T, Kim K (2013) Wormhole attack prevention mechanism for RPL based LLN network. In: 2013 Fifth international conference on ubiquitous and future networks (ICUFN). IEEE, pp 149–154
4. Levis P, Clausen TH, Gnawali O, Hui J, Ko J (2011) The trickle algorithm. RFC 6206
5. Miloslavskaya N, Tolstoy A (2019) Internet of Things: information security challenges and solutions. Clust Comput 22(1):103–119
6. Mitra D, Gupta S (2021) Data security in IoT using trust management technique. In: 2021 2nd International conference on computational methods in science & technology (ICCMST) (Los Alamitos, CA, USA, Dec 2021). IEEE Computer Society, pp 14–19
7. Perazzo P, Vallati C, Anastasi G, Dini G (2017) DIO suppression attack against routing in the internet of things. IEEE Commun Lett 21(11):2524–2527
8. Pongle P, Chavan G (2015) A survey: attacks on RPL and flowpan in IoT. In: 2015 International conference on pervasive computing (ICPC). IEEE, pp 1–6

9. Raoof AM (2021) Secure routing and forwarding in RPL-based internet of things: challenges and solutions. PhD thesis, Carleton University
10. Sedjelmaci H, Senouci SM, Taleb T (2017) An accurate security game for low-resource IoT devices. *IEEE Trans Veh Technol* 66(10):9381–9393
11. Tetcos. Tetcos: Netsim—network simulation software, India
12. Vasseur J (2014) Terms used in routing for low-power and lossy networks. RFC 7102
13. Wallgren L, Raza S, Voigt T (2013) Routing attacks and countermeasures in the RPL-based internet of things. *Int J Distrib Sens Netw* 9(8):794326
14. Weekly K, Pister K (2012) Evaluating sinkhole defense techniques in RPL networks. In: 2012 20th IEEE International conference on network protocols (ICNP). IEEE, pp 1–6
15. Winter T, Thubert P, Brandt A, Hui JW, Kelsey R, Levis P, Pister K, Struik R, Vasseur JP, Alexander RK et al (2012) RPL: IPv6 routing protocol for low-power and lossy networks. RFC 6550:1–157
16. Yavuz FY, Ünal D, Gül E (2018) Deep learning for detection of routing attacks in the internet of things. *Int J Comput Intell Syst* 12:39–58

Modelling Identity-Based Authentication and Key Exchange Protocol Using the Tamarin Prover



Srijanee Mookherji, Vanga Odelu, Rajendra Prasath,
Alavalapati Goutham Reddy, and Basker Palaniswamy

Abstract In real-time applications, authentication plays a vital role in enabling secure communications. The authentication protocols need to be formally verified under a defined threat model. Unless the protocols are verified for the intended security, the purpose of employing such protocols may eventually fail. There are multiple ways to formally verify the security of the authentication protocols including the use of automatic verification tools like the Tamarin Prover. The Tamarin Prover tool supports equational theories along with built-in functions. However, this tool does not support some mathematical operations such as elliptic curve point addition. It is necessary to have point addition in Identity-Based Encryption (IBE)-based authentication protocols. Chen–Kudla modelled the point addition operation in the Tamarin Prover using a technique based on concatenation. However, this technique is not applicable to all identity-based protocols including IBE-based authentication protocols. In this paper, we present a modelling technique known as normalised precomputation for point addition using a hash function. We analyse the security of a simple identity-based encryption-based key exchange protocol under extended Canetti and Krawczyk’s (eCK) adversary model. Our analysis shows that the proposed technique is secure and retains the properties of point addition. Therefore, the technique can be applied to different IBE-based authentication protocols where point addition operation is necessary.

S. Mookherji (✉) · V. Odelu · R. Prasath
Computer Science and Engineering Group, Indian Institute of Information Technology Sri City,
Chittoor, 630 Gnan Marg, Sri City 517646, Andhra Pradesh, India
e-mail: srijanee.mookherji@iiits.in

V. Odelu
e-mail: odelu.vanga@iiits.in

R. Prasath
e-mail: rajendra.prasath@iiits.in

A. G. Reddy
Department of Mathematics and Computer Science, Fontbonne University, St. Louis, MO 63105,
USA

B. Palaniswamy
VIT-AP University, Amravati 522237, Andhra Pradesh, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
S. J. Patel et al. (eds.), *Information Security, Privacy and Digital Forensics*,
Lecture Notes in Electrical Engineering 1075,
https://doi.org/10.1007/978-981-99-5091-1_9

Keywords The Tamarin Prover · Authentication protocol · Key exchange · eCK-Adversary model · Elliptic curve point addition

1 Introduction

A secure communication is a problem that the cryptography research community has been working on for a long time. Authentication and Key Exchange (AKE) is an essential part of a secure communication. The problem dealt with in an authenticated communication is that of an adversary \mathcal{A} which has the power to modify, delete, delay and introduce false messages or impersonate a participant in the communication. Key exchange in an authenticated communication allows two parties to generate a shared secret. Authentication protocols use various key exchange techniques like Diffie–Hellman Key Exchange (DHKE) protocol [5] to establish a secret session key [6, 17]. When it comes to a multi-server environment, such authentication protocols may have a major limitations such as the clients may need to store public keys of every single server [31]. To overcome the limitations, identity-based key exchange protocols were introduced [4, 15, 22]. The idea has been applied to design many key exchange protocols [3, 18]. In an identity-based cryptosystem, user identities are used as public keys. A trusted third-party generates a private key for the user using the user identity and a master key. The public key is the user identity, thus users do not need to store multiple public keys.

The extended Canetti–Krawczyk (eCK) [10] adversary model is a widely accepted adversary model. It is used to verify the various required security properties for AKE protocols. A protocol is considered as secure, if an adversary \mathcal{A} , who is in control of communication between two parties, is unable to distinguish session key from a random value. It can do so, only if it calls certain queries that reveal various secret information that are part of the protocol communication. In the eCK adversary model, the adversary is able to call *Ephemeral Key Reveal Query*, *Long-Term Key Reveal Query* and *Session Key Reveal Query*. The *Ephemeral Key Reveal Query* allows an \mathcal{A} to capture all the session-specific temporary secret information. The *Long-Term Key Reveal Query* reveals the long-term secret keys of a party to the adversary and *The Session Key Reveal Query* reveals the current session key between two parties. However, the adversary is allowed to call the queries one at a time.

The Tamarin Prover is an automatic formal security analysis tools which supports features like Diffie–Hellman, hashing, bilinear pairing and so on. The shortfall of the tool is that it does not support elliptic curve point addition [21]. The developers provide a modelling example for the Chen–Kudla protocol [3] where they use ordered concatenation in place of point addition. However, the same approach cannot be implemented for all AKE protocols using point addition operations. We introduce a generalised ID-based Authentication and Key Exchange (ID-AKE) protocol in this paper that uses point addition operations. We model the same in the Tamarin Prover tool using a different modelling technique and analyse it under the eCK adversary model.

In the upcoming sections, we define the required mathematical preliminaries in Sect. 2. Next, we describe the literature review on modelling AKE protocols using the Tamarin Prover in Sect. 3. We discuss the problem of replacing point addition operation with an ordered concatenation in Sect. 4. The contributions of the paper are presented in Sect. 5. In Sect. 6, the summary of a generalised ID-AKE protocol and its Tamarin Prover model is given. We demonstrate that the proposed modelling technique ensures that ID-AKE protocol is secure under the eCK adversary model and it retains the properties of point addition. Finally, Sect. 8 concludes the paper.

2 Mathematical Background

In this section, we discuss the required mathematical preliminaries used to design the ID-AKE protocol.

2.1 Bilinear Pairings

Bilinear pairings can be defined by assuming that G_1 is an additive cyclic group of prime order q , G_2 is a multiplicative cyclic group of prime order q . Let, P be the generator of G_1 , the bilinear pairing equation $e : G_1 \times G_1 \rightarrow G_2$ satisfies the following properties [29]:

- Bilinearity: $e(aP, bQ) = e(P, Q)^{ab}$, for all $P, Q \in G_1$ and for all $a, b \in \mathbb{Z}_q^*$.
- Computability: For all $P, Q \in G_1$, $e(P, Q)$ can be efficiently computed.
- Non-degeneracy: There exists $P, Q \in G_1$ with $e(P, Q) \neq 1$, where 1 is the multiplicative identity of G_2 .

2.2 Hash Function

A one-way hash function is a function $h : \{0, 1\}^* \rightarrow \{0, 1\}^n$ satisfying the following conditions [13, 16]:

- The input $x \in \{0, 1\}^*$ is of arbitrary length binary string and the output $h(x) \in \{0, 1\}^n$ is a binary string of fixed length with n bits.
- **One-wayness:** Given a $y = h(x) \in \{0, 1\}^n$, it is hard to compute x in $\{0, 1\}^*$.
- **Collision-Resistant:** Given $x \in \{0, 1\}^*$, finding $y \in \{0, 1\}^*$ where $x \neq y$ such that $h(x) = h(y)$ is infeasible to compute.

2.3 Message Authentication Code

Let $M \in \{0, 1\}^*$ be a message of variable length, $K \in \mathcal{K}$, where \mathcal{K} is the key space, be a secret key shared between two parties. We define a message authentication code, say, $MAC : \mathcal{K} \times \{0, 1\}^* \rightarrow \{0, 1\}^n$, and the function $C = MAC(K, M)$, where $C \in \{0, 1\}^n$ is a fixed length binary string. The MAC satisfies the following properties [24]:

- Without key K , it is hard to verify the message authentication code M .
- For a given C , it is hard to compute the M and K due to one-wayness of the MAC .

3 Related Works

Many AKE protocols have been designed till now using the concepts of the DHKE protocol [1, 7]. Shamir [22] introduced the identity-based key exchange protocol to overcome the problem of storing multiple public keys. Using the same concept, many ID-AKE protocols were proposed [3, 9, 14, 30]. The identity of the parties was used as their public keys. The DHKE protocol is a simple and essential protocol that is still being used widely to design AKE protocols. To study the modelling techniques used in the Tamarin Prover, studying the DHKE protocol model is of utmost importance. The DHKE protocol model is coded in [12] and its vulnerability is tested against the Man-In-The-Middle (MITM) attack under the eCK adversary model. In the MITM attack, an adversary \mathcal{A} is able to impersonate one or both the participants without their knowledge and finally establish a shared secret key with the participants [8].

In various Tamarin Prover documentations [11, 20, 21], the authors have described about a few AKE protocols like Joux protocol [9], Chen–Kudla protocol [3], RYY protocol [19], etc. that use point addition. The protocols are modelled using different modelling techniques and the codes are present in the ‘The Tamarin Prover Github Repository’[25]. For formalising the Joux protocol, they used the multiset of identities of the participants. A study of the technique is presented in Sect. 3.2. Next, as stated in [21], Chen–Kudla KA protocol is modelled using an ordered concatenation instead of point addition. The Chen–Kudla protocol’s modelling technique is thoroughly explained in Sect. 3.3. An exhaustive study of three protocols [3, 5, 9] is presented in order to look through the different modelling techniques incorporated to formalise them. Also, the paper focusses on their potential of being used in a generalised case. A comparative study of the protocols is presented in Table 1. The study summarises the protocols based on the cryptographic primitives used, the protocol specifications, the modelling techniques and various security properties used for verification.

```

lemma MITM:
  "All #i1 skey. (Session_created(skey)@i1)
  ==> (not (Ex #i2. K(skey)@i2))"

```

Fig. 1 A simple MITM lemma

3.1 Diffie–Hellman Key Exchange Protocol

The DHKE protocol is reviewed and the Tamarin Prover model is presented in this section. We consider two parties: *Alice* and *Bob*. Alice and Bob choose a secret $a \in (1 \leq a \leq p - 1)$, $b \in (1 \leq b \leq p - 1)$, respectively, over a finite field $GF(p)$ with prime p . $A = g^a \bmod p$ and $B = g^b \bmod p$ are computed by Alice and Bob, respectively. A is sent to Bob by Alice and B is sent to Alice by Bob. They compute the shared session key $SessK = SessKA = SesskB = g^{ab} \bmod p$.

Man-In-The-Middle Attack in the DHKE Protocol: The DHKE protocol is vulnerable to the MITM attack [8]. Let us assume that an adversary \mathcal{A} intercepts the message $M1 = \langle A \rangle$ from Alice and replaces it with $X = g^x$ finally sending $M2 = \langle X \rangle$ to Bob. Bob sends $M3 = \langle B \rangle$ which is again intercepted by the \mathcal{A} and passed onto Alice without any changes. At the end, \mathcal{A} establishes a session with Bob using $SessK = g^{xb}$, and therefore impersonating Alice. The DHKE protocol is modelled and tested using the Tamarin Prover in [12]. The model is tested for MITM vulnerability using the simple Lemma 1 which is shown and elaborated in Fig. 1.

Lemma 1 *For all cases, session keys that are created at an instance $i1$, the adversary, \mathcal{K} must not be able to compute a session key at an instance $i2$.*

The tool produced an analysis stating that the protocol is vulnerable to the MITM attack. The Tamarin Prover tool traces for the possibility of the MITM attack and is able to find a counterexample where an adversary, \mathcal{K} , is able to compute the session key, thus turning the lemma red. Here, the adversary sends a fake g^a . Therefore, we can conclude that an adversary is able to perform an MITM attack.

3.2 The Joux Protocol Using Signatures (SIGJOUX)

In this section, we review the three-party authentication protocol proposed by Joux [9], which is a variation of the Diffie–Hellman protocol. This uses bilinear pairing.

- Three parties: *Alice*, *Bob* and *Carol* participate in one round tripartite DHKE.
- Each of them select random values a , b and c , respectively. They also choose long-term keys ka, kb, kc , respectively, $\exists 1 \leq (a, b, c) \leq p - 1$ over the finite field $GF(p)$ with prime p . Finally, they compute $A = aP$, $B = bP$ and $C = cP$. Here, $e : G_1 \times G_1 \rightarrow G_2 \ni e(xP, yP) = e(P, P)_{xy}$.

- Alice, Bob and Carol simultaneously sign a message with their chosen long-term keys and send the following to two other parties : $SigA : sign_{ka}(ID_B, ID_C, A)$, $SigB : sign_{kb}(ID_A, ID_C, B)$ and $SigC : sign_{kc}(ID_A, ID_B, C)$.
- On receiving the same, each party is able to compute their own shared secret keys, $SessKA: h(e(B, C)^a, ID_A, ID_B, ID_C)$, $SessKB: h(e(A, C)^b, ID_A, ID_B, ID_C)$ and $SessKC: h(e(A, B)^c, ID_A, ID_B, ID_C)$.
- Finally, the shared secret is $SessK = SessKA = SessKB = SessKC = e(P, P)^{abc}$ where $e(P, P)^{abc} \in G_2$.

The entire code *Joux.spthy* is present in the Tamarin Prover Github repository [27]. The ephemeral key, ekA , is denoted with $\sim ekA$ which denotes that it is a fresh value and the rule *Proto1* will generate a fresh ekA for every session. The $!Ltk()$ ensures that the fact will remain constant at all times. While modelling the Session Key generation we can see that, for each party, the IDs are denoted by $\$$ which means that they are public. For each party, the IDs of the other two are added using the multiset operator ‘+’ as per the multiset rewriting rules for formalising a protocol.

The Tamarin Prover modelling formalises the protocol by using multiset rewriting rules. Alice, A , chooses her ephemeral key ekA . The other two parties are Bob, B , and Carol, C . The signing key $ltkA$ is used to sign his own public identity $\$A$, the multiset of the public identities of the two parties $\$B, \C along with the public ephemeral key $[ekA]P$. $Pstate(x, A, B + C)$, which is the protocol state fact, denotes that a session is executed. In the second rule, A checks the signatures of the other two parties, extracts their XB and XC which are the public ephemeral keys of B and C , respectively, and computes the shared key as $e(XB, XC)^{ekA}$. The protocol succeeds in providing Perfect Forward Secrecy with *Long-Term Key Reveal* model if A accepts the session key generated with B and C . However, it fails to provide the same if there is an *Ephemeral Key Reveal* modelled for the protocol.

3.3 Chen–Kudla Key Agreement Protocol

We study the Chen–Kudla Key Agreement Protocol in this section. It is an ID-based key exchange protocol that uses the concepts of bilinear pairing and point addition. A Key Generation Centre (KGC) is there that is responsible for the registration of users \mathcal{U} . The key exchange protocol is a two-party communication.

- In the KGC setup phase, the KGC randomly selects a secret key, Key , which acts as the long-term key and computes public key $Pub = KeyP \ni P \in G_1$. Here, P is a generator of G_1 and $Key \in Z_q^*$.
- In the key exchange phase, Alice (A) and Bob (B) are considered as two users. KGC computes $H_A = h(ID_A)$, $S_A = KeyH_A$ for Alice, $H_B = h(ID_B)$, $S_B = KeyH_B$ for Bob. KGC sends S_A and S_B to Alice and Bob, respectively. Here $H_A, H_B \in G_1$. $h()$ is the hash function $\ni \{0, 1\}^* \rightarrow G_1$.

- A computes $A = aP$ and B computes $B = bP$, where a and b are randomly selected ephemeral secrets. A is sent to Bob by Alice. B is sent to Alice by Bob.
- Alice then generates $SessKA = e(S_A, B)e(aH_B, Pub)$ and Bob generates $SessKB = e(S_B, A)e(bH_A, Pub)$ which results in the computation of $SessKey = SessKA = SessKB = e(bH_A + aH_B, Pub)$.

The Chen–Kudla protocol is modelled in the Tamarin Prover by replacing the point addition operation with an ordered concatenation. The complete code *Chen_Kudla.spthy* is available in the Tamarin Prover Github Repository [26]. The shared secret key $sessK = e(ex[hp(\$B)] + ey[hp(\$A)], P_s)$ is written as $(e(hp(\$B), mpk)^{ex} (e(sk_A, Y)))$ using the concepts of bilinearity [2] that states that $e(P + Q, Y) = e(PY)e(QY)$. The protocol model works aptly when the adversary \mathcal{A} is restricted from revealing the ephemeral key of the test session and its significant matching session. This is true even if no *Long-Term Key Reveal Query* is called by \mathcal{A} . On removing the *Ephemeral Key Reveal Query* restriction, the protocol fails to provide key secrecy.

Comparative Study: A comparative study of the protocols is presented in Table 1. The protocols [3, 5, 9] and the generalised ID-AKE protocol are compared with respect to the cryptographic primitives used to design the protocols, the Tamarin Prover modelling technique used and the various security properties achieved by the protocols.

Table 1 Comparative study of modelling protocols in the Tamarin Prover

Features	Comparative study of modelling protocols in the Tamarin Prover			
Protocol	DHKE [5]	Joux [9, 20]	Chen–Kudla KE [3]	Proposed ID-AKE
Cryptographic primitives	Finite field [GF(p)]	Bilinear pairing	Bilinear pairing, ECC point addition	Bilinear pairing
Protocol specifications	Not applicable	ID based, signature	ID based	ID based
Modelling technique	Simple	ID of other two parties as multiset (+ operator)	Bilinear terms concatenated instead of point addition operator	Pre-computed keys
MITM	Not secured	Secured	Secured	Secured
Perfect forward secrecy	Not applicable	Secured	Not secured	Secured
Ephemeral key secrecy	Not applicable	Not secured	Not secured	Secured

4 Problem with Ordered Concatenation in ID-AKE

The Tamarin Prover does not provide the provision of performing point addition. Also, it does not support computation of equalities such as $(c)[(a)P + (b)]P = [(ca)P + (cb)P]$ [21]. Here, for example, the Tamarin Prover model for Chen–Kudla Key Agreement Protocol (*Chen_Kudla.spthy*) present in the repository [26], bilinear terms having point addition are replaced with an ordered concatenation [21] as discussed in Sect. 3.3. There are many ID-AKE protocols that are designed using the point addition operation [14, 23, 28]. The same approach cannot be used in such cases where point addition is used to secure the master key of a Trusted Key Distribution Centre (TKGC). The point addition operation is used to generate the authentication message of the participants in the communication by using the public key of the TKGC. Using the concept of concatenation would not help in achieving security of the master key. This is because, for performing the concatenation operation, the master key needs to be used directly.

We present a generalised ID-based authentication and key exchange (ID-AKE) protocol and model it using the Tamarin Prover in this paper. The detailed description is presented in Sect. 6. The protocol uses the concept of bilinear pairing and point addition. Subsequently, to model the generalised ID-AKE protocol, we embrace the technique of normalisation and define a unary public function $hf/1$ that works similarly to a hash function in Tamarin Prover. Along with this, some precomputations need to be performed in order to ensure the KGC’s secret key security. The technique is illustrated in detail in Sect. 6.2.

5 Contributions of the Paper

The research contributions of the paper are as follows:

- We present a comparative study of the Tamarin Prover modelling techniques used to model authentication protocols that use point addition operations.
- We discuss a generalised ID-AKE protocol that uses point addition operation and present a technique using normalisation and precomputation to model the same.
- Under the eCK adversary model, we test the security properties using the proposed technique. The result shows that the proposed technique is able to achieve the properties of point addition without compromising security of the original protocol.

6 A Generalised ID-AKE Protocol

We provide a summary of the generalised ID-AKE protocol in this section. We discuss about the modelling strategy used and the normalisation and precomputations needed to successfully model the protocol.

Table 2 ID-AKE—registration phase

KGC	
Chooses a Secret Key, Key	
Computes Public Key, $Pub = Key \cdot P$	
KGC	User
Compute: $K_U = \frac{1}{h(ID_U) + Key} P$ $\langle ID_U, K_U \rangle$	

6.1 Authentication and Key Exchange Protocol

We begin with the Key Generation Centre setup phase as shown in Table 2. The Key Generation Centre (KGC) chooses a master private key, Key and generates public key $Pub = Key \cdot P$.

A user requests for registration in the user registration phase (Table 2). The KGC computes $K_U = \frac{1}{h(ID_U) + Key} P$ and sends $\langle ID_U, K_U \rangle$, which the user keeps safe. In the proposed protocol, we assume that two users Alice, A and Bob, B register with the KGC. The KGC sends $\langle ID_A, K_A \rangle$ to Alice and $\langle ID_B, K_B \rangle$ to Bob. Here, $K_A = \frac{1}{h(ID_A) + Key} P$ and $K_B = \frac{1}{h(ID_B) + Key} P$.

In the authentication and key exchange phase (Table 3), Alice chooses secret a and computes $A = a(h(ID_B)P + Pub)$. Alice then sends $\langle M1 = A \rangle$ to Bob. Similarly, Bob chooses secret b , computes $B = b(h(ID_A)P + Pub)$ and $SessKB = e(A, K_B)^b$. Bob then sends $\langle M2 = MAC(SessKB, A, B) \rangle$ to Alice. Thus, Alice authenticates Bob.

Alice further computes $SessKA = e(B, K_A)^a$ and sends $\langle M3 = MAC(SessKA, B, A) \rangle$ back to Bob. Hence, authenticating herself to Bob and establishing a secret session key $SessK = SessKA = SessKB = e(P, P)^{ab}$.

6.2 Modelling ID-AKE Using the Tamarin Prover

In this section, we describe about the normalisation and precomputations that are required in our modelling technique. We discuss the Tamarin Prover model and verify the protocol security under the eCK adversary model using the Tamarin Prover.

Normalisation and precomputation: In the designed ID-AKE, to model the point addition operation, normalisation needs to be performed. The public key for the participants needs to be pre-computed by the KGC. The importance of point addition in this protocol is that the operation is used to construct the ID-based public key without revealing the private key of the KGC. This is achieved by point adding the public key of the KGC. The point addition operation provides the required hardness to secure the private key.

For the normalisation of the initial point addition operation, we introduce a public unary function $hf/1$ that is against multiple inputs the function provides a single

Table 3 Alice and Bob authentication and key exchange phase

Alice	Bob
Choose secret a $A = a(h(ID_B)P + Pub)$ $\xrightarrow{M_1=(ID_A, A)}$	Choose secret b $B = b(h(ID_A)P + Pub)$ Compute : $SessKB = e(A, K_B)^b$ $Auth = MAC(SessKB, ID_A, ID_B, A, B)$ $\xleftarrow{M_2=(Auth, B)}$
Compute : $SessKA = e(B, K_A)^a$ $Conf = MAC(SessKA, B, A, ID_B, ID_A)$ Check: $Auth \stackrel{?}{=} Conf$ $\xrightarrow{M_3=(Conf)}$	Check : $Conf \stackrel{?}{=} Auth$
Shared Secret: $SessK = SessKA = SessKB = e(P, P)^{ab}$	

output. We use inv denoting field inverse and $pmult$ denoting point multiplication. The normalisation in the protocol is performed as follows:

- For $K_A = \frac{1}{h(ID_A)+Key}P$ the point addition part is normalised as $TempKa = hf(ID_A, Key)$. Next, K_A is computed as $K_A = pmult(inv(hf(ID_A, Key)), 'P')$.
- For $K_B = \frac{1}{h(ID_B)+Key}P$ the point addition part is normalised as $TempKb = hf(ID_B, Key)$. Next, K_B is computed as $K_B = pmult(inv(hf(ID_B, Key)), 'P')$.

In the AKE phase, $A = a(h(ID_B)P + Q)$ and $B = b(h(ID_A)P + Q)$ are the ephemeral public key that needs to be computed at Alice and Bob's end, respectively. In order to use the normalisations $TempKa$ and $TempKb$ for computation of A and B , Alice and Bob need to have the knowledge of 'Key' which is the long-term key of KGC. It is highly undesirable from protocol security point of view. Thus, we pre-compute the values $ap = pmult(TempKa, 'P')$ and $bp = pmult(TempKb, 'P')$ at the KGC's end and send it to Alice and Bob as public keys. With ap Bob computes $B = pmult(b, ap)$ and with bp Alice computes $A = pmult(a, bp)$ which are the ephemeral public key used for authentication.

The Tamarin Prover Code: The Tamarin Prover model for the generalised ID-AKE is explained below: The program *IDAKE.spthy* starts with the header 'theory *IDbasedAKE*' which is the theory name. The next line has 'begin' which means start of the protocol modelling. The third line calls the builtins that are required for the modelling. The fourth line describes the public functions `hf/1` and `mac/2` used to model the protocol. The code is shown in Fig. 2.

The rule 'TrustedKGC' is depicted in Fig. 3. It is defined to compute the public key and long-term key (ltk) of the KGC (key generation centre). For every session,

```

theory IDbasedAKE
begin
builtins: diffie-hellman, bilinear-pairing, hashing
functions: hf/1, mac/2

```

Fig. 2 ID-AKE—Tamarin Prover model—the Tamarin Prover code header

```

rule TrustedKgc:
let Pub = pmult(~ Key, 'P')
in [Fr(~ Key)]--[]->[!Ltk($TKGC, ~ Key), !PubK($TKGC, Pub)]

```

Fig. 3 ID-AKE—Tamarin Prover model—rule for KGC setup

```

rule AliceReg:
let TempKa = hf(h($IDA), Key)
ap = pmult(TempKa, 'P')
Ka = pmult(inv(TempKa), 'P')
in [!Ltk($TKGC, Key)]--[]->[!PubA($IDA, ap), !Ltk($IDA, Ka)]

rule BobReg:
let TempKb = hf(h($IDB), Key)
bp = pmult(TempKb, 'P')
Kb = pmult(inv(TempKb), 'P')
in [!Ltk($TKGC, Key)]--[]->[!PubB($IDB, bp), !Ltk($IDB, Kb)]

```

Fig. 4 ID-AKE—Tamarin Prover model—rule for user registration

the rule will generate a fresh persistent long-term key $\sim\text{Key}$ as registered with $\text{!Ltk}()$ which acts as the master key. Persistent fact $\text{!PubK}()$ is used to compute the public key.

The code presented in Fig. 4 shows the Rules ‘AliceReg’ and ‘BobReg’ which are used to model the long-term key generation for Alice and Bob using the master key of the KGC. According to the protocol, the key of the user is computed by using the concept of normalisation. Point addition is replaced by the singular public function $\text{hf}/1$ and TempKa and TempKb is computed accordingly. With the normalised value, the long-term key for Alice and Bob is computed and registered using $\text{!Ltk}()$. For the precomputation, the public key is registered using $\text{!PubA}()$ and $\text{!PubB}()$ which contains the normalised value.

Rules ‘Alice’ and ‘Bob’ present the computation of values of A and B . The computations are done using the Ephemeral keys a and b . The code is presented in Fig. 5. A and B (as per the protocol) are computed using the KGC’s public key. Thus, the long-term key of the KGC remains a secret. For modelling the same the normalised pre-computed values ap and bp have been used and ephemeral keys $\sim a$ and $\sim b$ have been registered using $\text{!Ephk}()$.

Rules ‘AliceSession’ and ‘BobSession’ as shown in Fig. 6 are used to generate the session keys sessska and sesskb using the persistent fact $\text{!SessionKey}()$, $\text{MAC}()$,


```

rule Alice:
let A = pmult(~ a, bp)
in [!PubB($IDB, bp), Fr(~ a)]--[]->[Alice(A), !Ephk( ~ a, ~ a )]

rule Bob:
let B = pmult(~ b, ap)
in [!PubA($IDA, ap), Fr(~ b)]--[]->[Bob(B), !Ephk( ~ b, ~ b )]

```

Fig. 5 ID-AKE—Tamarin Prover model—rule for generation of ephemeral public key

```

rule AliceSession:
let bilp = em(B, Ka)
  sesska = bilp ^ ~ a
  macmsgalice = mac(sesska, (<B, A, $IDA>))
in [Alice(~ a), !Ltk($IDA, Ka), In(<B, macmsgbob>), Alice(A)]
--[Session_created_A(sesska), Accept( ~ a, $IDA, $IDB, sesska )
, SessionID( ~ a, <'Alice', $IDA, $IDB, A, B> )
, MatchingSession(~ a, <'Bob', $IDB, $IDA, A, B> )
, Eq(macmsgalice, macmsgbob)]->[!SessionKey(sesska), Out(<A, macmsgalice>)]

rule BobSession:
let bilp = em(A, Kb)
  sesskb = bilp ^ ~ b
  macmsgbob = mac(sesskb, (<$IDA, A, B>))
in [Bob(~ b), !Ltk($IDB, Kb), Bob(B), In(<A, macmsgalice>)]
--[Session_created_B(sesskb), Accept( ~ b, $IDB, $IDA, sesskb )
, SessionID( ~ b, <'Bob', $IDB, $IDA, A, B> )
, MatchingSession( ~ b, <'Alice', $IDA, $IDB, A, B> )
, Eq(macmsgalice, macmsgbob)]->[!SessionKey(sesskb), Out(<B, macmsgbob>)]

```

Fig. 6 ID-AKE—Tamarin Prover model—rule for session key generation

which is used to authenticate each other is also computed. An equality check is done for the MAC() values that are exchanged using the equality restrictions (presented in Fig. 7). A SessionID() and a MatchingSession() is associated for every session created by the above rules. Session_Created() denotes that the rule ran and a session is created and Accept() fact states that the shared session key sesska and sesskb has been accepted [11].

Security Properties: To analyse the modelled protocol under the eCK adversary model, we design Lemmas 10 and 3. The lemma codes are presented in Figs. 8 and 9.

Lemma 2 MITM : *The lemma states that for all sessions created and the adversary, \mathcal{K} , has not called the Long-Term Key Reveal Query or the Session Key Reveal Query, it implies that \mathcal{K} is not able to compute the shared secret session key.*

Lemma 3 Session Key Secrecy: *The lemma states that there does not exist an accepted test session and the adversary, \mathcal{K} , does not have the shared session key. Also, the adversary has not called a Session Reveal Query. If the adversary has found a matching session it implies that the following queries have not been called:*

```

/*Restrictions*/
restriction Equality:
"All x y #i. Eq(x,y) @#i ==> x = y"

/* Key Reveals */
rule ltk_reveal:
[ !Ltk($TKGC, ~ Key) ]--[ LtkReveal($TKGC) ]-> [ Out(~ Key) ]
rule Sessionk_reveal:
[ !SessionKey(skey) ]--[ SesskeyReveal(skey)]-> [ Out(skey)]
rule Ephk_reveal:
[ !Ephk(~ s, ~ ek) ]--[ EphkeyReveal(~ s) ]-> [ Out(~ ek) ]

```

Fig. 7 ID-AKE—Tamarin Prover model—restrictions and key reveal models

```

lemma MITM:
"(All #i1 #i2 skey .
(Session_created_A(skey) @ i1 & Session_created_B(skey)
@i2 & not ( (Ex A #ia . LtkReveal( A ) @ ia )
| (Ex B #ib . SesskeyReveal( B ) @ ib )))
=> not (Ex #i3. K( skey ) @ i3 ))"

```

Fig. 8 ID-AKE—Tamarin Prover model—MITM lemma

```

lemma key_secretcy:
"not (Ex #i1 #i2 s A B k . Accept(s, A, B, k) @ i1 & K( k ) @ i2
& not(Ex #i4. SesskeyReveal(s) @ i4 ) & (All ss #i4 #i5 ms.
(SessionID( ss, ms) @ i4 & MatchingSession(s, ms) @ i5)
=> (not(Ex #i6. SesskeyReveal(ss) @ i6)
& not(Ex #i6 #i7. LtkReveal( A ) @ i6 & EphkeyReveal( s)@i7)
& not(Ex #i6 #i7. LtkReveal( B ) @ i6 & EphkeyReveal( ss)@i7)
& not(Ex #i6 #i7. LtkReveal( A ) @ i6 & LtkReveal( B)@i7)
& not(Ex #i6 #i7. EphkeyReveal( s) @ i6
& EphkeyReveal(ss)@i7))) & ((not(Ex ss #i4 #i5 ms.
SessionID( ss, ms) @ i4 & MatchingSession(s, ms) @ i5))
=> (not(Ex #i6. EphkeyReveal( s) @ i6 )
& not(Ex #i6. LtkReveal( B) @ i6 & i6 < i1 )))"

```

Fig. 9 ID-AKE—Tamarin Prover model—session key secrecy lemma

- *A Session Key Reveal Query for the obtained matching session.*
- *A Long-Term Key Reveal Query for Alice and Ephemeral Key Reveal Query for Alice’s matching session.*
- *A Long-Term Key Reveal Query for Bob and Ephemeral Key Reveal Query for Bob’s session ID.*
- *A Long-Term Key Reveal Query for both Alice and Bob.*
- *An Ephemeral Key Reveal Query for obtained matching session and the parties’ session ID.*

```

Running TAMARIN 1.6.1      Index      Download      Actions »      Options »

Proof scripts

theory IDbasedAKE begin

Message theory

Multiset rewriting rules and restrictions (12)

Raw sources (17 cases, deconstructions complete)

Refined sources (17 cases, deconstructions complete)

lemma MITM:
  all-traces
  "∀ #i1 #i2 skey.
    (((Session_created_A( skey ) @ #i1) ∧
      (Session_created_B( skey ) @ #i2)) ∧
      ¬((∃ A #ia. LtkRev( A ) @ #ia) ∨
        (∃ B #ib. SesskRev( B ) @ #ib)))) ⇒
    (¬(∃ #i3. K( skey ) @ #i3))"
  simplify
  by solve( Alice( ~a ) ▶o #i1 )

lemma eCK_PFS_key_secretcy:
  all-traces
  "∀ #i1 #i2 Test A B k.
    ((Accept( Test, A, B, k ) @ #i1) ∧ (K( k ) @ #i2)) ⇒
    (((∃ #i3. SesskRev( Test ) @ #i3) ∨
      (∃ MatchingSession #i3 #i4 ms.
        ((SessionID( MatchingSession, ms ) @ #i3) ∧
          (MatchingSession( Test, ms ) @ #i4)) ∧
          (∃ #i5. SesskRev( MatchingSession ) @ #i5))) ∨
      (∃ MatchingSession #i3 #i4 ms.
        ((SessionID( MatchingSession, ms ) @ #i3) ∧
          (MatchingSession( Test, ms ) @ #i4)) ∧
          ((∃ #i5 #i6. (LtkRev( A ) @ #i5) ∧ (EphkRev( Test ) @ #i6)) ∨
            (∃ #i5 #i6.
              (LtkRev( B ) @ #i5) ∧ (EphkRev( MatchingSession ) @ #i6))))))
  "

```

Fig. 10 The Tamarin Prover model visualisation for ID-AKE protocol

Finally, if the adversary did not find a matching session, it implies that there does not exist an Ephemeral Key Reveal Query for matching session. Also there does not exist a Long-Term Key Reveal call for Bob, thus stating that Bob has not been compromised.

Model Visualisation: Once Lemmas 2 and 3 are solved in the Tamarin Prover, the colour of the proof turns green as shown in Fig. 10. It is an indication that there were no traces found for any adversary computing the secret session key, k . Thus, we suggest that the designed ID-AKE protocol using the technique of precomputation and normalisation resists MITM attack and provides session key secrecy under the eCK adversary model.

7 Conclusion

In this paper, we study the designing techniques of security models using Tamarin Prover for various authentication protocols. It is observed that the point addition operation modelled in the literature is not applicable to many of the IBE-based protocols. In this work, we present a generalised IBE-based key exchange protocol and modelled it using the proposed normalised precomputation technique with the help of hash function. The Tamarin Prover simulations showed that the proposed technique provides security under the eCK adversary model. In conclusion, the proposed model can be applied to IBE-based protocols where the point addition operation is used.

References

1. Bresson E, Chevassut O, Pointcheval D (2007) Provably secure authenticated group diffie-hellman key exchange. *ACM Trans Inf Syst Secur (TISSEC)* 10(3):10–es
2. Chatterjee S, Sarkar P (2011) Identity-based encryption. Springer Science & Business Media
3. Chen L, Kudla C (2003) Identity based authenticated key agreement protocols from pairings. In: *Proceedings of the 16th IEEE computer security foundations workshop*. IEEE, pp 219–233
4. Das ML, Saxena A, Gulati VP (2004) A dynamic id-based remote user authentication scheme. *IEEE Trans Consum Electron* 50(2):629–631
5. Diffie W, Hellman M (1976) New directions in cryptography. *IEEE Trans Inf Theory* 22(6):644–654
6. Hölbl M, Welzer T (2009) An improved authentication protocol based on one-way hash functions and diffie-hellman key exchange. In: *2009 International conference on availability, reliability and security*. IEEE, pp 894–898
7. Huang LC, Chang TY, Hwang MS (2018) A conference key scheme based on the diffie-hellman key exchange. *Int J Netw Secur* 20(6):1221–1226
8. Johnston AM, Gemell PS (2002) Authenticated key exchange provably secure against the man-in-the-middle attack. *J Cryptol* 15(2):139–148
9. Joux A (2000) A one round protocol for tripartite diffie-hellman. In: *International algorithmic number theory symposium*. Springer, pp 385–393
10. LaMacchia B, Lauter K, Mityagin A (2007) Stronger security of authenticated key exchange. In: *International conference on provable security*. Springer, pp 1–16
11. Meier S, Schmidt B, Cremers C, Basin D (2013) The tamarin prover for the symbolic analysis of security protocols. In: *International conference on computer aided verification*. Springer, pp 696–701
12. Mookherji S, Odelu V, Prasath R (2021) Modelling ibe-based key exchange protocol using tamarin prover. *Cryptology ePrint Archive*
13. Odelu V, Das AK, Goswami A (2015) A secure biometrics-based multi-server authentication protocol using smart cards. *IEEE Trans Inf Forensics Secur* 10(9):1953–1966
14. Odelu V, Das AK, Wazid M, Conti M (2018) Provably secure authenticated key agreement scheme for smart grid. *IEEE Trans Smart Grid* 9(3):1900–1910. <https://doi.org/10.1109/TSG.2016.2602282>
15. Okamoto E, Masumoto H (1990) Id-based authentication system for computer virus detection. *Electron Lett* 26(15):1169–1170
16. Preneel B (1993) Analysis and design of cryptographic hash functions. Ph.D. thesis, Katholieke Universiteit te Leuven

17. Pu Q (2010) An improved two-factor authentication protocol. In: 2010 Second international conference on multimedia and information technology, vol 2, pp 223–226. <https://doi.org/10.1109/MMIT.2010.82>
18. Ruan O, Zhang Y, Zhang M, Zhou J, Harn L (2017) After-the-fact leakage-resilient identity-based authenticated key exchange. *IEEE Syst J* 12(2)
19. Ryu EK, Yoon EJ, Yoo KY (2004) An efficient id-based authenticated key agreement protocol from pairings. In: International conference on research in networking. Springer, pp 1458–1463
20. Schmidt B (2012) Formal analysis of key exchange protocols and physical protocols. Ph.D. thesis, ETH Zurich
21. Schmidt B, Sasse R, Cremers C, Basin D (2014) Automated verification of group key agreement protocols. In: 2014 IEEE Symposium on security and privacy. IEEE, pp 179–194
22. Shamir A (1984) Identity-based cryptosystems and signature schemes. In: Workshop on the theory and application of cryptographic techniques. Springer, pp 47–53
23. Shim KA (2012) A round-optimal three-party id-based authenticated key agreement protocol. *Inf Sci* 186(1):239–248
24. Stallings W (2006) *Cryptography and network security principles and practices*, 4th edn
25. Tamarin prover github repository. <https://github.com/tamarin-prover/tamarin-prover>. Accessed 20 Oct 2022
26. Tamarin prover github repository—chen-kudla protocol. https://github.com/tamarin-prover/tamarin-prover/blob/develop/examples/ake/bilinear/Chen_Kudla.spthy. Accessed 20 Oct 2022
27. Tamarin prover github repository—joux protocol. <https://github.com/tamarin-prover/tamarin-prover/blob/develop/examples/ake/bilinear/Joux.spthy>. Accessed 20 Oct 2022
28. Tsai JL, Lo NW (2015) A privacy-aware authentication scheme for distributed mobile cloud computing services. *IEEE Syst J* 9(3):805–815
29. Tsai JL, Lo NW (2015) Secure anonymous key distribution scheme for smart grid. *IEEE Trans smart grid* 7(2):906–914
30. Tseng YM, Huang SS, Tsai TT, Tseng L (2015) A novel id-based authentication and key exchange protocol resistant to ephemeral-secret-leakage attacks for mobile devices. *Int J Distrib Sens Netw* 11(5):898716
31. Wang C, Xu G, Li W (2018) A secure and anonymous two-factor authentication protocol in multiserver environment. *Secur Commun Netw* 2018

Sensor Fusion and Pontryagin Duality



S. Jayakumar, S. S. Iyengar, and Naveen Kumar Chaudhary

Abstract Boltzmann Machine (BM) and Brooks–Iyengar (BI) algorithm are solving similar problems in sensor fusion. Relationships between these two are established in detail. During 1984, BM was used as a toolset to solve posterior probability finding problems by Hinton (<https://youtu.be/kytxEr0KK7Q>, [10]). During 1996, Brooks–Iyengar algorithm was published (Brooks and Iyengar in Computer, [8]) and it was trying to have robust and yet precise computation of a parameter in a sensor network, where sensor network might include some faulty sensors as well. In this work, it has shown interesting results on BM and BI, when temperature is zero in BM. Dual space of sensor network is used as a space for sensor classification and also to find computability of measurement. Pontryagin duality (Dikranjan and Stoyanov in An elementary approach to Haar integration and Pontryagin duality in locally compact abelian groups 2011 [14]; Woronowicz in QuantumE(2) group and its Pontryagin dual 2000 [15]) is used to construct dual space for a given sensor network. For example, the Fourier transform can be considered as a dual space of the given sensor network. Kolmogorov complexity is used to model measurement problems into a problem of computability of elements in dual space. Sensor fusion problem is formulated as a problem of finding one program “ p ” which results in many strings as output. It appears that there is no known necessary sufficient condition on group (formed by using non-faulty sensors) for “ p ” to exist. And also, it is shown that quantum computing is a natural choice to find such a program “ p ” which produces many strings as output.

S. Jayakumar (✉)
Independent Consultant, Gandhinagar, India
e-mail: jk@jkuse.com
URL: <http://www.jkuse.com/>

S. S. Iyengar
Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, FL 33199, USA
e-mail: iyengar@cis.fiu.edu

N. K. Chaudhary
National Forensic Sciences University, Gandhinagar, India

Keywords Boltzmann machine · Brooks–Iyengar · Pontryagin duality · Quantum computin · Sensor fusion

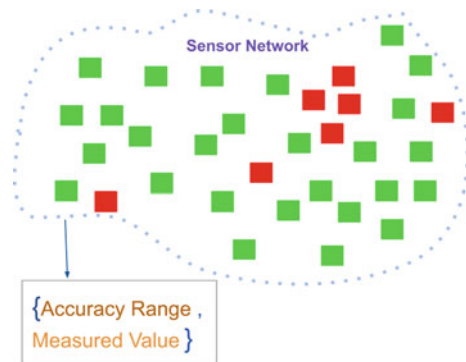
1 Introduction

Data collected in the real world is almost never representative of the entire population (how much collection is necessary and also sufficient!), and so by estimating distribution parameters from an observed sample population, it is possible to gain insight into unseen data. Measurements of any kind, in any experiment, are always subject to uncertainties or errors. Measurement process is, in fact, a process described by an abstract probability distribution whose parameters contain the information about the desired dynamical system. The results of a measurement are then samples from this distribution which allow an estimate of the theoretical parameters. In this view, measurement errors can be seen as sampling errors.

Sensor fusion is emerging as a critical requirement because more automation and associated machines are deployed to manage a given set of workflow. Sensor is designed to measure the value of a given parameter and provide an output for further processing. For example, reading from N sensor data, what are interpretations of N sensor data? Suppose “ N ” sensors are deployed to measure one parameter in a given area and assume that all these sensors are equipped with the ability to communicate with any other sensor in the same network of sensors. In the following Fig. 1, in sensor network, green color squares indicate good sensor and red color squares indicate faulty sensor. All these sensors are connected by using a given communication infrastructure.

Using many sensors for a signal parameter measurement provides the option to have a robust measurement method in place such that subsequent decisions can depend on the measured parameter. N sensors used in measurement systems and “ r ” sensors ($r \leq N$) failed to perform correct measurement. Challenge is to find those

Fig. 1 Red is faulty sensor, green is good sensor



“r” sensors that failed. Most importantly, in real-time finding, those failed sensors become an interesting problem to solve by using edge computing.

A Classical problem in distributed computing is the Byzantine Generals Problem. It attempted to formalize a definition for faulty nodes in a cluster and how for the system to mitigate it. Byzantine generals problem is providing methods to find faulty sensors in distributed computing. BI [2] algorithm had improved “Byzantine generals problem” by using measure over a period of time instead of using point measurement as in Byzantine generals problem. BI algorithm is an effort to find a way to solve the Byzantine generals problem. BI algorithm is using sensor classification to eliminate the faulty sensors. BI algorithm is using a data set from “measurements over an interval: instead of point measurement. In the BI algorithm, mostly, a heuristics method is used to choose good sensors in a given set of sensors which might include faulty sensors as well.

BI algorithm improves precision and accuracy of the interval measurements in sensor network. In fact, in BI algorithm’s above-mentioned performance, there is a presence of faulty sensors in sensor network. Specifically, the BI hybrid algorithm works with noisy data sets to combine Byzantine agreement with sensor fusion. Emerging deep learning segment is shown with Boltzmann machine [6] as a primary tool to perform signal classification. Interestingly, it is shown that BI is a special case of BM when temperature is zero. Computability of measurement has been using statistics-based algorithms before it had moved to modern methods that are part of toolset to handle dynamical system (by using observers in state space). It appears that Boltzmann machine-based observers might perform better compared to statistics-based methods and might come close to observers in dynamical system [1]. Finding such a “Probability distribution function” results in defining observer states and associated dynamical systems. Using the above-mentioned result is an early step in finding error-free sensors and forming a group (G) of sensors which are used in measurement decisions. Disturbance decoupling problem [1] is handled by using observer which can sense the state of dynamical systems.

Boltzmann distribution is using state’s energy and a temperature of a system to provide probability distribution of a sensor node (state). Boltzmann machine is used to get the digital twin of dynamical systems. Sensor network is used to observe dynamical systems in real-time and record samples which will be used in training Boltzmann machines. Sampling methods such as Gibbs, von Neumann entropy, Shannon, and Kolmogorov are used in sampling of data from a given sensor. BM generalizes BI algorithm and Byzantine generals problem to find faulty sensors. For example, BM can handle noise in measurement systems and provide a robust classification of sensors.

Kolmogorov complexity is used to measure entropy of the information source. If information source is a Markov process then Kolmogorov complexity is defined as “normalised by the length of information source”. Kolmogorov complexity converges almost surely (as the length of information source goes to infinity) to the entropy of information source, Shannon information handles random variables. Kolmogorov complexity is handling individual objects (sensors). By using observer, construct state space with topological group structure. Sensor network is formulated as a Group (G)

which is discrete and abelian. Pontryagin duality is useful to perform analysis on G . Result from Pontryagin duality [3] is used to form a problem in measurement and associated computability. Interestingly, it is shown that samples from every “error-free sensor” over time intervals result in measurement that is computable.

2 Mathematical Theory

2.1 Pontryagin Duality

Group with Discrete Topology G is a finite-discrete group with discrete topology. Each element in G is a k -tuple vector, which is a time series data from a given sensor for a given time interval. Let R be real numbers.

Let $f : G \rightarrow R$ and $g_1, g_2 \in G$ such that $g = (a_1, a_2, a_3, \dots, a_k)$ is k -tuple vector.

$$f(g) = \sum_{i=1}^k (a_i)^2 \tag{1}$$

$$f(g_1 \odot g_2) = f(g_1) + f(g_2)$$

$$f(g_1 \odot g_2) = f(g_1) + f(g_2) = f(g_2) + f(g_1) = f(g_2 \odot g_1)$$

G is an abelian group as well.

Where

$$f \in G^* \text{ such that } f : G \rightarrow R \tag{2}$$

$$f_i(g_i) = V_0^i + \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} (g_i(x))^2 dx \tag{3}$$

$$V_0^{ia} = \sum_{j=1}^{i-1} \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} (g_i(x) \times g_j(x)) dx$$

$$V_0^{ib} = \sum_{j=i+1}^N \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} (g_i(x) \times g_j(x)) dx$$

$$V_0^i = V_0^{ia} + V_0^{ib} \tag{4}$$

$$F : G^* \rightarrow R$$

$$F = (f_1, f_2, f_3, \dots, f_p)$$

$$F \in G^{**}$$

$$F = \frac{1}{p} \sum_{i=1}^p f_i(g_i) \tag{5}$$

G^{**} is isomorphic to G (Pontryagin Duality).

Locally Compact and Abelian Group In case G is not finite and discrete, but if G is a Locally compact abelian group, then the following will be used. G is the sample space of state measurements and states have a value in T via a function f .

$f : G \rightarrow T$ where T is S^1 and T is a circle group. G is sample space.

$\tilde{f} : G^* \rightarrow T$, Where G^* is a dual space of G

Pontryagin duality says G^{**} is isomorphic to G . The above definition of f can be extended as in the following:

$f : G \rightarrow T \times T$, where T is a circle (inline figure) in a complex plane. The following circles represent the same. Each element in G will take value in these two circles by using f .



$$\begin{aligned}
 g &= \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \in G \\
 g^2 &= \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \in G \\
 g^r &= \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \text{ (strong accept) } \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \dots \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \in G \\
 g^{r+1} &= \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = g \in G
 \end{aligned} \tag{6}$$

$f_x(g) = g(x)$, where x is 2 D Vector

$$f_x(g) = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in T \times T \tag{7}$$

$$f_x(g^r) = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}^r \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in T \times T \tag{8}$$

$$G = \{g^1, g^2, g^3, \dots, g^r\} \tag{9}$$

G is a finite abelian group.

$f_x(g_1 \odot g_2) = g_1(g_2(x))$ is a group homomorphism (strong accept)

$$f_x \in G^* = \text{Hom}(G, T \times T) \tag{10}$$

$$\begin{aligned} \tilde{f} : G^* &\rightarrow T \times T, \text{ where } G^* \text{ is a dual space of } G \\ \tilde{f}(f_x) &= f(f_x) \\ \tilde{f}(f_{1_x}) \odot f_{2_x} &= f_2(f_{1_x}) \text{ is a group homomorphism} \end{aligned}$$

$$\tilde{f} \in G^{**} \tag{11}$$

where G^{**} is a dual space of G^*

Sensor Network Model In classical case, G^* is a fourier transform. Pontryagin duality is used to construct G^* from given G . Most importantly, G^{**} is isomorphic to G . Computability of measurement is using above defined G^{**} for a given G .

V_0^i in equation (4) is the information received from other sensors. In this case, other $p - 1$ sensors had communicated V_0^i to a sensor node, where $i = 1, 2, 3, \dots, N$. Integral from T_1 to T_2 provides averaged value measured in that time interval ($T_2 - T_1$). Time series data from T_1 to T_2 is a real number.

2.2 Strictly Positive Real Functions of a Sensor

In dynamical systems, a state is a variable which has position and velocity where the mentioned position and velocity can be deterministic or stochastic. In the case of stochastic nature of state, outcomes of each possible measurement on a system result in probability distribution for position and also for velocity.

Interpretation of measurement might happen in different domains, for example, Fourier domain is one such domain. Measurement in interval is necessary for transforming a measured sample from time domain into frequency domain.

$g = (a_1, a_2, a_3, \dots, a_k)$ samples collected from a sensor. Deciding on time interval duration depends on the application and its dynamical system. Slow dynamics or fast dynamics or a mix of both is represented by following dynamical system:

$$g = (a_1, a_2, a_3, \dots, a_k) \text{ is transformed in to } H(z), \text{ where } z, \alpha_i, \beta_i \in C. \tag{12}$$

$$H(z) = \sum_{i=1}^k \frac{\alpha_i}{z + \beta_i} \tag{12}$$

$H(z)$ is a Strictly Positive Real function (SPR) if and only if $H(z)$ is analytic in the real part of $H(z) \geq 0$

SPR is very useful to classify a given sensor:

$$|\beta_1| \leq |\beta_2| \leq |\beta_3| \leq \dots \leq |\beta_i| \tag{13}$$

The above can be used to find an agreement between sensors in terms of frequency response of each sensor data.

2.3 *Byzantine Generals Problem*

The Byzantine Generals Problem (BGP) is a classical problem in distributed computing. BGP is using point data from each sensor node to provide inference on sampled point data from all generals. The BGP model is to formalize a definition for faulty (traitorous) sensors in a cluster and how much system to mitigate it. BI algorithm provides a method to solve BGP.

BI algorithm is using measurements over an interval. Each node communicates with another node to learn the validity of its measurement. Each sensor updates its value to other sensors in a given sensor network. Once again all nodes need to run through interaction till each node converges to a value. This needs to go on till value on each node is converged. At this stage, the given node is qualified to pass its output to other sensors or not.

Each sensor node has measured data in a given interval and then same data set is shared between all sensors of network. Weighted average of the midpoints of all the intervals is computed and same is used to perform sensor fusion.

BI algorithm is more toward deterministic in nature and BI appears to be a good fit in a situation where there is a need to perform sensors without any training infrastructure. Computability of given measurements is an emerging challenge in large-scale sensors used in measuring a particular parameter.

2.4 *Boltzmann Distribution*

Bayesian networks are probabilistic graphical models of directed acyclic graphs, Bayesian sensor is a node and each node is an observable or latent node. Edges represent conditional dependencies of nodes that are connected. Some sensor nodes might not connect with other sensor nodes. In this case, they are conditionally independent of each other. Each node is associated with a probability function.

Input to each node is a set of values from the node's parent variables. Output of nodes is probability distribution.

Probabilistic graphical model is used to model a set of variables and their conditional dependencies via a directed acyclic graph. Boltzmann distribution is a form of a probability distribution. Boltzmann distribution provides the probability of state, where each state is a function of the state's energy and a temperature of a system [13]. In the Boltzmann distribution, joint distribution is not known explicitly or is difficult to sample from directly. Gibbs sampling is useful to get the value of joint distribution in which the conditional distribution of each node is known. It generates an instance from the distribution of each node, by using conditional on the current values of the other nodes. For posterior distribution of a Bayesian network, Gibbs sampling is useful. Bayesian sensor networks using collection of conditional distributions of each node.

$$p_i = \frac{e^{-\frac{E_i}{kT}}}{\sum_{j=1}^N e^{-\frac{E_j}{kT}}} \quad (14)$$

where E_i is the energy, T is the temperature, and k is the Boltzmann constant. In sensor network, term T is associated with noise generation term in measurement process. If T is 0, then measurement is clean (which may not be true often in real world of sensing).

In the Boltzmann machine, each node is connected with other nodes in a given network. Because of this each undirected edge of node represents dependency, where E_i is the probability of a certain state of system p_i and N is the number of sensors in a given sensor network. In the above example, there are three hidden units and four visible units. This is not a restricted Boltzmann machine. Input is fed to all nodes (green and yellow) during interval 0 to T_1 .

Step 1 (input to Boltzmann machine): Supply external input and train network with all systems that are subjected to system temperature T (same for all, different per state is in question).

Step 2 (output from Boltzmann machine): Measure each state that are supporting measurable conditions. The maximum possible energy of the above system will provide provision to quantize energy levels.

Output energy E is a combination of all four energies from each node “ v ”. Each energy level provides a possible state of each node. But in above E , there is no contribution from hidden nodes but there is an indirect contribution which needs to be estimated by using a model of dynamical system [1].

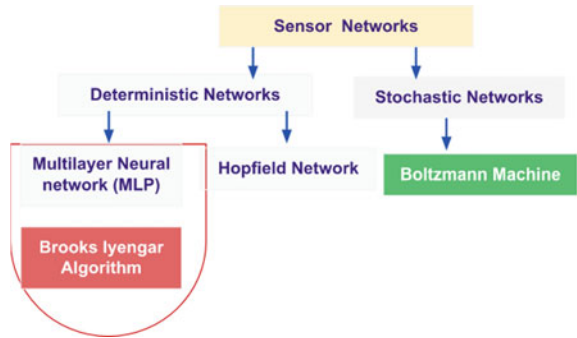
Input vector v is used to predict of hidden value of h in $p(h|v)$. As a next step, use predicted hidden value of h in the prediction of new input value of v by using $p(v|h)$. This process is repeated k times. After k iterations, input vector v_k is recreated from the original input value v_0 . Target states include all possible states of sensor. Each sensor in a network is associated with a finite set of possible states.

Cost of assigning a label to a node sensor is defined as unary cost. Cost of assigning a pair of labels to two different node sensors is defined as pairwise cost. Sensor fusion requires assigning costs to all sensor nodes that are in a distribution.

2.5 Sensor Networks

Sensor networks modeled as deterministic networks or stochastic networks. Info-graphic Fig. 2 shows sensor networks in which both deterministic networks and stochastic networks are shown. In the deterministic networks section, Multilayer Perceptron (MLP) Sensor network and Hopfield networks are included. MLP is used to estimate the conditional average of an input data set. MLP can be viewed as a “fully connected feedforward artificial neural network (ANN)”. Supervised learning is used in training MLP sensor networks. Mentioned training is using backpropagation algorithm which is a generalization of the least mean squares algorithm. Each

Fig. 2 Brooks–Iyengar algorithm in sensor networks



piece of data set is used to train perceptrons by changing connection weights by using backpropagation algorithms.

Infographic Fig. 2 shows that BI Algorithm is part of MLP sensor Network.

The Hopfield network is part of a deterministic recurrent neural network. Lyapunov function is used in the evolution of dynamical states of the Hopfield network. For a given initial condition on each sensor node, there is a well-defined evolution which is following the Lyapunov function. The Hopfield network is used to solve combinatorial problems and also model time series data.

Stochastic sensor networks include Helmholtz and Boltzmann machines. In these stochastic sensor networks, given an input to a node, state of node does not converge to an ensemble distribution (not to one unique state). BM provides a probability distribution of sensor node of network. BM is the stochastic equivalent of the Hopfield Sensor network.

2.6 Reduction of Boltzmann Machine to Hopfield Mode

Boltzmann network energy level is a function of temperature. If temperature is high, then energy also will be high in the Boltzmann network. If $T = 0$ (temperature), then Boltzmann network reaches an energy level which is in equilibrium (energy level need not be zero). In a sense at $T=0$, the Boltzmann network becomes deterministic network. In particular, Boltzmann network becomes the Hopfield network, because Hopfield is having Lyapunov function which can be considered as a constraint (as it comes from energy). In the case of MLP, there is no Lyapunov function and thus no constraint as well.

BI algorithm is closer to MLP because BI algorithm is deterministic and does not have Lyapunov function. Training MLP network [2], backpropagation algorithm is used.

BI algorithm is similar to backpropagation to arrive at convergence in node value. Brooks–Iyengar algorithm performs the following: Every node is given its input

and also average values from other nodes (average over T). Nodes jointly provide deterministic output.

In the above, it is clear that no Lyapunov or temperature is used in BI and thus BI is a special case of Hopfield network where there is no constraint from Lyapunov. BI is another version of MLP where network provides deterministic output by using conditions of other nodes.

3 Sensor Classification

Each sensor is having an energy level at a given time period. Energy level of other sensors also expected to have energy in a similar range. However, it is not expected to have too much of difference in energy level from sensor to sensor.

3.1 Brooks–Iyengar Algorithm in Sensor Classification

Sensor fusion by using BI algorithm [8] is using a Processing Element (PE) to compute accuracy range and also measured value estimation. Let sensor j be used t sec duration to record k samples. And also let sensor j receive measured values from other sensors in a given network. PE of a given sensor j is using

- Recorded k samples (0 to t sec) in sensor j .
- Measured values from other sensors from 1 to N but not sensor j .

Above workflow is part of each PE at a given sensor from 1 to N . Measurement is done for a duration t sec in a given sensor and same measured data is used in PE, where PE is using these samples to compute its “measure value” by using measured values from other sensors.

Assumption is that the measured value from other sensors is a proper time sequence. Keeping time stamp in each measured value is another area of research and that is handled well in IEEE 1588 standard [7].

It is assumed that uniform interval is used to collect each sample in a given sensor. And also, all sensors are synchronized with the clock to start the sample collection process.

BI algorithm is removing sensors with faulty conditions and using only sensors with no error. BI is using heuristic algorithms or variance-based algorithms to classify sensors.

If sensor is providing an “image signal” then BM can handle it with ease and perform sensor classification work. BI appears to be having issues in handling “image” as an input, thus converting image signal as time series data might help BI algorithm. However, BI algorithm extensions to handle image data can use BM and keep temperature equal to zero. $T = 0$ in equation (14) results in $p_i = 1$ for all i . Having

$p_i = 1$ for all sensors for all time is not good and the same results in not a good model of sensor network. Thus, making $T = \epsilon$ small value results in a model which can be used for deterministic algorithms like BI algorithm.

3.2 Boltzmann Machine in Sensor Classification

Maxwell Boltzmann statistics is applicable to identical, distinguishable particle. The molecule of gas is a particle of this type.

- Find a model of a given binary data.
- What is the need to model binary data?
- What can one do with model?
- What can be done with model that emerged from binary data?
- How probability is assigned to each binary vector ?
- How above is connected to the weights of Boltzmann Machine?

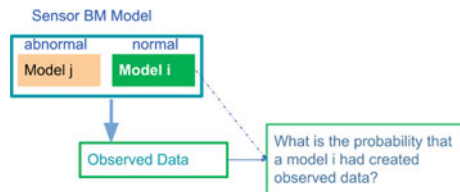
Particles which are regulated by Maxwell–Boltzmann Statistics have to be distinguishable each other. Reconstruction of sensor nodes is different from classification. Reconstruction requires probability distribution of the input values to a given sensor node. Classification of a sensor is associating a sensor (to an input to a given sensor node) with a continuous or discrete value (Fig. 3).

Sensor networks are assumed to be Bayesian networks, while a sensor network is described as a collection of conditional distributions of each sensor node. Gibbs sampling is used to sample sensor networks. Boltzmann machine is used to model a given “set of Binary Vectors”. Fractional belief propagation is used in algorithms for Inference.

Trained Boltzmann machine is deployed to find out distribution of input stream. Sensor classification work is a special case of the above-mentioned process in Boltzmann machines.

Example in [5] is using restricted BM is used as a deep learning network to classify nodes in TCP/IP network.

Fig. 3 Sensor and its model



3.3 Pontryagin Dual Space in Sensor Classification

A locally compact abelian group (sensor network) is useful to model BM. Discrete topology with abelian group (sensor network) is useful to model sensor network with BI algorithm. Equation (3) is used to define dual space (of sensor network) which has elements that are handling BI and BM. Pontryagin dual ([3, 15]) used in obserables. Fourier transform is understood through Pontryagin duality.

Theorem 1 Peter–Weyl’s (Abelian case) [14] *If G is a compact abelian group, then G^* separates the points of G .*

Pontryagin–van Kampen duality theorem says that this functor is an involution, i.e., $G^{**} = G$ for every G which is locally compact abelian.

From above, compact groups to discrete ones and vice versa. It defines a duality between the subcategory G^* of compact abelian groups and the subcategory G of discrete abelian groups. Using the fact that every continuous function on a compact group G^* is almost periodic, this invariant mean gives also a (unique) Haar integral for compact abelian groups.

Definition in Eq. (2) is used to drive $f_1, f_2, \dots, f_k \in G^*$. and $g_1 \in G$. By using Peter–Weyl theorem, it is possible to map functions $f_1, f_2, \dots, f_k \in G^*$ to $g_1 \in G$, i.e., many elements in G^* are mapped to a single element in $g_1 \in G$. Locally compact set in G^* is mapped to a single element in $g \in G$. Observations by using many sensors ($f_1, f_2, \dots, f_k \in G^*$) and result of fusion is a sensor in (or node) G .

In physics and signal processing, the Fourier transform is understood as a transformation of signals from time domain to other domains (frequency). Fourier transform is defined on locally compact groups.

4 Computability of Measurement

4.1 Kolmogorov Complexity for a Given Sensor Reading

Kolmogorov complexity is defined as the length of the shortest binary program from which the object can be effectively reconstructed.

Kolmogorov complexity definition is using probability theory and information theory. In fact, philosophical notions of randomness are also included in definition of Kolmogorov complexity. It appears that Kolmogorov complexity is intimately related to problems in probability theory and also in information theory.

Algorithmic information content of a given object is defined by using Kolmogorov complexity. For example, Kolmogorov complexity is combining definition of computability and statistics to express the complexity of a finite object,

Let “p” is a program, T be a Turing machine and “s” be the output of T.

Let “N” be number of sensors used in sensor network and “r” be the number of faulty sensors.

What is a shortest binary representation of a program from which an information source can be reconstructed by using N-r sensors?

where $K_T(s)$ is used to detect regularities of a given sensor data in order to find new information from a given sensor.

For example, expression “K” is computable if and only if there is an effective procedure that, given any (k-tuple) x of natural numbers, will produce the value $f(x)$.

$$f : N^k \rightarrow R$$

In agreement with this definition, computable functions take finitely many natural numbers as arguments and produce a value which is a single natural number.

4.2 Entropy Computation for a Given Sensor Reading

Construction of dual space of given sensor network G is the first step in getting the “shortest binary representation of a program” (Fig. 4).

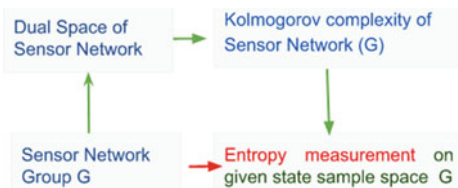
In Sect. 2, there are illustrations that provide steps to construct dual space. Function “f” definition is key in the construction of dual space. But “f” definition needs to have physical relevance to measurement process which is using an N-r good sensor.

To measure same, first Dual space of G is constructed. And Kolmogorov complexity of G is estimated. From Kolmogorov complexity of G, entropy measurement of given “G”. The above method is indirect but uses well-defined Pontryagin duality result and Kolmogorov complexity result.

5 Research Problems

Smart cities are expected to have a large number of sensors in a given network and deployed sensor readings are expected to provide necessary data to an engine that drives decision. For a decision to be correct, it is necessary to include readings from good sensors and not include faulty sensor reading. Thus, there is a need to look for algorithms that can handle sensor data which has measurement noise. In this direction, Boltzmann machine appears to be a natural fit. But the issue is that

Fig. 4 Entropy measurement of sensor



training a Boltzmann machine can be a challenge, if there is no data set available for training Boltzmann machine.

By using the Pontryagin duality of a given group (which is constructed by using a sensor network) of sensors, computability of measurement (in the sense of Kolmogorov complexity) is formulated as Turing machine problem.

5.1 Kolmogorov Complexity of Sensor Data

Measured Samples from “N” sensors are used in the following. There are k samples from each sensor. It is assumed that the first N-r sensors are good and other r sensors are faulty.

$$g_i = \{a_1^i, a_2^i, a_3^i, \dots, a_k^i\} \quad (15)$$

g_i is samples from sensor i and also g_i is mapped to s^i via T, for $i = 1, 2, \dots, N$
Turing machine T is providing the following strings for N-r sensors:

$$K_T(s^i) = \min \{ |p|, T(p) = s^i \} \quad (16)$$

“p” is a program which results in the above given strings via machine T.

Problem 1 What are necessary and sufficient conditions on strings s^1, s^2, \dots, s^{n-r} for the existence of program “p”.

If there exists a program “p” for problem 1, then same “p” computable or not?

If there exists “p” for problem 1 and “p” is computable, then what is method and apparatus to compute such a program “p” ?

One program “p” is proving one string output. As an inverse problem, it is a challenge to show the existence of “p”.

Program “p” resulting in strings s^1, s^2, \dots, s^{n-r} via Machine T. In the normal case, the program “p” is expected to produce one string “s”. But in the above case one program “p” produces “N-r” strings. Observed strings “s” are associated with Group G . Finding such a program “p” is equivalent to performing sensor fusion in a group G . In Quantum computing , one program “p” can result in many observables.. Thus above is a good example and application of quantum computing. Create a quantum computing program “p” which results in many observables.

6 Conclusion

BI Algorithm is shown to be equivalent to BM, when BM is deterministic and also temperature is zero. It appears that the mentioned result might open up research in the selection of initial weight in BM. Maybe the BI algorithm is useful for setting up the

initial weights of the BM, else random weights are used in BM as a starting weight. Problem of measurement is transformed into the problem of finding computability of a function which is an element in dual space of a sensor network. Kolmogorov complexity is used to model measurement problems into a problem of computability of elements in dual space which is constructed by using Pontryagin duality. Most interestingly, it is shown that sensor fusion is a problem in quantum computing in which one program “ p ” generates many output strings.


Acknowledgements Thanks to Shri Jagadeesha Chinagudi for his support during the early days of in-person discussion on Sensor Fusion and BI algorithm. And also thanks to Prof S S Iyengar for all the support during in-person discussion on sensor fusion problem. Thanks to Mistral Solutions, Bilva infra, SunPlus Software, and Apollo Tyres for providing the opportunity to work with their research team.

References

1. Wonham WM (1979) Linear multivariable control, a geometric approach, 2nd edn. Springer-Verlag, New York
2. Ao B, Wang Y, Yu L, Brooks RR, Iyengar SS (2016) On precision bound of distributed fault-tolerant sensor fusion algorithms , Vol. 49, No. 1, 2nd edn, Article 5. ACM Computing Surveys
3. Stefano G, Willam Z (2018) Fourier Transforms from strongly complementary observables
4. Su D (2016) The fourier transforms for locally compact abelian groups
5. Jayakumar S, Kanshi for TCP IP network safety . <https://www.jkuse.com/dltrain/deploy-in-edge/kanshi>
6. Jayakumar S, Boltzmann machine. <https://www.jkuse.com/home/deep-learning/boltzmann-machine>
7. Cho H, Jung J, Cho B, Jin Y, Lee SW, Baek Y (2009) Precision time synchronisation using IEEE 1588 for wireless sensor networks. Department of Computer Engineering, Pusan National University, Busan, Republic of Korea
8. Brooks RR, Iyengar SS (1996) Robust distributed computing and sensing algorithm. Computer
9. Soler-Toscano F, Zenil H, Delahaye JP, Gauvrit N (2014) Calculating Kolmogorov complexity from the output frequency distributions of small turing machines , <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0096223>
10. Hinton G, Lecture by Prof Geoffery Hinton, <https://youtu.be/kytxEr0KK7Q>
11. Lonkar S, Training an MLP from scratch using backpropagation for solving mathematical equations, <https://medium.com/swlh/training-an-mlp-from-scratch-using-backpropagation-for-solving-mathematical-equations-91b523c24748>
12. Jayakumar S (1998) Simultaneous stabilization of feedback systems, PhD Thesis, Systems and Control Engineering, Indian Institute of Technology Bombay
13. Wolf W (2018) A thorough introduction to boltzmann machines , <https://willwolf.io/2018/10/20/thorough-introduction-to-boltzmann-machines/>
14. Dikranjan D, Stoyanov L (2011) A An elementary approach to Haar integration and Pontryagin duality in locally compact abelian groups , Department of Mathematics and Computer Science, Universite of Udine, Via delle Scienze, 208, Loc Rizzi, 33100 Udine, Italy
15. Woronowicz SL (2000) Quantum E(2) group and its Pontryagin dual, Department of Mathematical Methods in Physics, Faculty of Physics, University of Warsaw Hoza 74, 00-682 Warszawa, Polska and ETH Zürich June 16, 2000

Lightweight Malicious Packet Classifier for IoT Networks



Seyedsina Nabavirazavi , S. S. Iyengar, and Naveen Kumar Chaudhary

Abstract Although the Internet of Things (IoT) devices simplify and automate everyday tasks, they also introduce a tremendous amount of security flaws. The current insufficient security measures for smart device protection make IoT devices a potential victim of breaking into a secure infrastructure. This research proposes an on-the-fly intrusion detection system (IDS) that applies machine learning (ML) to detect network-based cyber-attacks on IoT networks. A lightweight ML model is trained on network traffic to defer benign packets from normal ones. The goal is to demonstrate that lightweight machine learning models such as decision trees (in contrast with deep neural networks) are applicable for intrusion detection achieving high accuracy. As this model is lightweight, it could be easily employed in IoT networks to classify packets on-the-fly, after training and evaluation. We compare our lightweight model with a more complex one and demonstrate that it could be as accurate.

Keywords Internet of things · IoT · Networks security · Machine learning · Fault detection · Decision trees · Neural networks

1 Introduction

The popularity of the Internet of Things (IoT) has significantly increased. The forecast for the total number of connected IoT devices in 2025 is 27.1 billion. Today, it seems inevitable having a smart device in our homes. The proliferation of smart devices is not only within the domestic environment but it is also the driving force behind the development of an interconnected knowledge-based world; economy, society, and machinery of government. However, IoT devices come with a tremendous amount of

S. Nabavirazavi (✉) · S. S. Iyengar
Florida International University, Miami, FL 33199, USA
e-mail: sina.nabavi16@gmail.com

N. K. Chaudhary
National Forensic Sciences University, Gandhinagar, Gujarat, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
S. J. Patel et al. (eds.), *Information Security, Privacy and Digital Forensics*,
Lecture Notes in Electrical Engineering 1075,
https://doi.org/10.1007/978-981-99-5091-1_11

139

security risks [3]. Synopsys has a blog post that reveals, back in 2017, a lack of confidence in the security of medical devices with 67% of manufacturers. They believed an attack on a device is probable within a year, and only 17% of manufacturers take steps to prevent them.

The protocols designed for the IoT protocol stack are different from those of the IP stack. IoT designers have to choose between WiFi, Bluetooth, ZigBee, Z-Wave, and LoRaWan as their network protocol. IoT devices are power-constrained and have specific functionality. Using general-purpose protocols for every IoT device would result in more battery consumption and less quality of service. Hence, there is no standard approach for handling security issues in IoT, such as IPSec or SSL for the Internet.

The insufficient security measures and lack of dedicated anomaly detection systems for these heterogeneous networks make them vulnerable to a range of attacks such as data leakage, spoofing, denial of service (DoS/DDoS), etc. These can lead to disastrous effects, damaging the hardware, unavailability of the system, and compromising sensitive data privacy. For example, a deauthentication attack performed on a device with critical significance, such as a steering wheel in a wireless car, can pose a threat to human life. There is a gap between security requirements and the security capabilities of currently available IoT devices.

The traditional IT security ecosystem consists of static network defenses (firewalls, IDS), the ubiquitous use of end-point defenses (e.g., anti-virus), and software patches from vendors. We can't either employ these mechanisms due to the heterogeneity in devices and their use cases. This means that traditional approaches for discovering attack signatures (e.g., honeypots) will be insufficient or non-scalable [13].

Traditional anomaly detection systems are also ineffective within IoT ecosystems since the range of possible normal behaviors of devices is significantly larger and more dynamic than in traditional IT environments. IDSs such as SNORT and Bro only work on traditional IP-only networks as they are static and use signature-based techniques [13]. To address this issue, multiple intrusion detection systems have been introduced in the context of IoT networks. Yet, the majority of them focus on detecting a limited set of attacks, in particular, routing attacks and DoS. However, there are IDSs that introduce dynamicity by employing machine learning to detect malicious behavior. Soltani et al. focused on 3 deep learning algorithms and applied them to the context of intrusion detection [12]. They also introduced a new deep learning-based classifier. Amouri et al. employed supervised machine learning for IDS [1, 5]. Shukla et al. proposed an IDS that uses a combination of machine learning algorithms, such as K-means and decision trees to detect wormhole attacks on 6LoWPAN networks [11]. Cao et al. proposed a machine learning intrusion detection system for industrial control systems [6]. Bangui et al. have employed random forests for IDS in vehicular ad hoc networks. The aim of this paper is similar. Yet, its focal point is to design the machine learning IDS as lightweight and efficient as possible. When the IDS is lightweight and effective, it may be embedded within certain IoT devices. We summarize the contributions of our work below.

- We analyze a public dataset for malicious IoT packet classification.
- We convert the raw traffic of the captured data (PCAP) to a comprehensible format for training. (Lightweight models are unable to classify malicious packets by inspecting raw bytes)
- We develop a Python code to instantiate, train, and test the specified machine-learning models.
- We document and report the evaluation metrics for these models and compare them.

The rest of the paper follows this outline. Section 2, discusses the phase of data collection and how the collected data is comprehended to train the machine learning models. For both training and testing the models, the open-source Aposemat IoT-23 dataset [7] will be used. Later, in Sect. 3, we provide clarification on which machine learning models we use, which features are selected, and how the data is labeled. Section 4 provides the results and the evaluation of these models. In this section, the performance of lightweight models is compared with a more complicated model, a neural network. Section 5 concludes the project's paper and Sect. 6 opens the door for future work.

2 Data Collection

2.1 Dataset

We use the public dataset of **Aposemat IoT-23** for our models [7]. The packets are grouped into different chapters (scenarios). The dataset contains 20 captures of malware traffic in the IoT network and 3 captures of benign traffic. The dataset contains more than 760 million packets and 325 million labeled flows with more than 500h of traffic. The IoT-23 dataset consists of 23 captures overall, called scenarios, of different IoT network traffic. We summarize this information in Figs. 1 and 2.

The malicious scenarios were created executing a specific malware in a Raspberry Pi. In the dataset, the researchers have included traffic from Mirai, Torii, Hide and Seek, and Hajime attacks. The network traffic capture for the benign scenarios was obtained by capturing the network traffic of three different IoT devices: a Philips HUE smart LED lamp, a Somfy Smart Door Lock, and an Amazon Echo home intelligent personal assistant. We should mention that these three IoT devices are actual devices and not simulated. Both malicious and benign scenarios run in a controlled network environment with an unrestrained internet connection like any other real IoT device.

#	Name of Dataset	Duration (hrs)	#Packets	#ZeekFlows	Pcap Size	Name
1	CTU-IoT-Malware-Capture-34-1	24	233,000	23,146	121 MB	Mirai
2	CTU-IoT-Malware-Capture-43-1	1	82,000,000	67,321,810	6 GB	Mirai
3	CTU-IoT-Malware-Capture-44-1	2	1,309,000	238	1.7 GB	Mirai
4	CTU-IoT-Malware-Capture-49-1	8	18,000,000	5,410,562	1.3 GB	Mirai
5	CTU-IoT-Malware-Capture-52-1	24	64,000,000	19,781,379	4.6 GB	Mirai
6	CTU-IoT-Malware-Capture-20-1	24	50,000	3,210	3.9 MB	Torii
7	CTU-IoT-Malware-Capture-21-1	24	50,000	3,287	3.9 MB	Torii
8	CTU-IoT-Malware-Capture-42-1	8	24,000	4,427	2.8 MB	Trojan
9	CTU-IoT-Malware-Capture-60-1	24	271,000,000	3,581,029	21 GB	Gagfyt
10	CTU-IoT-Malware-Capture-17-1	24	109,000,000	54,659,864	7.8 GB	Kenjiro
11	CTU-IoT-Malware-Capture-36-1	24	13,000,000	13,645,107	992 MB	Okiru
12	CTU-IoT-Malware-Capture-33-1	24	54,000,000	54,454,592	3.9 GB	Kenjiro
13	CTU-IoT-Malware-Capture-8-1	24	23,000	10,404	2.1 MB	Hakai
14	CTU-IoT-Malware-Capture-35-1	24	46,000,000	10,447,796	3.6G	Mirai
15	CTU-IoT-Malware-Capture-48-1	24	13,000,000	3,394,347	1.2G	Mirai
16	CTU-IoT-Malware-Capture-39-1	7	73,000,000	73,568,982	5.3GB	IRCBot
17	CTU-IoT-Malware-Capture-7-1	24	11,000,000	11,454,723	897 MB	Linux,Mirai
18	CTU-IoT-Malware-Capture-9-1	24	6,437,000	6,378,294	472 MB	Linux.Hajime
19	CTU-IoT-Malware-Capture-3-1	36	496,000	156,104	56 MB	Muhstik
20	CTU-IoT-Malware-Capture-1-1	112	1,686,000	1,008,749	140 MB	Hide and Seek

Fig. 1 Summary of the malicious IoT scenarios

#	Name of Dataset	Duration(-hrs)	#Packets	#ZeekFlows	Pcap Size	Device
1	CTU-Honeypot-Capture-7-1 (somfy-01)	1.4	8,276	139	2,094 KB	Somfy Door Lock
2	CTU-Honeypot-Capture-4-1	24	21,000	461	4,594 KB	Philips HUE
3	CTU-Honeypot-Capture-5-1	5.4	398,000	1,383	381 MB	Amazon Echo

Fig. 2 Summary of the benign IoT scenarios

2.2 Data Preparation

Our lightweight machine learning model would not be able to classify raw bytes of network traffic. Hence, we convert the raw PCAP files into Packet Description Markup Language (PDML) format. PDML conforms to the XML standard and contains details about the packet layers. We then simply represent the PDML files in Comma-separated Values (CSV) by only selecting our desired features from each PDML packet. We have implemented this in Python.

2.3 Feature Selection

As the feature space is relatively large (Table 4), all packet features may not be relevant. We have manually selected 13 features that appeared to have the highest correlation based on Eirini's research [2]. We state the name of the features below.

length, caplen, frame-encapType, frame-timeShift, ip-flags, ip-flagsMF, ip-flagsDF, ip-ttl, ip-fragOffset, tcp-flagsAck, tcp-flagsSyn, tcp-flagsPush, icmp-code

When a feature is missing in a packet, for example, the `tcp.flag` in a UDP packet, we replace the non-existing value, *None*, with -1 . The idea is to use a unique value for missing features (i.e., not used by existing features) so the model learns the impact of a missing feature as well.

Sarhan et al. focused on feature extraction in machine learning-based IDS more deeply and have concluded that the choice of datasets significantly alters the performance of feature extraction techniques [10].

2.4 Sample Size Reduction

According to the scale of our paper, we have selected one malicious and one normal scenario to train and test our models. The benign scenario contains **75356** packets, whereas the malicious scenario contains **83068**.

3 Machine Learning Models

The Python library we use for employing machine learning is **scikit-learn**. We use decision trees as our lightweight model [4, 8, 9]. Decision Trees are a supervised learning non-parametric method for classification and regression problems. Their purpose is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

A problem to address when training a machine learning model is overfitting. The model, in order to keep the accuracy as high as possible, may overfit the dataset and lose its ability to generalize. We use validation curves to measure the overfitting of our model. We evaluate this metric with respect to the depth of the decision tree. We can see in Fig. 3 that when the depth excels 10, the model starts to overfit. Hence, we keep the hyperparameter of `max_depth` less than or equal to 10. There are other ways of dealing with overfitting in decision trees, such as ensemble techniques and pruning.

After finding the maximum possible depth for our tree (10), we then instantiate a *DecisionTreeClassifier* class and train it on the dataset. We set the maximum depth

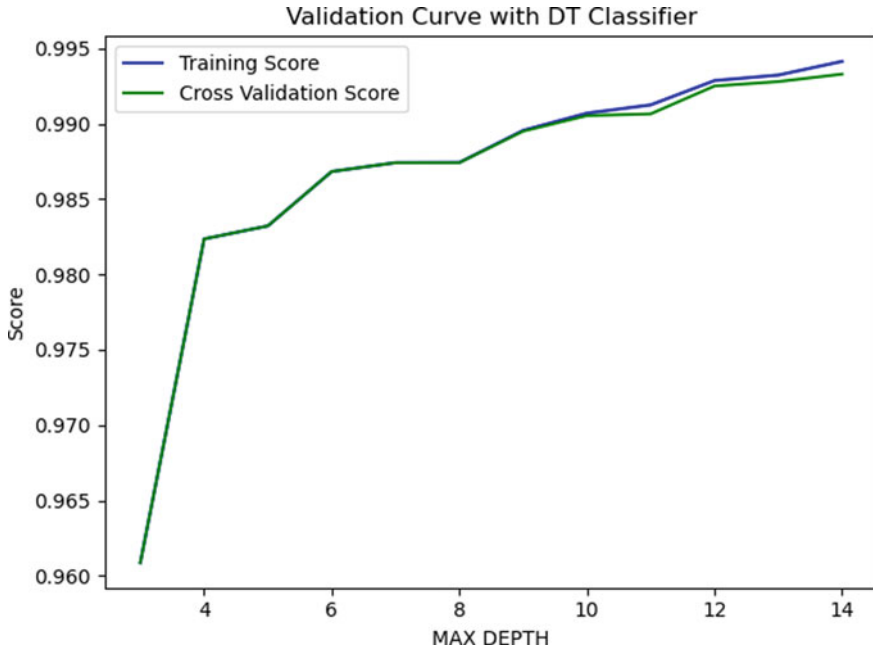


Fig. 3 Validation score with respect to the tree depth

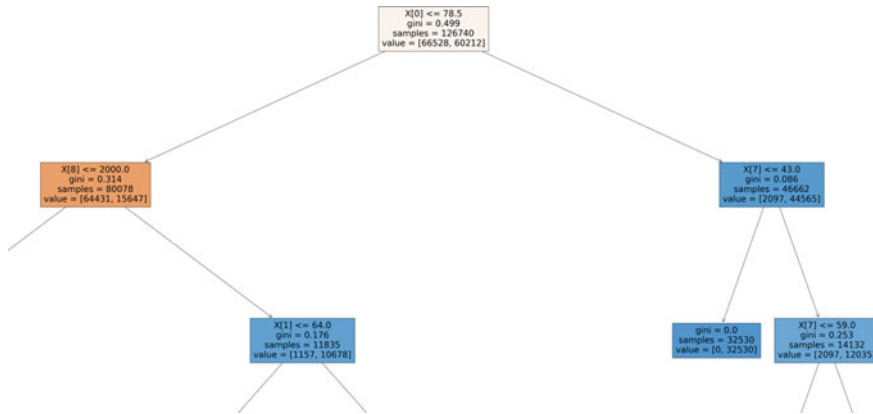


Fig. 4 A subtree of the final decision tree

of the decision tree to 10. To be able to see the decision rules and the number of samples satisfying them more clearly, we can refer to Fig. 4 which depicts a subtree of the final model.

Figure 5 illustrates a bigger subtree of our model. We can see in the subtree that the feature *len* ($X[0]$) appears in many logical rules and is an essential parameter for our decision tree.

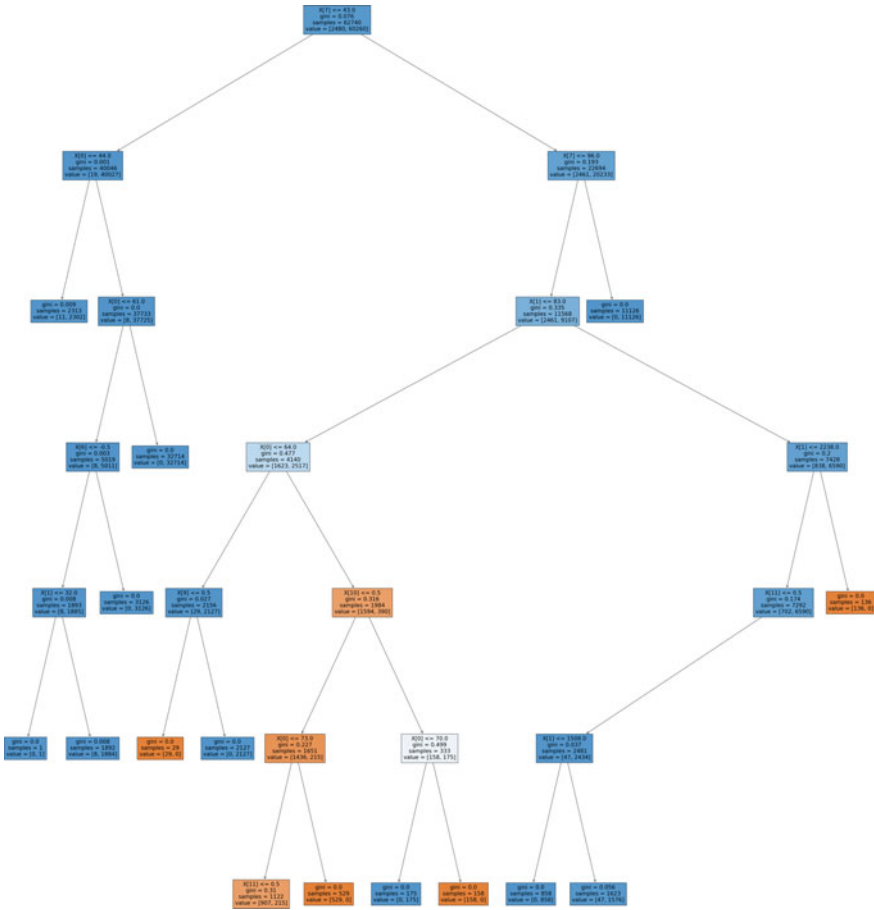


Fig. 5 Detailed subtree

4 Evaluation

At first, we describe the evaluation metrics of our decision tree in Table 1. To show that our lightweight model’s performance is satisfactory in this use case, we develop a neural network using the same library and compare its metrics with those of the decision tree. Table 2 briefly describes our neural network.

The neural network is **96.6%** accurate, which is even less than our decision tree. With that in mind that we have avoided overfitting, we can claim that our model outperforms classical neural networks in this scenario. It is important to note that the deep learning approaches may reach higher accuracy (99% and above), especially when the task includes the detection of the attack type [12] (Table 3).

Table 1 Decision tree evaluation metrics

Metric	Result
Accuracy	0.993
True positive (correctly classified as malicious)	16400 packets
False positive	58 packets
True negative	15087 packets
False negative	142 packets

Table 2 Neural network parameters

Parameter	Value
Solver	lbfgs
Hidden layers	4
Hidden layer dimensions	(5, 5, 5, 2)
Activation function	tanh

Table 3 Neural network evaluation metrics

Metric	Result
Accuracy	0.966
True positive (correctly classified as malicious)	15857 packets
False positive	219 packets
True negative	14771 packets
False negative	839 packets

5 Conclusion

In this paper, we presented an on-the-fly malicious packet classifier. It employs decision trees to capture IoT network packets and label them as normal or malicious. We demonstrated that our lightweight model outperforms complex neural networks while keeping the processing and storage requirements at a minimum.

6 Future Work

For future work, we may integrate this model with a packet capturer to automatically label all (or some) of the traffic in the network.

We selected our features manually. In future research, one might also automate feature selection using statistical or ML methods (intrinsic, wrapper methods, etc.)

One possible contribution to this research would be attack classification. The group of malicious packets found by the decision tree can be fed to a more complex machine learning model to detect the type of attack happening in the network. We may use several IoT attack categories for classification.

- **Denial of Service (DoS):** aims to make IoT devices unavailable by overloading the network and disrupting the services.
- **Distributed Denial of Service (DDoS)/Botnets:** an adversary compromises many IoT devices to employ a significant DoS.
- **Spoofing:** The attacker tries to manipulate an authenticated identity by forging.
- **Man-In-The-Middle:** The communication channel is compromised. The attacker can act after this attack as a proxy to read, write, and modify the packets.
- **Insecure Firmware:** After the control over an IoT device is gained, the device is used to attack other devices.
- **Data Leakage:** If the data is not encrypted, the privacy of the user data is compromised and may be used by an attacker to access the private network.
- **Botnets:** An adversary controls a network of connected IoT devices by which he performs malicious actions.
- **Brute Force Password Attacks:** A potential attacker can gather substantial processing power to try every possible secret a device possesses.

Acknowledgements Research was sponsored by the Army Research Office and was accomplished under Grant Number W911NF-21-1-0264. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

Appendix

A Full List of Features

The following table includes all the features that we collected, from which 13 were selected (Table 4).

Table 4 Appendix: feature list

len	icmp.respin
caplen	icmp.respto
frame.encaptype	data.len
frame.offsetshift	ssl.record.content-type
frame.len	ssl.record.version
frame.cap-len	ssl.record.length
frame.marked	arp.hw.type
frame.ignored	arp.proto.type
eth.lg	arp.hw.size
eth.ig	arp.proto.size
ip.version	arp.opcode
ip.hdr-len	http.response.code
ip.dsfield.dscpl	http.content-length
ip.dsfield.enc	http.response
ip.src	http.response-number
ip.dst	http.request
ip.len	http.request-number
ip.flags	classicstun.type
ip.flags.rb	classicstun.length
ip.flags.df	udp.srcport
ip.flags.mf	udp.dstport
ip.frag-offset	udp.length
ip.ttl	udp.chksum.status
ip.proto	udp.stream
ip.checksum.status	dns.flags.response
tcp.srcport	dns.flags.opcode
tcp.dstport	dns.flags.truncated
tcp.stream	dns.flags.recdesired
tcp.len	dns.flags.z
tcp.seq	dns.flags.checkdisable
tcp.nxtseq	dns.flags.rcode
tcp.ack	dns.flags.queries
tcp.hdr-len	dns.count.answers
tcp.flags.res	dns.count.authr
tcp.flags.ns	dns.qry.name.len
tcp.flags.cwr	dns.count.labels

(continued)

Table 4 (continued)

len	icmp.respin
tcp.flags.ecn	dns.resp.type
tcp.flags.urg	dns.resp.class
tcp.flags.ack	dns.resp.ttl
tcp.flags.push	dns.resp.len
tcp.flags.reset	igmp.version
tcp.flags.syn	igmp.type
tcp.flags.fin	igmp.max-resp
tcp.window-size-value	igmp.checksum.status
tcp.window-size	ntp.flags.li
tcp.window-size-scale-factor	ntp.flags.vn
tcp.checksum.status	ntp.flags.mode
tcp.urgent-pointer	ntp.startum
tcp.options.nop	ntp.ppoll
tcp.options.mss-val	ntp.root-delay
tcp.options.sack-perm	ntp.rootdispersion
tcp.analysis.bytes-in-flight	ntp.precision
tcp.analysis.push-bytes-sent	bootp.type
tcp.payload	bootp.hw.type
icmp.type	bootp.hw.len
icmp.code	bootp.hops
icmp.ident	bootp.secs

References

1. Amouri A, Alaparthi VT, Morgera SD (2018) Cross layer-based intrusion detection based on network behavior for IoT. In: 2018 IEEE 19th wireless and microwave technology conference (WAMICON), pp 1–4
2. Anthi E, Williams L, Słowińska M, Theodorakopoulos G, Burnap P (2019) A supervised intrusion detection system for smart home IoT devices. *IEEE Internet Things J* 6(5):9042–9053. <https://doi.org/10.1109/JIOT.2019.2926365>
3. Anthi E, Williams L, Burnap P (2018) Pulse: an adaptive intrusion detection for the internet of things. In: *Living in the internet of things: cybersecurity of the IoT*, pp 1–4. <https://doi.org/10.1049/cp.2018.0035>
4. Bilge L, Kirda E, Kruegel C, Balduzzi M (2011) Exposure: finding malicious domains using passive DNS analysis
5. Buczak AL, Guven E (2016) A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Commun Surv Tutor* 18(2):1153–1176. <https://doi.org/10.1109/COMST.2015.2494502>
6. Cao Y, Zhang L, Zhao X, Jin K, Chen Z (2022) An intrusion detection method for industrial control system based on machine learning. *Information* 13(7):322
7. Garcia S, Parmisano A, Erquiaga MJ (2020) Iot_23: a labeled dataset with malicious and benign IoT network traffic. <https://doi.org/10.5281/zenodo.4743746>

8. Kruegel C, Toth T (2003) Using decision trees to improve signature-based intrusion detection. In: Proceedings of the 6th International workshop on the recent advances in intrusion detection (RAID'2003), LNCS vol 2820. Springer Verlag, pp 173–191
9. Salzberg SL (1994) C4.5: programs for machine learning by J. Ross Quinlan, Morgan Kaufmann Publishers, Inc. 1993. Mach Learn 16(3):235–240. <https://doi.org/10.1007/BF00993309>
10. Sarhan M, Layeghy S, Moustafa N, Gallagher M, Portmann M (2022) Feature extraction for machine learning-based intrusion detection in IoT networks. Digital Commun Netw
11. Shukla P (2017) MI-ids: a machine learning approach to detect wormhole attacks in internet of things. In: 2017 Intelligent systems conference (IntelliSys) pp 234–240
12. Soltani M, Ousat B, Siavoshani MJ, Jahangir AH (2021) An adaptable deep learning-based intrusion detection system to zero-day attacks. arXiv preprint [arXiv:2108.09199](https://arxiv.org/abs/2108.09199)
13. Yu T, Sekar V, Seshan S, Agarwal Y, Xu C (2015) Handling a trillion (unfixable) flaws on a billion devices: rethinking network security for the internet-of-things. In: Proceedings of HotNets, 5p. Philadelphia, PA

Cyber Security Issues and Challenges on Non-fungible Tokens



N. Kala

Abstract Blockchain Technology helps buy and sell digital assets known as Non-Fungible Tokens (NFT). It is an asset that is not fungible. It is a different type of asset that exists in the virtual world and uses cryptocurrencies. Unique individual cartoons, like the 'Bored Ape Yacht Club', are made available through the NFT collection with unlimited, choices. The owners of such NFTs exhibit them with pride as avatars on social media. These tokens are not merely images. They include assets in the form of art, music, video game items, collectibles, sports memorabilia, domain names, and tweets, on various websites such as 'Decentraland' and the 'Sandbox'. It is interesting to note that even land in the virtual world using NFT. This process of recording the ownership details of a digital asset is sure to revolutionize the thinking in the minds of people. This paper aims to study the overview of NFTs in detail. It further focuses on the advantages and limitations of NFTs. An attempt has been made to compare NFTs with Blockchain and cryptocurrencies. This paper also deals with the standards and protocols used for NFTs. Further, the issues and challenges with cyber security and privacy in NFTs are discussed. This paper focuses on the future scope of NFTs in the context of the 'Metaverse'.

Keywords Non-fungible tokens · Cryptocurrencies · Blockchain · Metaverse

Abbreviations

NFT	Non-fungible tokens
ID	Identification
ERC	Ethereum Request for comments
ACID	Atomic, Consistent, Isolated, and Durable

N. Kala (✉)

Centre for Cyber Forensics and Information Security, University of Madras, Chennai 600005, India

e-mail: nkala@unom.ac.in

STRIDE	Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege
DoS	Denial of Service
EoP	Elevation of Privilege
NIFTY	National Stock Exchange FIFTY
IPFS	Inter Planetary File System
DeFI	Decentralized Finance

1 Introduction

Non-fungible tokens (NFT) [1] are cryptographic tokens that live on a blockchain, each identified with a unique identification code and metadata that cannot be replicated. NFTs are digital representations of physical assets. They are often compared to digital passports since each token contains a unique non-transferable identity to distinguish it from other tokens. The distinct construction of NFTs has the potential for several use cases. The market for NFTs is centered on collectibles. For example, digital artwork, sports cards, and rarities. Because NFTs are based on blockchains, they can be bought, sold, and traded more efficiently. This can be achieved without intermediaries. This simplifies transactions reduces the probability of fraud and creates new markets. But there is a key difference between crypto and NFTs. Like cryptocurrencies, digital currencies are fungible meaning they can be traded or exchanged one for another. For example, one bitcoin is always equal to another bitcoin. NFTs shift the crypto paradigm by making each token unique and irreplaceable which makes one NFT to be equal to another. NFTs are in the evolution of the simple concept of cryptocurrencies. Today's modern finance systems consist of sophisticated trading and loan systems for different asset types of physical systems. NFTs are a step forward in the reinvention of this infrastructure through digital representations.

To sum up, Non-fungible Tokens (NFTs).

- NFTs are unique cryptographic tokens that exist on a blockchain and cannot be replicated.
- NFTs can represent real-world items like artwork and real estate.
- “Tokenizing” these real-world tangible assets makes buying, selling, and trading them more efficient while reducing the probability of fraud.
- NFTs can also function to represent individuals' identities, property rights, and more.
- Collectors have sought NFTs as their value initially soared, but has since moderated.

2 Examples of NFTs

NFTs have the potential for several use cases. For example, they are an ideal vehicle to digitally represent physical assets like real estate and artwork. Perhaps the most famous use case for NFTs is that of cryptokitties [2]. Launched in November 2017, cryptokitties are digital representations of cats with unique identifications on Ethereum’s blockchain. Each kitty is unique and has a price in ether. They reproduce among themselves and produce new offspring, which have different attributes and valuations compared to their parents.

Non-fungible tokens are also excellent for identity management. Consider the case of physical passports that need to be produced at every entry and exit point. By converting individual passports into NFTs, each with its unique identifying characteristics, it is possible to streamline the entry and exit processes for jurisdictions. NFTs can serve as an identity management purpose within the digital realm as well [3].

3 Origin of NFT [4, 5]

The origin of NFT is summarized in Table 1.

4 Working of an NFT

Key Technical Components

Blockchain—A blockchain is an immutable growing list of records linked together by cryptography, it is decentralized as it is managed by a peer-to-peer network of nodes for it to be ultimately used as a publicly distributed ledger.

Smart Contracts—The basic idea of smart contracts is that many kinds of contractual clauses—such as liens, terms, delineation of property rights, etc.

Address and Transaction—In the case of sending a crypto asset or currency, the user has to prove the ownership of the corresponding private key and then send the asset or currency to another address with a valid digital signature.

Data Encoding—Converting data from one form to another is called data encoding. Usually, it is employed to compress to save disk space or to expand for better resolution.

Web3 Wallet—A Web3 wallet is a tool to interact with Web 3.0 applications. The Web3 wallet is capable of storing crypto assets.

Non-fungible tokens (NFT) are digital assets that can be purchased and sold on blockchain technology on specialized platforms just like cryptocurrencies. They are basically online objects. Clear ownership is provided over such digital items. Most NFTs are a part of the Ethereum blockchain. Ethereum is a cryptocurrency like

Table 1 Summary of origin of NFT

Year	Origin of NFTs
<i>2016—The early history of NFTs</i>	
Colored coins	Meni Rosenfield introduced colored coins in the year 2012. This led to the use of colored coins for the bitcoin blockchain. Colored coins help to describe The use of colored coins was aimed to portray the ways of exposition of the real-world assets on blockchain platform. This helps to ascertain the ownership of such assets which are quite similar to those of regular bitcoins. However, the use of concept of ‘token’ helps in determining its use and in the process makes them separate and yet remain unique
NFT—Quantum	NFT—Quantum was first created by Kevin McCoy in the year 2014 NFT ‘Quantum’ was the first known coin minted in namecoin block digital by artist Kevin during 2014. ‘Quantum’, represents a digital image of a pixelated octagon’ which gives perceptible changes in colors resembling that of an octopus
Counterparty	Bitcoin 2.0 namely the Counterparty platform emerged as a platform that helped in the creation of digital assets
Spells of genesis	Spells of genesis followed the Counterparty NFTs. It helped in the issue of ‘in-game assets’
Memes start	Memes start emerged in the year 2016. It worked on the counterparty platform which an emergence of rare pepes NFTs. Which is a variation on the internet meme ‘Pepe the Frog’. It id based on non-fungible tokens (NFTs) recorded on the CounterParty platform
<i>2017–2020 NFTs go mainstream</i>	
NFTs shift to ethereum	There has been a big shift for NFTs to Ethereum. It introduces tokens with an aim of introduction of standards for tokens
Token standards	Token standards got introduced. This helps in the process of creation, issue, and deployment of tokens in the blockchain technology
Cryptopunks	John Watkinson and Matt Hall were influenced by the London punk culture and the cyber punks created by Cryptopunks
Cryptokitties	In the hackathon conducted for the Ethereum ecosystem, Axiom Zen introduced Cryptokitties. It is a virtual game based on Ethereum blockchain. In this game, the players, are allowed to adopt crypto kitties, breed them and also trade them. The Kitties are also stored on crypto wallets. It helped people to earn huge profits
NFT gaming and metaverse	NFT gaming and metaverse is a decentralized platform on Ethereum blockchain using R programming. It essentially helps its players to earn and build and perhaps own on a blockchain. Platforms such as Decentraland allow the players to explore, play games, build, collect items, and build own blockchain
Token standards	Token standards got introduced. This helps in the process of creation, issue, and deployment of tokens in the blockchain technology
Platforms and games	Enjin Coin (ENJ), helps its developers to tokenize and thereby making the process of earning for the in-game items a value in the real-world. Axie Infinity (AXS) is a game that emerges as partly owned and operated by its players

(continued)

Table 1 (continued)

Year	Origin of NFTs
<i>2021—The Year of the NFT</i>	
2021	2021 is the year of NFT as there was an explosive trend in demand and supply of NFT. It paved the way for online auctions in the NFT art whereby auction houses like Christie’s and Sothey’s also got indulged in the process of auctioning online
Christie’s art	The process of Christie’s sale led to a noteworthy NFT marketplace. It has paved the way for the emergence of new NFTs and their platforms. The entire process helps in the achievement of new standards and thereby helps create authentic digital assets with absolute unique identity
Facebook goes into meta platform	There is a process of facebook getting rebranded as meta and the adoption of metaverse has eventually increased the demand

Bitcoin. NFTs are unique crypto tokens that are managed on a blockchain. NFTs have a code, unique ID and other metadata which is non-duplicated and blockchain acts as the decentralized ledger which traces the ownership and history of transaction of each NFT. Using Ethereum one can create an NFT through a contract-enabled blockchain with appropriate tools and support. This process is known as ‘minting’. Ethereum was one of the first widely used standards.

5 Protocols for the Establishment of NFT

The fundamental requirement for the establishment of NFT is to have a distributed ledger for records and interchangeable transactions or trading in a peer-to-peer network. Researchers in their technical report [6] have come up with two approaches.

1. Top-to-bottom approach

This is very simple but it is a classical path. In this approach, NFTs are built from the ‘initiator’, and then sell them to the ‘buyer’.

2. Bottom to top approach

The second approach on the contrary reverses this path by setting an NFT template, so that every user can create its own unique on-top.

The comparison between the topto bottom and bottom to top approaches is summarized in Table 2.

The workflow of NFT is illustrated in Fig. 1.

Table 2 Comparison between Top to bottom versus Bottom to top approach

Top to bottom	Bottom to top
For the first design (e.g., CryptoPunks [7]), an NFT protocol consists of two roles namely: 1. NFT owner 2. NFT buyer	For this design (e.g., Loot [8]), the protocol consists of two roles namely: 1. NFT creator 2. NFT buyer In most cases, a buyer can also act as a creator and vice versa because an NFT product is created based on random seeds when a buyer bids for it. This encompasses the functions that are customizable and user friendly
NFT digitize	Template create. A smart contract is created by the initiator through a template in order to setup using several basic rules and features
The owner of an NFT owner digitizes raw data into an appropriate format and checks the completeness and accuracy the of file, title, and description are	Such as character style, weapons, or accessories in a game
NFT sign. A transaction including the hash is signed by the owner of NFT data, and then directs the transaction to a smart contract	NFT Randomize: The process involved bidding by the buyer for an NFT and customize the product based on certain additional features which are randomly chosen from a predefined database at that state
NFT mint &trade. The process of minting and trading begins after the smart contract receives the transaction with the NFT data. Token standards is the main mechanism behind NFTs is the logic	NFT mint &trade. The process of minting and trading begins when starts once the corresponding smart contract is initiated
NFT confirm. The process of minting completes upon confirmation of the transaction. By this approach, There will be a unique blockchain address for an NFT which will be as persistence evidence	NFT confirm. NFTs will be persistently stored on-chain once the consensus procedures are completed. All the procedures are conducted through smart contracts

6 Prevailing Token Standards

NFT related Token standards are listed below:

- ERC-20 [9]—There are several ERC20-compliant tokens deployed on the Ethereum network. Various teams have written different implementations. These implementations have different trade-offs: from gas saving to improved security.
- ERC-721 [10]—Any NFT on Ethereum works with wallet/broker/auction applications through a standard interface. Other standards that work with applications include the ERC-20 token standard. The ERC-721 standard’s token ID is a single non-fungible index and the group of these non-fungibles is deployed as a single contract with settings for the entire collection.

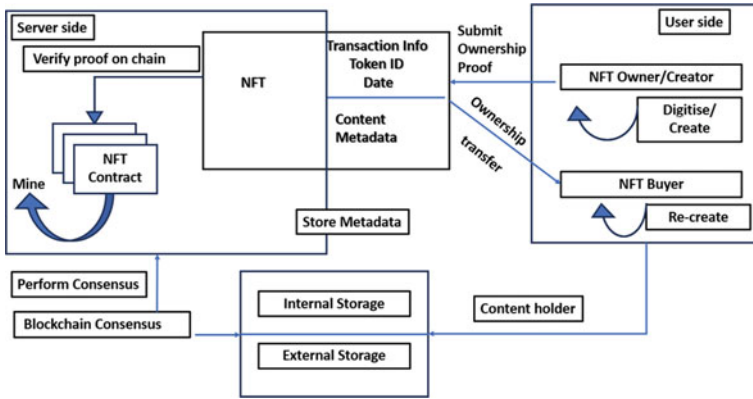


Fig. 1 Workflow of NFT [6]

- ERC-1155 (Multi Token Standard) [11]—It is a standard that provides an interface for contracts that manage multiple token types. Deployment of a single contract includes any combination of configurations such as fungible or non-fungible tokens may also include semi-fungible tokens. ERC-1155 is any number of fungible or non-fungible tokens whereas existing standards such as ERC-20 require the deployment of separate contracts per token type. Contrastingly, the multi-token Standard allows each token ID to represent a new token type that can be configured. This may have its own metadata, supply, and other attributes.

Properties desired properties are tabulated hereunder—Table 3.

NFT schemes are essentially decentralized applications [12], and thus enjoy the benefits/properties from their underlying public ledgers.

7 NFTs Digital Disruption, Opportunities and Transformation

NFT ecosystem is still in a nascent stage and has to mature fully. This section discusses different types of issues associated with NFT. Different entities are involved in NFT transactions. They are:

- **Users**
 - Content Creator
 - Seller
 - Buyer
- **Market Places**—are Application platforms where NFTs (assets) are traded. There are typically two main components of an NFT:

Table 3 Properties of NFT

S.No	Property	Description
1	Verifiability	Ownership of NFT along with its token metadata can be verified publicly
2	Transparent execution	Minting, selling and purchasing are the activities of NFTs that are publicly accessible
3	Availability	It is to be noted that the NFT system never goes down meaning they are available always and all the tokens and issued NFTs are always available for selling and buying
4	Tamper resistance	Trading records of the NFT metadata are stored persistently. Once the transactions are confirmed they cannot be manipulated
5	Usability	NFTs are usable in the sense that they are having up-to-date information about the ownership that is user-friendly with clarity of information
6	Atomicity	Trading NFTs can be completed in a durable transaction. It is called ACID transaction and it means one atomic, consistent, isolated, and durable transaction. The shared execution state is the way that the NFTs can run
7	Tradability	It is possible for every NFTs and their analogous products can be traded and exchanged indiscriminately

- User-facing front-end in web
- SmartUser-facing is a smart contracts collection that interacts with the blockchain. They are of two types:

Firstly ‘Marketplace contracts’ are used to implement the part of the NFT protocol that interacts with the blockchain, and
Secondly ‘Token contracts’, are used to manage NFTs.

Marketplaces allow users to perform the following activities:

- User Authentication.
- Token Minting.
- Token Listing and
- Token Trading.

The token-related activities are collectively called events. Three types of NFTM protocol designs are possible depending upon where these events are stored. They are.

- On-chain: In an On-chain market place all the events take place live on the blockchain (e.g. Axie, CryptoPunks, Foundation, and SuperRare).
- Off-chain: In an Off-chain market place the events are recorded in a centralized, off-chain database managed by the NFTM (Nifty).
- Hybrid: In a Hybrid market place the events are stored either on-chain or off chain depending on their type (OpenSea and Rarible).

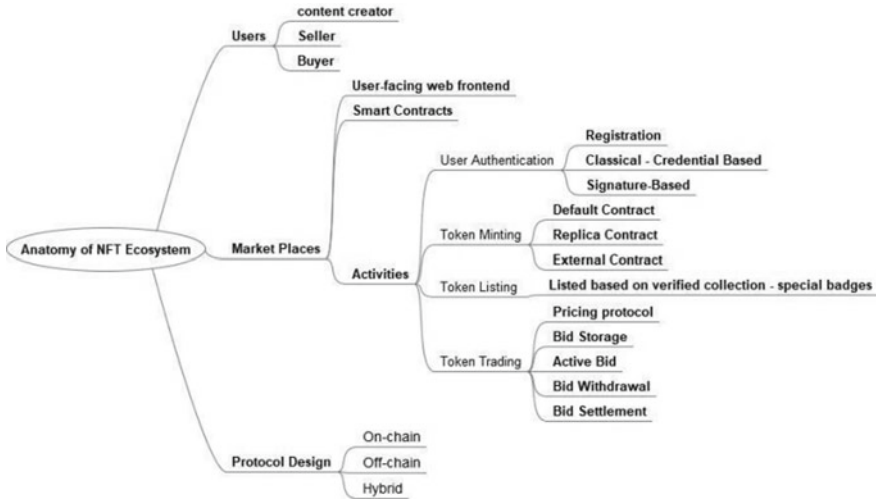


Fig. 2 Anatomy of NFT ecosystem

Anatomy of NFT ecosystem is illustrated in Fig. 2.

NFTs are used to represent digital assets [13]. They are unique and cannot be replaced. Digital game items, collectible art pieces, or virtual real estate are some of the common examples of NFTs. Storage of digital assets of NFTs takes place on a blockchain platform. Transactions are recorded on a blockchain which is similar to the recording of cryptocurrencies. Even though both store the transaction in blockchain, there is a key difference between NFT and Cryptocurrencies. NFTs store the information about the asset itself on the blockchain rather than just storing details of the transaction. There are infinite opportunities for NFT. It is not only limited to the artwork but also allows tokenization and commercialization of digital assets. Some examples include gaming, virtual events, captured moments, collectibles, trading cards, music, memes, e-tickets, domain names, metaverse, and also property. NFT also finds its application in banking and finance. A wide range of opportunities and the current industry with examples are summarized in Table 4.

8 Cyber Security Issues and Challenges

8.1 Cyber Attacks on NFT

In order to understand the cyber security attacks first one must understand Cyber Security. The following are the basic elements of cyber security. These include confidentiality, integrity, availability, non-repudiation, authenticity and availability. NFT is a combination of technology comprising of Blockchain, Storage and Web

Table 4 Wide range of opportunities and the current industry with examples

Industry	Disruption and transformation	Examples
Gaming industry	<ul style="list-style-type: none"> • An interesting feature is the ‘breeding’ mechanism in the gaming industry • Raising pets personally • Opportunities for spending time in breeding new offspring • Purchasing rare and limited editions of virtual pets and selling them at a high price • It attracts investors to join the games • Make NFTs prominent • Ownership of records provided benefits developers • Provides royalties 	Cryptogames: <ul style="list-style-type: none"> • CryptoKitties • Cryptocats • CryptoPunks • Meebits • Axie infinity • God unchanged
Digital collectibles	<ul style="list-style-type: none"> • The industry is undergoing an unprecedented transformation with many benefits from NFTs • Increased Movement and Tradability of Assets • Protect the Value of Tokenized Assets • Bring the assets to unified form with possibilities to evolve 	<ul style="list-style-type: none"> • Trading cards • Wines • Digital images • Videos • Virtual real estate • Domain names • Diamonds • Crypto stamps • Real intellectual properties
Metaverse	<ul style="list-style-type: none"> • Lifelike gaming with real-world gains • Virtual shopping that mimics In Real-Life (IRL) • Exclusive design fashion • Events without price, time commitment • Enhanced social community experiences 	<ul style="list-style-type: none"> • Decentraland • Cryptovoxels • Somniumspace • Megacryptopolis • Sandbox
Virtual events	<ul style="list-style-type: none"> • Attendee check in • Participant engagement measurement • Sourcing simple meetings • Digital marketing and signage • Abstract and presentation management • Second screen and engagement • Exhibitor and Floor plan management • Association management 	<ul style="list-style-type: none"> • Mobile apps • AR/VR • Chatbots • Streaming • Sourcing • Registration • Attendee tracking CRM • Data management • Matchmaking

Application. Since it is based on web application and storage, it makes the whole NFT ecosystem vulnerable to cyber-attacks. Every aspect and components may become an attack surface. In order to evaluate risk evaluation and the security in NFT ecosystem STRIDE model of threat intelligence can be applied. STRIDE is a model for identifying computer security threats. It provides a mnemonic for six categories of cyber security threats. The threats are Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service and Elevation of Privilege. These threats are violations of a property for a system and the same is given in Table 5.

Table 5 Threat against violation of a desired property

Threat	Desired property
Spoofing	Authenticity
Tampering	Integrity
Repudiation	Non-repudiation
Information disclosure	Confidentiality
Denial of service	Availability
Elevation of privilege	Authorization

8.1.1 Spoofing

Spoofing is the ability of an adversary to impersonate another entity. For example, it could be someone or something pretending to be someone else or something else. By doing so an attempt can be made by the adversary to gain confidence, get access to systems, steal data, steal money, or spread malware. In a system, it corresponds to authenticity. A malicious attacker may exploit authentication vulnerabilities. A user’s private key can be exploited to steal the assets and illegally transfer the ownership of NFTs whenever a user is trying to interact in order to mint or sell the NFTs.

8.1.2 Tampering

Manipulating the application data that is exchanged between client and server by modifying the parameters. Tampering refers to the malicious modification of NFT data, which violates integrity. The metadata and ownership of NFTs cannot be maliciously modified after the transaction is confirmed. However, the data stored outside the blockchain may be manipulated.

8.1.3 Repudiation

Repudiation refers to denying an act. For example, the author of a statement cannot dispute [14]. This is related to the security property of non-repudiation [15]. In the context of NFT, a particular user after sending NFT to another user cannot deny it. This is assured and achieved by the security of the blockchain and hence it cannot be forged. However, malicious hackers can tamper with hash data.

8.1.4 Information Disclosure

Unauthorized users can compromise confidentiality and thereby information Disclosure occurs [15]. In the NFT system, the instruction code in smart contracts is entirely transparent and any changes can be publicly accessible. A malicious attacker can

easily exploit the linkability of the hash and transaction only if the user puts the NFT hash into the blockchain.

8.1.5 Denial of Service (DoS)

DoS attack [16] is a type of network-based attack. In this attack, a malicious attacker aims to render a server unavailable. The intended users will not be able to access normal functions, instead, the services will be interrupted. DoS attacks will target and violate the availability and break down the NFT service. The attacker then will use the resources in an unauthorized way. Since blockchain is the backbone of NFT, it guarantees the availability of users' operations thereby legitimate users can use the resources. This prevents data loss either due to accidental errors or attacks. However, DoS to NFT service attacks can happen when centralized web applications or the raw data is outside of the blockchain.

8.1.6 Elevation of Privilege

An attacker can elevate the privilege once they compromise the NF ecosystem. Elevation of Privilege [17] is a property that relates to authorization. In this type of threat, an attacker may do a privilege escalation than that was initially granted. In the NFT system, the selling permissions are managed by a smart contract. A poorly designed smart contract may make NFTs lose such properties.

Potential Security Issues and Corresponding Solutions of NFTs are summarized in Table 6.

9 Cyber Security Issues in NFT

This section focuses cyber security issues in the NFT ecosystem. When discussing about NFT there are three entities that have been identified namely the users, marketplaces, and external entities. Correspondingly the concerns that arise in mind with respect to these entities and the issues thereof are:

- How do NFTs operate?
- Is there any vulnerability in NFTs?
- Are these vulnerabilities exploitable?
- What are the ways in which external entities pose a threat to NFTs?
- Is there a possibility of users getting involved in fraudulent activities?
- Does malpractice result in monetary loss?

Popular NFT marketplaces include OpenSea, AXIE, CryptoPunks, Rarible, SuperRare.

Table 6 Cyber security issues and their countermeasures based on STRIDE model

Stride model	Desired protection	Cyber security issues	Countermeasure for the cyber security issues
Spoofing	Authenticity	<ul style="list-style-type: none"> • There is a possibility of exploitation of authentication vulnerabilities by a hacker • There is also a possibility of an attacker stealing the private key of a user 	<ul style="list-style-type: none"> • Formal verification process to be ensured on the smart contract • Cold wallet can be used to prevent the private key leakage
Tampering	Integrity	There is a possibility of manipulation of the data stored outside the purview of the blockchain	The integrity of data transfer can be ensured by sending both the original data and hash data to the NFT buyer when the actual trading takes place
Repudiation	Non-repudiation	It is diligent to bind the hash data with an attacker’s address	Non-repudiation can be mitigated using a multi signature contract
Information disclosure	Confidentiality	Confidentiality of transactions can be compromised by an attacker and exploit the hash and transaction in order to establish a connection with a specific buyer or seller of NFT	Replacing smart contracts with privacy preserving smart contracts in order to protect the privacy of user
Denial of service	Availability	In case of the asset is stored outside the blockchain, then there is a possibility of NFT data becoming unavailable	Countermeasure for denial of service can be achieved using hybrid blockchain architecture with a consensus algorithm
Elevation of privilege	Authorization	An Elevation of privilege attack occurs due to an inadequately designed smart contract that results in the loss of certain properties of NFT	Elevation of privilege can be achieved through a formal verification process of the smart contracts [18]

Foundation, and NIFTY. There are several issues with respect to marketplace relating to user authentication, Token Minting, Token listing, and Token Trading. In a market place, some of the above mentioned operations are optional, some are mandatory, others are partial, and some exist while some do not exists.

9.1 Issues with Respect to User Authentication

9.1.1 Money Laundering

Money laundering schemes is one of the major issue in the physical world especially in the trading of Art objects. This makes it much easier in an NFT transaction of art work as there is no necessity to transport the art work and the trading is done by anonymous users. Some of the crypto exchanges are highly regulated such as Coinbase and Binance US. It is a mandatory requirement that in these crypto exchanges, one has to create an account and provide Personally Identifiable Information (PII) such as residential address, Social Security Number and supporting documents as well. Verification of user identity is the first step to deter cyber criminals who try to access anonymously and perform money laundering activities. Verification such as KYC makes it impossible to use such platforms by imposing regulatory restrictions.

9.1.2 Two-Factor Authentication [2FA]

Traditional fintech companies such as banks, brokerages, and cryptocurrency exchanges (e.g. Coinbase) provide two-factor authentication as a security measure as an option. Some NFTs offer 2FA as an option and not as a default. Some allow the use of wallets (e.g. Ethereum). Similarly NIFTY allow 2FA as optional. However recently a number of accounts in NIFTY where 2FA was optional. Hackers compromised the accounts in March 2021 [19].

10 Issues with Respect Token Minting

A token can be created or minted (6) by calling the suitable method of token contract. There are certain standards in order to create a token minting such as ERC-721 ERC-1155. The single token minted or created is the one that manages the ownership of several NFTs. Each token, therefore, has the following attributes.

They can be uniquely identified as token contract address-token id pairs on the blockchain. NFT can be created or minted in the following ways:

- Default contract.
- Replica contract.
- External contract.

A family of NFTs with a common theme is called a Collection (e.g. CryptoPunks). There can be two issues with respect to token listing namely:

- Verifiability of Token Contracts
 - A token is verifiable. The external token contracts must be verified as they can be malicious.
 - A malfunctioning contract can exhaust gas without doing any work.
 - Countermeasure—NFT should make the source token contract to be available for public scrutiny before minting
- Token Metadata can be tampered

Each asset holds a pointer corresponding to the asset which has the metadata of a token. There will be a loss of significance of the token if there is a change in the metadata. NFT such as a piece of art represents a particular asset that is sold in the market. Buyer's belief is infringed when there is a change in metadata. Generally, the metadata's content and location are decided upon during minting. Tampering of metadata can be made possible by a malicious owner or creator by manipulating the metadata URL and altering the metadata itself. This happens post mining.

Countermeasures for manipulating metadata by a creator can be protected by disallowing at the contract level. Alteration of metadata can be protected by hosting metadata in Inter Planetary File System (IPFS). This is possible because the Universal Resource Locator of an object is stored in IPFS including the hash of its content. In this way, the metadata cannot be tampered. In some other internal token contracts there is no way to update the metadata URL.

10.1 Issues with Respect Token Listing

The process of listing assets by owners after having created one is called Token Listing. In this process, in order to sell an asset in NFT platforms such as NIFTY, SuperRare, and Foundation, mandate verification by either a seller or collection. Some NFTs are open such as OpenSea, Rarible are verified, thereby provide credibility and raise the confidentiality of the buyers.

10.1.1 Violation of the Principle of Least Privilege

During Token Listing there can be a violation of least privilege. When an NFT listing takes place the ownership can be transferred from owner to buyer. In order to achieve this the NFT does one of the following:

- The owner of NFT transfers the asset to an escrow account which is maintained by an outsider who engages in facilitating to withhold of valuables, like money, deeds of the property, and documents relating to individuals' finance. This is done

for the two agreeing parties who meet during a transaction involving finance. It is interesting to note that in this connection, an operator of an Ethereum account shall manage all the NFTs on behalf of the owner.

- This is a risk because one single escrow account managed by NFT can hold the all the assets and thereby violates the principle of least privilege.
- There another issue when there is a marketplace hack.
- Additionally, issue of compromise of private key of the seller (owner) happens when the key gets compromised.

10.1.2 Violation Invalid Caching

During listing of NFTs, during the sale of an asset, some NFTs such as OpenSea, Rarible, etc. leverage by avoiding repeated requests through caching in order to fetch the associated assets in the form of images. When cache goes out of sync, due to an image getting updated or disappears, the buyer could be tricked to purchasing a nonexistent asset or an asset that is displayed in the token listing as stale cache.

10.1.3 Verification of Seller and Collections

Buyer communities are receiving not only preferential treatment through verified selling and collections but also getting attracted with greater attention. Verification comes with financial benefits. It also brings greater trust to the NFT. This helps in the community identity authentic creator and content. For example, a blue checkmark badge on an account means it has been verified for authenticity by Open Sea. An account must have certain criteria such as a user name, a profile picture, a verified email address, and a connected Twitter account. In the absence of verification, it is abused in different ways:

- **Forging verification badge**

Scammers indulge in forging and manipulating the profile pictures making the profiles appear visually indistinguishable by superimposing them through an image verification badge.

- **Impersonation**

Scammers abuse weak verification procedures by creating fake profiles. They create fake profiles without actually proving the ownership of the corresponding accounts by just submitting social media handles.

- **Wash Trading**

Wash trading happens when a trader purchases and sells the same securities many times within a short time in order to confuse other market participants about the price of an asset or liquidity position. An attacker can take the control of fictitious trade

which is activated among multiple accounts in order to inflate the quantum of sales value.

10.2 Issues with Respect to Token Trading

Issues with respect to Token Trading involve Lack of transparency, Fairness in Bidding, Royalty distribution, and marketplace fees are the issues with respect to token trading and are discussed below.

10.2.1 Lack of Transparency

Crypt tokens are digital assets built on cryptocurrency's blockchain [20]. A blockchain is a digital ledger that stores information in blocks which are linked. Information stored could be transaction records or they could be programs that operate on the blockchain. Such a program is called a smart contract that resides within the blockchain. The information therein can be accessed as and when needed. Hence the smart contract ensures that the information stored therein is transparent and immutable [21]. An NFT smart contract is a mechanism for implementing a sale agreement between the NFT owners and the Buyers. They are self-executing and capable of checking the contract terms have been satisfied. In this process, the terms can be executed without the need for an intermediary or a centralized authority. When an NFT is purchased a unique token is issued. This has the information and details of the smart contract. The owner of the NFT can display the asset listing and sell it which proves the authenticity of the owner. Hence it prevents counterfeiting of NFT and can be verified publicly. Each transaction in the blockchain includes the following information.

- Owners (Seller) Address.
- Buyers (Current) Address
- The proceeds of selling NFT
- The time taken for transfer of ownership.

At the same time, if the records and transition are stored off-chain, verification of transactions, trading, and ownership is impossible. Moreover, spurious forged sales can abuse off-chain records and inflate the volume of trading as well as genuine activity. Further, there can be tampering, and censorship, and may be prone to disappear from the NFT database in an off-chain record.

Fig. 3 On-chain versus Off-chain bidding



10.2.2 Fairness in Bidding

Implementation of bidding in NFT can be of two types and is illustrated in Fig. 3.

- **On-chain bidding:**

On-chain bidding happens through a smart contract. It asks for the bid amount to be deposited while bidding.

- **Off-chain bidding:**

In Off-chain bidding, the order book can be maintained without the requirement of prior payment. This can be abused both by the NFT as well as the users. There can also be hype and inflation of the bid volume, as the bids are invisible from the blockchain. Hence, the NFTs are vulnerable to ‘bid pollution’.

10.2.3 Royalty Distribution and Fee Evasion in Marketplace

A creator creates an artwork, for example, he owns the artwork. He or she can also sell Royalties in NFT can be set for the creator-owner and it is offering a share of income every time the artists’ work is sold and this can be retained by the artists perpetually. Every time the asset is traded the original creator receives royalty recursively. Artists, musicians, content providers, and creators of all types can receive royalties. Programs such as smart contracts self-execute and enforce commercial agreements to regulate the distribution of royalties. Abuse of royalties has also been identified during implementation such as cross-platform, non-enforcement, and post-enforcement. Potential abuse of distribution of Royalties is listed in Table 7.

Table 7 Potential abuse of royalty distribution

S.No	Implementation	Abuse
1	Cross-platform	<ul style="list-style-type: none"> ● Enforcement of royalty is either due to marketplace contract or through the application ● The royalty information is not shared with each other by NFT ● This means royalty set on one platform is not visible to another platform ● Lack of coordination is exploited by malicious sellers in an NFT platform that will evade royalty while trading
2	Non-enforcement	<ul style="list-style-type: none"> ● ERC token contracts are used for transactions ● It does not enforce either royalty or marketplace fees ● A malicious seller can: <ul style="list-style-type: none"> – Evades payment either by transferring the NFT to the buyer – Settling the payment off the platform
3	Post-sales modification	<ul style="list-style-type: none"> ● Creator modifies the royalty amount after effecting the primary sale ● Calculation of royalty is based on price listing ● A malicious creator can abuse by trapping the buyer which is achieved by fixing a low royalty and later on escalating the price post-sales. In such a situation, the buyer may end up paying more royalty than what was advertised initially by mere oversight

10.3 Issues with Respect to Fraudulent User Behaviors

DappRaddaris [22] is a trusted industry source that projects accurate insights with high quality on decentralized applications to a global audience. Global prosumers are discovering dapps and managing their NFT/and DeFi portfolios with DappRadar. Decentralized Finance (DeFi) is an emerging financial technology that is based on secure distributed ledgers comparable to those used by cryptocurrencies. The control from intermediaries such as banks and institutions on money, financial products, and financial services [23] is removed from the system.

DappRadar in their recent report titled Blockchain user behavior report from NFT perspective has given a detailed analysis on the user behavior from NFT perspective. They have analyzed the global aspects and trends such as geography and devices used for interaction. Further, their report has shown how DeFi users relate with NFT users interchangeably. The key findings of their analysis are summarized in Table 8.

There can be other attacks due to the user behavior such as digital scarcity, NFT giveaways, and front running attacks.

10.3.1 Counterfeit NFT Creation

Smart contracts validate the authenticity of an NFT and manage the collection. Verification is an essential practice that needs to be ensured when trading and checking whether the token one is buying is legitimate or not. Lack of awareness of the buyers

Table 8 Country-wise user behavior in investing in NFT and blockchain

User behavior	User's behavior
Interest in investing in NFT	User's in countries like the United States of America, Indonesia, Russia, and India show their interest gauged by NFT
Interest in investing in DeFi	Uses express their interest gauge by DeFi in the United States of America, the extent be maximum. This rank in this regard is followed by countries like Brazil, Thailand, Russia, and China
Interest in gaming dapps	In this regard, the maximum number of uses country-wise in hierarchical order is the Philippines India Brazil Venezuela and the United States of America
Interest in investing in the mobile market	There exists a difference in this regard between mobile users and desktop uses in May 2021 it was found that 53% of the users so true of the mobile while on the other hand 46% of the users surf true their desktops
DeFi Verus NFT user interaction	It is interesting to note that among users 85% of them have used NFT there has been an inter ever-increasing trend in this regard
NFT collections	There has been ever-increasing trend in this regard
Whale concentration index (WCI)	It together with the user's ID ratio are the metrics which are avidly adopted by the users there exists a robust negative correlation among these metrics WCI confirms the details of the total numbers that are being processed

pertaining to the existence of counterfeit NFTs. Also, the users have limited knowledge about the verification process with respect to authentication. Malicious users make it possible to offer counterfeit NFTs. There can be two types of counterfeiting. One is a similar collection name and the other is doxing. In similar collection names, counterfeiting can happen where an NFT can use the name of a collection or individual piece of object. In doxxing (doxing) is the act of publicizing the identity and or personal details of a person. It is a form of cyberbullying. In the NFT space, some people choose to communicate without showing their faces lest using their real names. Due to this, there is an increase in the number of scams and rug pulls. Crypto developers abandon a project and run away with investors' funds [24]. This is called *Rug pull* which is a malicious manipulation in the cryptocurrency industry. The people could involuntarily get doxxed in NFTs thereby revealing private personal information of an individual or an organization.

10.3.2 Trading Malpractice

Forbidden trading practices are made possible by malicious actors. Illicit practices include wash trading, shill bidding, and bid shielding, and these are summarized in Table 9.

Table 9 Illicit trading practices

Illicit NFT trading practices	Description of Illicit NFT trading
Wash trading	In wash trading, the buyer and seller connive with each other in order to increase the volume of trading of an asset through fraudulent activities
Shill bidding	In this common auction fraud, the bidders collude with each other and artificially hike the price thereby deceive the honest and making them forcibly pay the higher amount through spurious bids
Bid shielding	Malicious bidders collude with each other by guarding the low bid and thereby dissuade the legitimate bidder from further bidding

10.4 Issues with Respect to the Privacy

There are many legal issues about different aspects of NFTs. Only relatively a little has been talked about with respect to privacy laws. NFTs differ from traditional assets. The transactions in the real world where payment is made in currency and ownership is recorded in a public ledger. Examples include—shares, real estates, art works, and cultural assets or in a non-public record such as a customer database. In NFT trading, privacy risks [25] do not arise as they are stored and transacted in blockchain and as crypto assets. They are not directly associated with an individual’s personal identity. NFTs are acquired using cryptocurrencies and they are stored on a decentralized network. It is through a personal identity which is essentially an anonymous wallet. Current privacy legislation was not drafted with blockchain and Web 3.0 in mind. There exists a lacuna in the legal framework relating to crypto assets and Web 3.0 in owning, acquiring, and disposing of NFTs which could give rise to privacy risks in Web 3.0 and also on the following:

- Online identifiers.
- Avatars.
- Addresses of Blockchain.
- Activities in the Transactions.
- Location information.

11 Future Scope of NFT in the Context of Metaverse

NFTs allow artists to monetize their artwork and protect their copyright. NFTs also allow artists to receive royalties every time their creations change hands via smart contracts. Additionally, NFT creates huge earning opportunities for artists and the things like Metaverse are creating lucrative career paths for many digital artists. The Metaverse predicted as the future of social media and digital space is a shared virtual space for all types of digital activities. In general, a variety of techniques, such as augmented reality and the Internet, are used to construct the virtual world.

The concept is from recent decades and has a great future due to blockchain's rapid expansion. While the metaverse is still in the early stages of development, many businesses are already dabbling in the digital realm and NFTs will play a crucial role in the ecosystem of the metaverse [26]. It has great potential for career opportunities for designers, programmers, and information security professionals. NFTs are also important in the metaverse for creating exclusive environments and enhancing the digital communities and social experiences.

12 Conclusion

In this paper, the overview of emerging NFTs prevailing in the blockchain market has been studied. Firstly, the origin and evolution of NFT have been listed. Key technical components driving the working of NFT are discussed. Protocols for the establishment of NFT are focused and prevailing standards are also discussed and compared along with various properties. This paper has further discussed various issues and challenges with respect to cyber security and privacy are discussed. Finally, the study also focuses on the future scope of NFTs in the context of 'Metaverse'.

References

1. <https://www.investopedia.com/non-fungible-tokens-nft-5115211> [KB1]. Accessed on 09 Sep 2022
2. <https://www.cryptokitties.co/>. Accessed on 06 Sep 2022
3. Financial Times. NFTs: Identity in the metaverse. <https://www.ft.com/partnercontent/crypto-com/nfts-identity-in-the-metaverse.html>. Accessed on 10 Sep 2022
4. <https://www.zenofineart.com/blogs/news/the-beginning-of-nfts-a-brief-history-of-nft-art>. Accessed on 10 Mar 2022
5. <https://www.digitaltrends.com/computing/what-are-nfts-non-fungible-tokens-historyexplained/>. Accessed on 04 Mar 2022
6. Qin W et al (2021) Non-Fungible Token (NFT): Overview, Evaluation, Opportunities and Challenges (Tech ReportV2)
7. Cryptopunks. (2021) Accessible: <https://www.larvalabs.com/cryptopunks>. Accessed on 11 Sep 2022
8. Loot contract code (2021). <https://etherscan.io/address/0xff9c1b15b16263c61d017ee9f65c50e4ae0113d7#code>. Accessed on 10 Aug 2022
9. Fabian V, Vitalik B (2015) Eip-20: Erc-20 token standard. Accessible: <https://eips.ethereum.org/EIPS/eip-20>. Accessed on 11 Sep 2022
10. William E, Dieter S, Jacob E, Nastassia S (2018) Eip-721: Erc-721 non-fungible token standard. Accessible: <https://eips.ethereum.org/EIPS/eip-721>. Accessed on 15 Sep 2022
11. Witek R et al (2018) Eip-1155: Erc-1155 multi token standard. Accessible: <https://eips.ethereum.org/EIPS/eip-1155>. Accessed on 20 Sep 2022
12. Buterin V et al (2014) A next-generation smart contract and decentralized application platform. White Paper 3(37):2
13. <https://www.c-sharpcorner.com/article/understanding-the-workflow-of-nft-systems/>. Accessed on 12 Sep 2022

14. Zhou J, Gollman D (1996) A fair non-repudiation protocol. In: Proceedings 1996 IEEE symposium on security and privacy. pp 55–61. IEEE
15. Menezes AJ, Van Oorschot PC, Vanstone SA (2018) Handbook of applied cryptography. CRC Press
16. Moore D, Voelker G, Savage S (2006) Inferring internet denial-of-service activity. *ACM Trans Comput Syst* 24:115–139
17. Shostack A (2008) Experiences threat modeling at microsoft. MODSEC@ MoDELS 2008
18. Wang Q, Li R (2021) A weak consensus algorithm and its application to high performance blockchain. In: IEEE INFOCOM 2021-IEEE Conference on Computer Communications (INFOCOM). IEEE
19. Hackers stole NFTs from nifty gateway users. <https://www.theverge.com/2021/3/15/22331818/nifty-gateway-hack-steal-NFTs-credit-card>. Accessed on 18 Sep 2022
20. <https://www.fool.com/investing/stock-market/marketsectors/financials/cryptocurrency-stocks/crypto-tokens/>. Accessed on 09 Sep 2022
21. <https://cyberscrilla.com/nft-smart-contracts-explained>. Accessed on 17 Sep 2022
22. <https://dappradar.com/blog/blockchain-user-behavior-report-nftperspective#Key%20Findings>. Accessed on 20 Sep 2022
23. [https://www.investopedia.com/decentralized-finance-defi-5113835~:text=Decentralized%20finance%20\(DeFi\)%20is%20an,financial%20products%2C%20and%20financial%20services](https://www.investopedia.com/decentralized-finance-defi-5113835~:text=Decentralized%20finance%20(DeFi)%20is%20an,financial%20products%2C%20and%20financial%20services). Accessed on 22 Sep 2022
24. <https://coinmarketcap.com/alexandria/glossary/rug-pull>. Accessed on 24 Sep 2022
25. <https://www.jdsupra.com/legalnews/nfts-privacy-issues-for-consideration-7804114/>. Accessed on 24 Sep 2022
26. <https://101blockchains.com/nfts-and-metaverse/>. Accessed on 28/01/2022

The Rise of Public Wi-Fi and Threats



Prateek Bheevgade, Chirantan Saha, Rahul Nath, Siddharth Dabhade, Haresh Barot, and S. O. Junare

Abstract The rise of public Wi-Fi is increasing daily, and as it expands, it comes with many new opportunities and challenges. In this paper, you will find out why most intruders use public Wi-Fi to conduct cyber-criminal activities and how intruders can easily access your data or download your data as you have connected through malicious public Wi-Fi networks. In this research work, a survey has been done to determine why people prefer public Wi-Fi networks or private Wi-Fi. Experimental work focuses on intruders using ATHEROS (Hardware) and Wireshark & N-Map (Software) to extract data from open public Wi-Fi. At the same time, using public Wi-Fi, how to secure your data and its safety measurements and tips are given.

Keywords ATHEROS · Wireshark · Data privacy · Public Wi-Fi · Private Wi-Fi · Threats · Malicious attack

1 Introduction

As we all know, public Wi-Fi is increasing daily; we see this open public Wi-Fi in Airports, Amusement Parks, Coffee shops, shopping malls, and many other locations. Most people prefer free internet, which allows them to download resources like movies, watch YouTube videos, Instagram Reels, and most social media activities. The internet policy has allowed everyone to access the internet for ethical and reasonable purposes. But in today's world, every country and all cities have public Wi-Fi in their area location. According to internet policy, most people use the internet for unlawful activities, which may damage the personal security of any person connected to the networks. Intruders mostly use tools and software like Kali Linux, which comes under the devices that allow Intruders to gain resources over Wi-Fi. Intruders use "ATHEROS" and N-Map to gain access.

P. Bheevgade · C. Saha · R. Nath · S. Dabhade (✉) · H. Barot · S. O. Junare
School of Management Studies, National Forensic Sciences University, Gujarat, India
e-mail: dabhade.siddharth@gmail.com

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
S. J. Patel et al. (eds.), *Information Security, Privacy and Digital Forensics*,
Lecture Notes in Electrical Engineering 1075,
https://doi.org/10.1007/978-981-99-5091-1_13

175

Atheros is a developer of semiconductor chips for network interchanges, especially remote chipsets. N-Map (“Network Mapper”) is a free, open-source utility for network revelation and security auditing. Numerous frameworks and organization directors likewise find it helpful for inventory, overseeing administration overhaul schedules, and observing host or administration uptime.

Some threats regarding the virtually open Wi-Fi hotspots include sensitive data leaks and malware contaminations. In this situation, intruders behind public Wi-Fi networks can catch the information when you use it for productive purposes. In the following problem, cybercriminals can divert the user utilizing the web into a fake website instead of a simple one, fooling them and downloading malware inside their systems.

Those not aware of the most effective method to involve public Wi-Fi securely could end up in a difficult situation. To avoid this, you ought to constantly stay on high alert against the accompanying risks of open wireless networks.

- Identity theft: at the point when somebody utilizes someone else’s private distinguishing data without taking their anxiety, known as Identity Theft.
- Data breach: A data breach is an occurrence wherein data is taken or taken from a framework without the information or approval of the framework’s proprietor.
- Malware infection: Malware, or malicious software, is a program or file which makes computer, network, or server vulnerable.
- Packet sniffing or eavesdropping: The act of stealing data packets across the computer network is called packet sniffing.
- Man-in-the-middle attack: A man-in-the-middle (MITM) attack is a general term for when a culprit positions himself in a discussion between a client and an application.

2 Related Work

2.1 Global Scenario

As per statista.com, the worldwide public Wi-Fi areas of interest Conjectures (2016–22) propose that by 2019 there will be assessed of 362 million public Wi-Fi areas of interest accessible around the world. This figure would address a close to quadrupling of public Wi-Fi areas of interest beginning around 2016, showing the fast ascent in these organizations around the world. This pattern toward development is supposed to go on into basically the mid-2020s.

2.2 Literature Review

Smart Wireless networks are assisted by technologies such as Zigbee and SmartWAZ, which can be used for Wi-Fi and Zigbee Interfaces [1]. Information leakage [2] is a bigger challenge in today's world. The limitless number of people using Wi-Fi networks in public areas comes into contact with threats like War-Driving attacks, which result in sensitive information leakage. WIFI's trustworthiness is developing as its usage spreads to public areas and individual home networks. Despite broad information on potential weaknesses connected with public Wi-Fi, many individuals still connect with these unfamiliar networks [3]. Furthermore, 60% of the members knew nothing about any dangers related to utilizing an untrusted network and didn't care about their privacy [4]. Several Metropolitan cities [5] of India, like Delhi, Kolkata, Mumbai, and Chennai, have public Wi-Fi services in bus stations, airports, railway stations, etc., where most people use their devices like smartphones, tablets, laptops, etc., which contains some sensitive and personal information [6] of them. In these areas, people use the internet either for entertainment, for business purposes, or for some kind of social interaction [7] among other users. To examine areas of interest privacy policies and terms-of-purpose documentation, this unveils the specialist organization's information and privacy practices [8] Iso, they accept the user's policy [9] shown while connecting the public Wi-Fi networks. While using public networks. But the users are unaware of getting attacked by attackers or intruders such as phishing mail, malware, adbots, etc., which affects their authentication [10] of their devices.

Hidden Dangers

The infinite-expanding growth of public hotspots in public areas is making various customers unaware of the upcoming dangers they will face. The public networks are handled mainly by intruders who can target the users and try to capture their sensitive information from their systems [11].

Smartphone Security Practices

The world of smartphones has evolved over the past centuries, from phones where the users are only allowed to call to smartphones where the user can store documents, accounts, images, various applications, etc. The increasing amount of sensitive personal data on smartphones has made intruders an attraction to steal data. Most users connect the public networks, unaware that intruders might be watching their activities [12].

Public Wi-Fi During Travel

There are scenarios where people use public Wi-Fi while they are traveling. These networks are available on trains, buses, airplanes, etc. While they are using the web, there are chances that the intruders can also be going with them. The intruder can make Wi-Fi connections with the same name but in different locations. Once the user visits the area, there is a chance that the user's personal information can be collected without knowing them. This way, the user's IP addresses are more vulnerable, which can increase potential threats [13].

Location Privacy

When users use public Wi-Fi while traveling, they automatically connect with several public Wi-Fi but with the same name. This can make the user's IP address vulnerable, which can help intruders access the victim's devices. This is a kind of trap where several users get trapped by intruders and help them get the victim's GPS location, making the location's privacy very low [14].

Secure Authentication

Wi-Fi has become a necessity of the new era, and everyone wants to use it. Most people, including business people, use public Wi-Fi for business or office work, like meetings, checking emails, etc. They are using their login credentials to obtain access to their accounts, which led the attackers to steal their credential information while the users are accessing public Wi-Fi. This causes a massive drop in their authentication [15].

Self-Protective Behaviors

The use of public Wi-Fi is good, but not for every purpose. People using public Wi-Fi, especially in public areas, should be aware of what to do and what not to do while using public Wi-Fi. They must know of the risks while using sensitive credentials on public Wi-Fi. The users should use VPN to keep them safe as VPNs change the Mac address and location of the devices [16].

3 Methodology

Our research consists of *two main categories*. The *first part* investigates the public motive for using public Wi-Fi. Also, we provide the most possible and secure public Wi-Fi architecture for the industry. The *second part* consists of the Intruder or Fraudster's motive to create a Public open Wi-Fi.

3.1 The Public Motive for Using Public Wi-Fi [17]

- **Free**

Most open Wi-Fi networks are accessible free of charge. In this manner, you will want to get a good deal on information charges and do all your web-related tasks quickly on your singular cell phone at a sufficiently lovely speed. Contrasted and a Wired network connection, wireless network organization offers essential advantages regarding cost and work. Especially while presenting another Wi-Fi network, you can cleave down the costs in wiring and support.

- **Productivity**

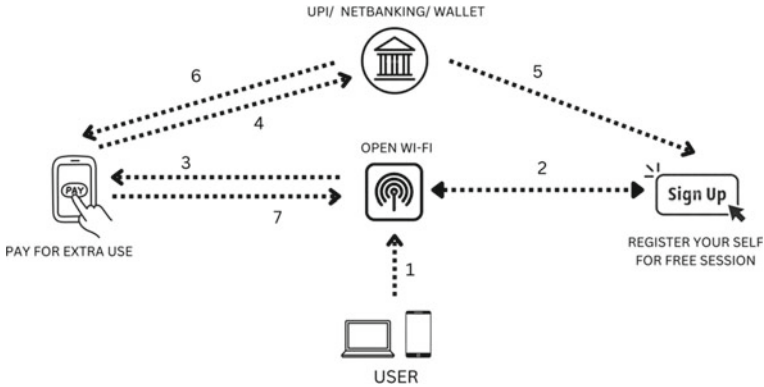


Fig. 1 Block diagram of secure public Wi-Fi

Whatever maintains flexibility progresses productivity. Public Wi-Fi is conventionally open in generally business places, for instance, air terminals, malls, cafes, rail stations, emergency clinics, and other practically identical zones, to propel movability among clients and people in general. As per the point of view of associations, this is of most outrageous importance in light of the fact that their delegates will need to deal with their undertakings peacefully in a rush, given they have the contraption to get themselves connected with the public Wi-Fi organization. In this way, it could be said public Wi-Fi can emphatically and straightforwardly impact working environment efficiency. This significant potential gain of public Wi-Fi is unquestionably quite excellent.

- **Emergency**

You can't foresee the event of a crisis. It can happen anyplace, whenever, without earlier notification. Public Wi-Fi behaves like nothing else and can carry many advantages to managing what is happening. *For instance:* Through the guide of public Wi-Fi, individuals will want to illuminate their area and security status to the experts overall and their precious ones on the occasion of crises, stimulating up the hunt and salvage process (if any).

- **Benefits for Students**

This one's an obvious point, I presume. I had previously featured the reality that most open Wi-Fi administrations are accessible free of charge. Thus, understudies on a limited spending plan can benefit enormously from it. The web association has become pretty much a piece of the essential conveniences, particularly that of an understudy. A free Wi-Fi administration can generally help him more than anything (Fig. 1).

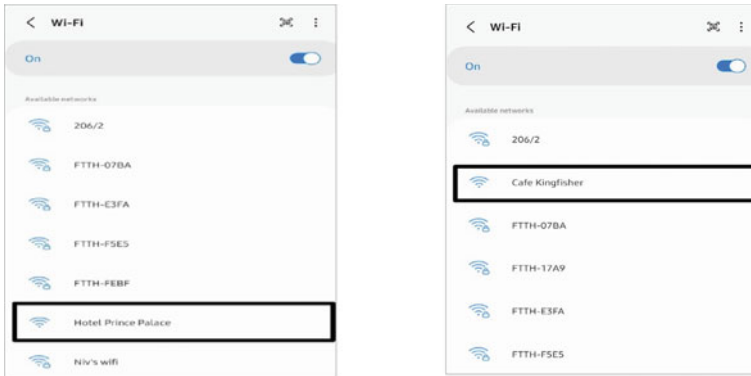


Fig. 2 Public Wi-Fi made by intruder

3.2 Intruder or Fraudster Motive to Create Public Wi-Fi [18]

Setup of Malicious Public Wi-Fi

Let's say the intruder has created a malicious Public Wi-Fi, which may affect those users who are connected to that Wi-Fi. (Refer Fig. 2) which tells that intruders can easily create a public Wi-Fi; no one can even guess which one is malicious.

Collecting All Sensitive Information of User Devices [18]

Once the malicious public Wi-Fi is created and as much traffic starts to join the open Wi-Fi networks, the intruder will begin executing the next step, which is collecting public user IP addresses and MAC address of particular devices from the users. The user only accesses your devices but also contains the personal information of the user and may try to ask for a ransom to delete the user's personal information.

In the scenario (Refer to Fig. 3), the user wants to access the website named <https://www.bank.org> using public Wi-Fi. But the user was unaware of an intruder inside the public Wi-Fi. Now, when the user types the website <https://www.bank.org>, the intruder interrupts the connection and manipulates the website name as <https://www.bunk.org>. The user is unaware of this manipulation and is redirected to the website that the intruder makes. The intruder's website is the same as the original website, with the same user interface and some functionality that helps trick the user. The user tries logging into the malicious webpage using their credentials. After redirecting, the user gets a "404 error" (Access Error/Server Error). But at this point, the intruder gets the user's credentials and can use them for their intention (Figs. 4, 5, 6, 7, 8, 9).

Analysis and Collecting Data [18].

Here, we have selected a Particular MAC Address BSSID to view details.

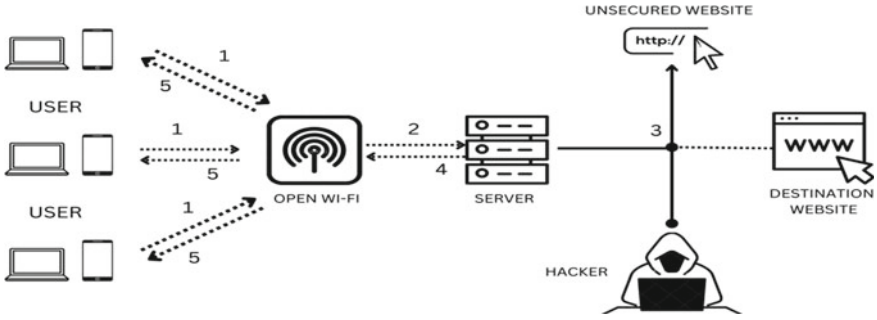


Fig. 3 Block diagram of intruder or fraudster to create public Wi-Fi [19, 20]

```

root@Prateek: ~# airodump-ng wlan0mon
CH 3 ][ Elapsed: 42 s ][ 2022-09-12 15:23

BSSID          PWR  Beacons  #Data, #/s  CH  MB  ENC  CIPHER  AUTH  ESSID
4C:AE:1C:71:FE:BF -66    74        3    0  9  270  WPA2  CCMP   PSK   FTTH-FEBF
4C:AE:1C:71:F5:E5 -47    41       107    0  4  270  WPA2  CCMP   PSK   FTTH-F5E5
4C:AE:1C:71:17:A9 -54    34        0    0  11 270  WPA2  CCMP   PSK   FTTH-17A9
4C:AE:1C:71:FE:F6 -56    27        1    0  11 270  WPA2  CCMP   PSK   Niv's wifi
4C:AE:1C:71:FF:9B -67    16        3    0  1  270  WPA2  CCMP   PSK   Darshan's wifi
4C:AE:1C:72:08:07 -61    29        0    0  1  544  WPA2  TKIP   PSK   206/2
4C:AE:1C:71:E3:FA -70    23        0    0  7  270  WPA2  CCMP   PSK   FTTH-E3FA
4C:AE:1C:6F:D3:CE -66    23        0    0  1  270  WPA2  CCMP   PSK   211-D
4C:AE:1C:72:0A:13 -68    18        0    0  11 270  WPA2  CCMP   PSK   FTTH-0A38
4C:AE:1C:71:E6:F1 -72    24        0    0  7  270  WPA2  CCMP   PSK   Dhruv's wifi
4C:AE:1C:72:07:DB -76    7         0    0  9  270  WPA2  CCMP   PSK   FTTH-07DB
4C:AE:1C:71:88:B6 -76    29        0    0  11 270  WPA2  CCMP   PSK   FTTH-88B6
4C:AE:1C:72:0B:07 -77    17        0    0  4  270  WPA2  CCMP   PSK   211
4C:AE:1C:6F:A1:A6 -78    5         0    0  8  270  WPA2  CCMP   PSK   FTTH-A1A6
4C:AE:1C:71:C9:CF -80    5         0    0  2  270  WPA2  CCMP   PSK   FTTH-C9CF
4C:AE:1C:72:07:BA -81    5         0    0  1  270  WPA2  CCMP   PSK   FTTH-07BA
4C:AE:1C:70:5E:82 -82    7         0    0  1  270  WPA2  CCMP   PSK   FTTH-5E82
4C:AE:1C:72:1F:BC -85    3         0    0  4  544  WPA2  TKIP   PSK   110/2
4C:AE:1C:71:E3:29 -85    4         0    0  1  270  WPA2  CCMP   PSK   ANUFOREVER
4C:AE:1C:71:F6:A0 -86    4         0    0  5  270  WPA2  CCMP   PSK   Room101
4C:AE:1C:72:09:BF -88    3         0    0  1  270  WPA2  CCMP   PSK   FTTH-09BF
4C:AE:1C:71:1E:05 -88    2         0    0  11 270  WPA2  CCMP   PSK   FTTH-E505
4C:AE:1C:72:07:C5 -89    1        10    0  9  270  WPA2  CCMP   PSK   FTTH-07C5
54:47:E8:39:2E:D9 -90    3         0    0  6  130  WPA2  CCMP   PSK   Cranberry-2.4G
4C:AE:1C:50:A2:BC -90    3         0    0  7  270  WPA2  CCMP   PSK   FTTH-42BC
4C:AE:1C:71:1E:BA -90    4         0    0  6  270  WPA2  CCMP   PSK   FTTH-FEB4
4C:AE:1C:72:08:5F -91    5         0    0  11 270  WPA2  CCMP   PSK   FTTH-085F
4C:AE:1C:71:C8:0C -93    7         0    0  6  270  WPA2  CCMP   PSK   FTTH-C80C
4C:AE:1C:71:E5:B2 -89    2         0    0  1  270  WPA2  CCMP   PSK   FTTH-E5B2

BSSID          STATION          PWR  Rate  Lost  Frames  Notes  Probes
(not associated) 9C:29:76:F0:D5:A7 -80    0 - 1    0    1
(not associated) 7E:C5:99:A1:41:60 -66    0 - 1    0    1
(not associated) 32:FE:CD:AA:9D:80 -82    0 - 1    0    1
4C:AE:1C:71:FE:BF BA:18:D5:C7:22:C5 -42    0 - 1    0   230
4C:AE:1C:71:FE:BF F2:0A:87:0B:8B:BF -56    0 - 1    0    3
4C:AE:1C:71:F5:E5 E4:9C:FD:80:C5:F1 -71    240-240 319  164
4C:AE:1C:71:FE:F6 C8:94:02:39:70:07 -81    0 - 1    0    3
4C:AE:1C:71:FF:9B C4:23:60:C9:2F:B1 -83    50- 60  46
4C:AE:1C:71:E3:FA EE:61:C5:0D:C6:68 -83    0 - 1    0    4
4C:AE:1C:71:FE:BA 8B:E7:DA:6D:1B:77 -76    0 - 1    0    1
4C:AE:1C:71:E6:F1 0E:7A:07:25:7F:02 -73    0 - 60  24
4C:AE:1C:71:88:B6 D2:59:2E:86:92:ED -86    0 - 1    0    4
4C:AE:1C:70:5E:82 0A:95:C1:00:99:FA -82    0 - 1   146  5
quitting...

```

Fig. 4 Collecting all the Wi-Fi network user details

3.3 Analysis of Collected Data Through Wire-Shark [19, 21]

Wireshark is the world's premier and most broadly utilized network convention analyzer. Wireshark catches the information coming or going through the NICs on its gadget utilizing a hidden parcel catch library. As a matter of course, Wire-shark catches on-gadget information just, however it can catch practically every one

```
CH 2 ][ Elapsed: 3 mins ][ 2022-09-12 15:30
BSSID          PWR RXQ Beacons #Data, #/s CH MB ENC CIPHER AUTH ESSID
4C:AE:1C:71:FE:BF -70 18 578 17 0 9 270 WPA2 CCMP PSK FTTH-FEBF
BSSID          STATION          PWR Rate Lost Frames Notes Probes
4C:AE:1C:71:FE:BF 62:E6:76:DC:68:FC -1 1e-0 0 3
4C:AE:1C:71:FE:BF 38:68:93:60:92:80 -1 1e-0 0 2
```

Fig. 5 Getting details of particular I.P. or MAC address

```
(root@Prateek)-[~]
# ls
capture-01.cap          capture-02.kismet.csv  capture-03.log.csv    capture-05.csv
capture-01.csv         capture-02.kismet.netxml capture-04.cap        capture-05.kismet.csv
capture-01.kismet.csv  capture-02.log.csv    capture-04.csv       capture-05.kismet.netxml
capture-01.kismet.netxml capture-03.cap        capture-04.kismet.csv capture-05.log.csv
capture-01.log.csv     capture-03.csv       capture-04.kismet.netxml
capture-02.cap        capture-03.kismet.csv capture-04.log.csv
capture-02.csv        capture-03.kismet.netxml capture-05.cap
```

Fig. 6 All the files collected from the user device I.P. address

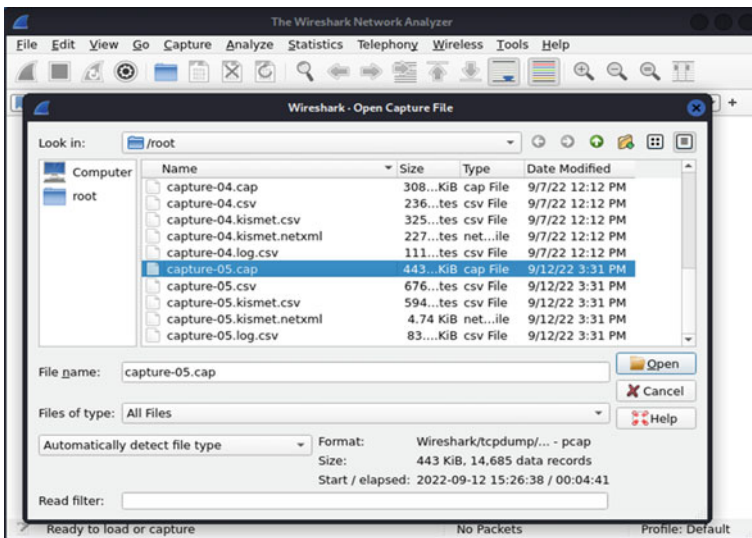


Fig. 7 Using wire-shark capture collected files to view details information [19]

of the information on its LAN whenever run-in unbridled mode. Presently, Wireshark utilizes NMAP's Parcel Catch library (called Npcap), the libpcap library of the Windows version. Wireshark uses this library to capture live network data on windows.

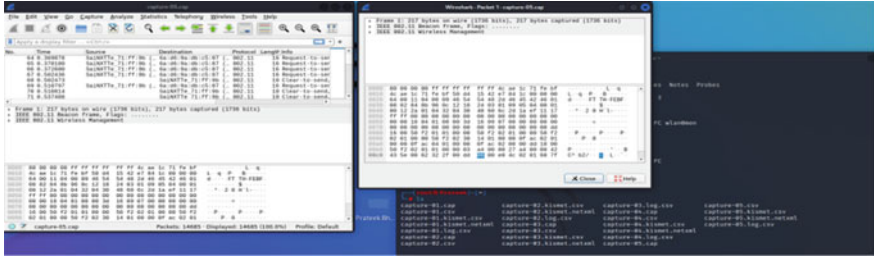


Fig. 8 Showing captured data packets which collected through the user I.P. and MAC address [19]

```

0000  3c 3f 78 6d 6c 20 76 65 72 73 69 6f 6e 3d 22 31 <?xml ve rsion="1
0010  2e 30 22 20 65 6e 63 6f 64 69 6e 67 3d 22 49 53 0" enco ding="15
0020  4f 2d 38 38 35 39 2d 31 22 3f 3e 0a 3c 21 44 4f 0-8859-1 "> <!D
0030  43 54 59 50 45 20 64 65 74 65 63 74 69 6f 6e 2d CTYP E de ction-
0040  72 75 6e 20 53 59 53 54 45 4d 20 22 68 74 74 70 run SYST EM "http
0050  3a 2f 2f 6b 69 73 6d 65 74 77 69 72 65 6c 65 73 :/kismet wirele s
0060  73 2e 6e 65 74 2f 6b 69 73 6d 65 74 2d 33 2e 31 s.net/ki smet-3.1
0070  2e 30 2e 64 74 64 22 3e 0a 0a 3c 64 65 74 65 63 .0.dtd"> <detec
0080  74 69 6f 6e 2d 72 75 6e 20 6b 69 73 6d 65 74 2d tion-run kismet-
0090  76 65 72 73 69 6f 6e 3d 22 61 69 72 6f 64 75 6d version="airodum
00a0  70 2d 6e 67 2d 31 2e 30 22 20 73 74 61 72 74 2d p-ng-1.0 " start-
00b0  74 69 6d 65 3d 22 4d 6f 6e 20 53 65 70 20 31 32 time="Mo n Sep 12
00c0  20 31 35 3a 32 36 3a 33 38 20 32 30 32 32 22 3e 15:26:3 8 2022">
00d0  0a 0a 09 3c 77 69 72 65 6c 65 73 73 2d 6e 65 74 . . .swire less-net
00e0  77 6f 72 6b 20 6e 75 6d 62 65 72 3d 22 31 22 20 work num ber="1"
00f0  74 79 70 65 3d 22 69 6e 66 72 61 73 74 72 75 63 type="in frastruc
0100  74 75 72 65 22 20 66 69 72 73 74 04 74 69 6d 65 ture" fi rst-time
0110  3d 22 4d 6f 6e 20 53 65 70 20 31 32 20 31 35 3a ="Mon Se p 12 15:

```

Fig. 9 ASCII values of particular captured data

4 Output Analysis of Public Wi-Fi [Survey] [21]

In this part, we depict the consequence of breaking down the gathered information from individuals associating their devices to the internet through open Wi-Fi, confidential WIFI, or portable internet using a survey. We put together the analysis of the study into three segments: in the initial segment, we describe the result of the study of people giving their opinions on using the internet from various states of India; in the second segment, we describe the preference of internet among the users, and the last part describes the purpose of using public Wi-Fi in public areas.

4.1 Survey of People Giving Their Opinions on Using the Internet Across States of India [21]

In this analysis, we gained the public intention of using the kind of internet (public Wi-Fi, private Wi-Fi, or their internet) across various states of India. We gathered responses using the online survey distributed among people residing in different states of India, and the survey result depends upon the usage of the type of internet whenever they are in public.

The above graph (Refer to Fig. 10) shows the percentage of people giving feedback in the survey. We gathered information from seventeen states, and the ratio

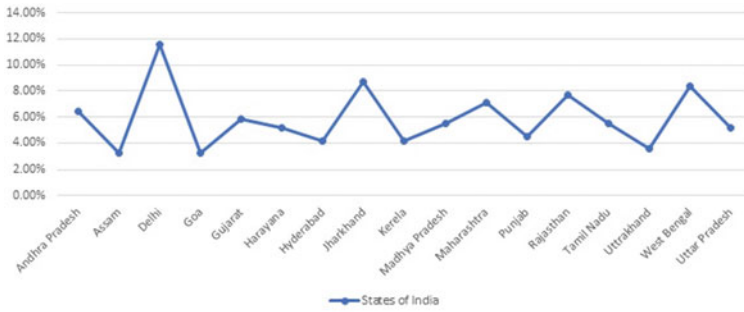


Fig. 10 Survey of people giving their opinions on using the internet across states of India

indicates the number of individuals who filled out the survey. The graph shows that the maximum response came from states like Delhi (around 12.0%), and the minor response was gathered from states like Assam and Goa (about 3.0%).

4.2 Priority of Users Using Internet [12]

In this analysis, we gathered the number of users using which type of internet in public areas from the survey. We categorized them into public Wi-Fi, private Wi-Fi, and their own internet.

The (Refer Fig. 11) pie chart above shows the level of individuals using public Wi-Fi, private Wi-Fi, and their internet. From the result of our survey, 17.15% of people prefer using public Wi-Fi, 39.28% prefer using private Wi-Fi, and 43.57% prefer using their mobile internet in public areas.

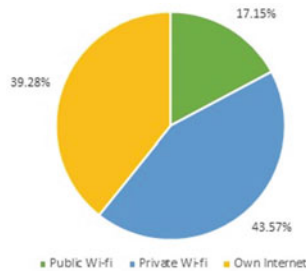


Fig. 11 Pie-chart regarding the survey of priority of users using the internet

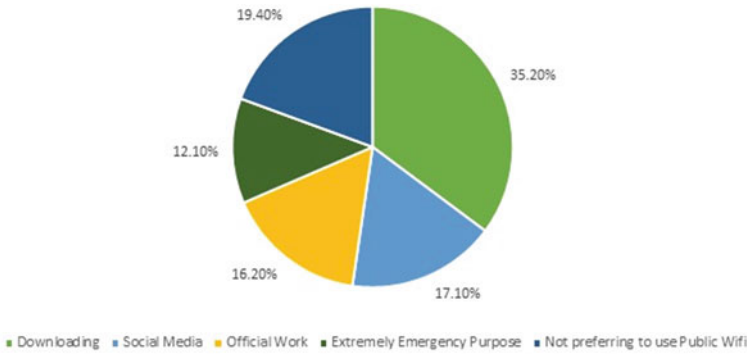


Fig. 12 Pie-chart regarding the purpose of using Wi-Fi in public areas

4.3 Purpose of Using Wi-Fi in Public Areas

In this analysis, we gathered the number of users using public Wi-Fi in public areas for some purpose from the survey.

The pie chart (Refer to Fig. 12) above shows the percentage of people consuming public Wi-Fi in public areas for some or other purpose. 35.20% of people use public Wi-Fi for downloading stuff, 17.10% of people use it for social media, 16.20% of people use it for office work, and 12.10% of people use it for extremely emergency purposes like using Maps in case they lost the path, booking cabs, etc. 19.40% of people don't prefer to use public Wi-Fi in public areas.

5 Rise of Public Wi-Fi and Its Threats [9, 10, 15]

The rising network of public wireless access points and the ascent of wireless computing devices like laptops, tablets, and mobiles have made it more straightforward for individuals to get data on the internet. Such networks have adjusted to the expected security and protection chances. There are a few dangers that mean being unsafe while involving public Wi-Fi here.

Identity Theft

Perhaps the most serious and regular danger concerns the robbery of individual data. Personal information occurs in many formats:

- Login credentials.
- Financial and business information.
- Personal data like date of birth and government authority number.
- Pictures, videos, documents, etc.

When intruders assemble someone's information, they can get access to your pages and cause harm to their funds and notoriety. On the off chance that they don't get full admittance to the information on devices, they might, in any case, steal data sharing over the networks [22].

Rogue Wi-Fi Networks

The client could be tricked into using a renegade Wi-Fi set up by an intruder. That organization named "Free Wi-Fi" may be set up unequivocally to harvest your sensitive data [23].

Man-In-The-Middle Attacks

Communicating with free open Wi-Fi allows you to have your sensitive and individual data hindered by outsiders. This is because intruders can arrange between your laborers utilizing the Wi-Fi and the connection point [24].

Eavesdropping

Anyone related to a comparative Wi-Fi network can listen to what clients send and get using a packet analyzer or sniffer. These instruments give the probability to see all that is sent across the Wi-Fi network if it wasn't encoded. Packet sniffers let intruders explore affiliation issues as well as execution issues with their remote networks to impede other clients' information and take anything from their devices [25].

Malware Distribution

Another threat that can occur while using public Wi-Fi is the implementation of malware on your device. Malware exists in many forms:

- Viruses.
- Worms.
- Trojan horses.
- Ransomware.
- Adware, etc.

Intruders on similar public Wi-Fi could plant malware on your device if it isn't safeguarded as expected. A suspect Wi-Fi supplier could utilize the hotspot to contaminate your devices with at least one of these threats [21, 26].

6 Safety Measures to Secure Your Data [16, 27, 28]

Some safety measures will help users to secure their data while connecting public networks.

- Switch off programmed availability highlights, including Bluetooth, before you sign in to public Wi-Fi. Switch off all highlights on your telephone, P.C., or tablet that will permit your device to interface with different gadgets or public remote networks.

- Search for HTTPS toward the start of a site address. This implies the association between the browser and the web server is encoded, so any information submitted to the site will be protected from listening in or altering. Most browsers likewise incorporate a padlock towards the start of the site's location to show on the off chance that the site utilizes encryption procedures.
- Input no private data on any sites while visiting public Wi-Fi. Indeed, even with HTTPS, it's as yet hazardous to enter confidential data while utilizing public Wi-Fi. Since numerous intruders are sitting behind the public Wi-Fi, they will figure out how to take out your data.
- Continuously forget the network after you utilize public Wi-Fi. Make a point to tap the "forget network" choice on your network preferences once you use public Wi-Fi. This will keep your device from interfacing with it again without your consent.
- Limit Airdrop and Record Sharing when you're on a public network. You'll need to remove the highlights that empower frictionless document sharing on your devices.
- Utilize a VPN (Virtual Private Network) as it is the most dependable choice to ride on open networks. It is one of the most valuable tools to assist people in keeping their data secure when signed on to public networks.

The data which was acquired from the survey gave an assumption that public Wi-Fi is not safe for any personal use. Most study participants preferred private Wi-Fi/own network over public Wi-Fi in public places. But some people connect to public Wi-Fi for entertainment purposes. So they are still unaware of the consequences and dangers of getting viruses or data theft via connecting to public Wi-Fi.

7 Conclusion

While doing a detailed analysis of open Wi-Fi, we have found that intruders can access your information effectively by utilizing different hardware and tools, making the intruder's work simpler. As well as we likewise finished an open Wi-Fi survey where we observed that there are 40% of individuals who inclined toward free web public Wi-Fi, and the rest of the 60% of individuals use their information as it is not secure wireless networks that will make the electronic transaction completely safe. We also studied public user motives and intruders' motives to create a malicious Wi-Fi portal so that the intruder will easily access your devices through MAC and IP addresses.

Consider more secure options in contrast to public Wi-Fi, similar to utilizing limitless information anticipate your smartphone or using your smartphone as a private hotspot. This way, even if an intruder obtains the user's credentials, such as username and password, they won't be able to access your accounts. The intruder can access public Wi-Fi like airports, amusement parks, coffee shops, shopping malls, and other locations. While working on research for public Wi-Fi, we have found how people like to use public Wi-Fi to complete their necessary needs like downloading or

surfing the internet. The research also concludes that intruders misuse public Wi-Fi as a phishing system to quickly gain access to the devices and extract their personal and sensitive information, like how we experimented on how Atheros was used to find the channels [25, 27].

References

1. Zhang B, Zuo J, Mao W (2019) SmartWAZ: design and implementation of a smart WiFi access system assisted by Zigbee. *IEEE Access* 7:31002–31009. <https://doi.org/10.1109/ACCESS.2019.2901051>
2. Elhamahmy ME, Sobh TS (2011) Preventing Information Leakage Caused by War Driving Attacks in Wi-Fi Networks. <https://doi.org/10.13140/RG.2.1.1494.9280>
3. Choi HS, Carpenter D (2013) Connecting to unfamiliar Wi-Fi hotspots - a risk taking perspective. In: *AMCIS 2013 proceedings*, 13. <https://aisel.aisnet.org/amcis2013/ISSecurity/RoundTablePresentations/13>. Access Date-November 2022
4. Ali S, Osman T, Mannan M, Youssef A (2019) On privacy risks of public WiFi captive portals. Workshop on data privacy management (DPM, co-located with ESORICS 2019), September 26–27, 2019, Luxembourg. arXiv version: July 3, 2019
5. Vidales P, Manecke A, Solarski M (2009) Metropolitan public WiFi access based on broadband sharing (invited paper). In: *Proceedings of the Mexican International Conference on Computer Science*. pp. 146–151. <https://doi.org/10.1109/ENC.2009.22>
6. Hammonds M, Muhammad J (2019) Does connecting to public Wi-Fi have an effect on your personal information and privacy, at *ADMI-2019*, April: 11–14, 2019, Memphis, TN
7. Hampton KN, Gupta N (2008) Community and social interaction in the wireless city: Wi-fi use in public and semi-public spaces. *New Media Soc* 10(6):831–850. <https://doi.org/10.1177/1461444808096247>
8. Ali SAA (2020) A large-scale evaluation of privacy practices of public Wifi captive portals. Thesis, Master of Applied Science in Information Systems Security, Concordia University Montréal, Québec, Canada
9. Spacey R, Muir A, Cooke L, Creaser C, Spezi V (2017) Filtering wireless (Wi-Fi) internet access in public places. *J Librariansh Inf Sci* 49(1):15–25. <https://doi.org/10.1177/0961000615590693>
10. Kindberg T, Bevan C, O'Neill E, Mitchell J, Grimmett J, Woodgate D (2009) Authenticating ubiquitous services: A study of wireless hotspot access. In: *ACM International Conference Proceeding Series*. pp 115–124. <https://doi.org/10.1145/1620545.1620565>
11. Shahin E (2017) Is WiFi worth it: the hidden dangers of public WiFi. *Catholic Univ J Law Technol* 25(1)7:205–230. Available at: <https://scholarship.law.edu/jlt/vol25/iss1/7>
12. Breitinger F, Tully-Doyle R, Hassenfeldt C (2020) A survey on smartphone user's security choices, awareness and education. *Comput Secur*. <https://doi.org/10.1016/j.cose.2019.101647>
13. Cheng N, Oscar Wang X, Cheng W, Mohapatra P, Seneviratne A (2013) Characterizing privacy leakage of public WiFi networks for users on travel. In: *Proceedings - IEEE INFOCOM*. pp 2769–2777. <https://doi.org/10.1109/INFCOM.2013.6567086>
14. Raghunath MT, Narayanaswami C, Narayanaswami C (2003) IBM research report a practical approach to location privacy in public WiFi networks a practical approach to location privacy in public WiFi networks
15. Watts S (2016) Secure authentication is the only solution for vulnerable public wifi. *Comput Fraud Secur* 2016(1):18–20. [https://doi.org/10.1016/S1361-3723\(16\)30009-4](https://doi.org/10.1016/S1361-3723(16)30009-4)
16. Maimon D, Becker M, Katz J (2017) Self-protective behaviors over public WiFi networks. *LASER 2017 HotCRP*. <https://www.consumer.ftc.gov>

17. Sombatruang N, Sasse MA, Baddeley M (2016) Why do people use unsecure public Wi-Fi? An investigation of behaviour and factors driving decisions. In: ACM International Conference Proceeding Series, vol Part F130652, pp 61–72. <https://doi.org/10.1145/3046055.3046058>
18. Baray E, Kumar Ojha N (2021) WLAN Security Protocols and WPA3 Security Approach Measurement through Aircrack-ng Technique. In: Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021, pp 23–30. <https://doi.org/10.1109/ICCMC51019.2021.9418230>
19. Suharyanto CE, Simanjuntak P (2017) Potential threat analysis hypertext transfer protocol and secure hypertext transfer protocol of public WiFi users (Batam Case). *Int J Sci Eng Res* 8(2):320–326. ISSN 2229–5518
20. Lotfy AY, Zaki AM, Abd-El-Hafeez T, Mahmoud TM (2021) Privacy issues of public Wi-Fi networks. pp 656–665. https://doi.org/10.1007/978-3-030-76346-6_58
21. Nazir R, Iaghari AA, Kumar K, David S, Ali M (2022) Survey on Wireless Network Security. *Archives of Computational Methods in Engineering*, vol 29, no 3. Springer Science and Business Media B.V., pp 1591–1610. <https://doi.org/10.1007/s11831-021-09631-5>
22. BJ Koops R Leenes 2006 Identity theft, identity fraud and/or identity-related crime *Datenschutz und Datensicherheit - DuD* 30 9 553 556 10.10f07/s11623-006-0141-2
23. Vijay BP, Pranit T, Swapnil DD (2012) Protecting Wi-Fi networks from rogue access points. In: Fourth International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom2012), Bangalore, India. 119–122. <https://doi.org/10.1049/cp.2012.2508>
24. Tareq M, Safeia A (2021) Man in the middle attack in wireless network. <https://www.researchgate.net/publication/356290033> Access on November 2022
25. Tuli R (2020) Packet sniffing and sniffing detection. *Int J Innova Eng Technol* 16(1):22–32. ISSN: 2319-1058. <https://doi.org/10.21172/ijiet.161.04>
26. Kim S (2020) Anatomy on malware distribution networks. *IEEE Access* 8:73919–73930. <https://doi.org/10.1109/ACCESS.2020.2985990>
27. McShane I, Gregory M, Wilson C (2016) Practicing safe public Wi-Fi: assessing and managing data-security risks (December 22, 2016). Available at SSRN: <https://ssrn.com/abstract=2895216> or <https://doi.org/10.2139/ssrn.2895216>
28. Maimon D, Howell CJ, Jacques S, Perkins RC (2022) Situational awareness and public Wi-Fi users' self-protective behaviors. *Secur J* 35(1):154–174. <https://doi.org/10.1057/s41284-020-00270-2>

Digital Forensic Investigation on Ponzi Schemes



Babu Madhavan and N. Kalabaskar

Abstract The Ponzi scheme is an economic offence to lure investors by giving false assurance of huge returns with less risk. The money is provided to the older investors from the new investors in the form of payouts. Mostly the scheme would collapse or in the verge of collapse when the reduction of new investors. The scheme would try to retain later or the latest customers to reinvest or roll over the payouts into the scheme. Ponzi schemes are also performing as “Smart Ponzi Scheme” [1] where new technology (blockchain) cryptocurrency has been used indirectly. Some Ponzi scheme adopts a hybrid model without the knowledge of investors. Fugazzi financial securities are also given in the form of bonds with some security values, especially payouts are enticed to roll over again. The Ponzi scheme would initially be started and pretend to be a formal and genuine financial business and would certainly know the scheme would collapse. The scheme’s brainchild or brainchildren would counter and prepare for the easy way outs. The accumulated money is invested in real estate, movable properties, Panama papers, gold investments, cryptocurrencies, smart contracts (Cryptocurrency), and offshore investments. The new and the latest investors were victimized a lot. The financial victimization will be more from the new investors to the initial or oldest investors. The fairness of identifying actual financial loss incurred by the investors has to be justified for a fair settlement. The nature of the Ponzi scheme itself is a discrete business and the brainchild/brainchildren behind the scheme have constructed business infrastructure in such a way that they cannot be caught or tracked or detected. They have chosen complex technological infrastructure to make the investigation process difficult. The Ponzi scheme is accomplished by complex infrastructure in a way digital forensics investigation process made so difficult to detect. Ponzi scheme identification and intelligence about people and infrastructure are to be collected properly else the break in the detection chain would end up in fragile evidence collection. Understanding of infrastructure of the Ponzi scheme model is crucial to gather all information and quantifying the actual amount and people who were involved in the Ponzi scheme. The magnitude

B. Madhavan (✉) · N. Kalabaskar

Department of Cyber Forensics and Information Security, University of Madras, Chennai, Tamil Nadu, India

e-mail: bob4u1985@gmail.com

of the Ponzi scheme scam would only be identified by the proper digital forensic investigation process. This paper discusses the complex infrastructure adopted by the Ponzi schemes. The hurdles and challenges faced by the investigation team and digital forensics Investigation team to detect the magnitude of the scam involved. This paper also addresses the lacuna of policy, enforcement, and regulatory lens on the Ponzi scheme with respect to the existing monitoring system and infrastructure.

Keywords Digital forensics · Economic offence · Ponzi scheme

1 Introduction

The Ponzi scheme is an economic offence to lure investors by giving false assurance of huge returns with less risk. The money is provided to the older investors from the new investors in the form of payouts. Mostly the scheme would collapse or in the verge of collapse when to the reduction of new investors. The scheme would try to retain later or the latest customers to reinvest or roll over the payouts into the scheme.

Ponzi schemes are also performing as “Smart Ponzi Scheme” [1] where new technology (blockchain) cryptocurrency has been used indirectly. Some Ponzi scheme adopts a hybrid model without the knowledge of investors. Fugazzi financial securities are also given in the form of bonds with some security values especially payouts are enticed to roll over again.

The Ponzi scheme would initially be started and pretend to be a formal and genuine financial business and would certainly know the scheme would collapse. The scheme’s brainchild or brainchildren would counter and prepare for the easy way outs.

The accumulated money is invested in real estate, movable properties, Panama papers, gold investments, cryptocurrencies, and smart contracts (cryptocurrency).

The complex infrastructure of the Ponzi scheme is to be unraveled and busted with help of digital forensic investigation. Understanding of Ponzi scheme model and gaining actual information pertaining to the scheme to arrive real magnitude of the scam. The real magnitude of investors and money would certainly help fair disperse among invested victims.

1.1 History

The fraudulent scheme was named after “Charles Ponzi” [2, 3] who ran schemes such as “Postal reply coupon” selling with 100% return in 90 days. Though such fraud might occur prior to Charles Ponzi’s contemporary this was well notified economical crime after Charles Ponzi was Busted.

1.2 Famous Ponzi Scheme Cases

Bernie Madoff [4] Ponzi scheme scam: Bernie Madoff was an American financial manager who ran a Ponzi scheme for more than two decades with a magnitude of more than 60 billion US dollars.

The center said that various programs and campaigns have been run by the government to sensitize people about various fraudulent and Ponzi Schemes.

Over the last 3 years, 8 cases involving 84 companies have been assigned to the Serious Fraud Investigation Office (SFIO) to inspect Ponzi schemes/multi-level marketing/chit fund activities, Minister of the State of Corporate Affairs, Rao Inderjit Singh, said in the Rajya Sabha on Tuesday [5, 6].

2 The Technology Adopted by Ponzi Schemes

The complex infrastructure model has been adopted by Perpetrators who ran Ponzi schemes in India. The infrastructure model is apparent that it inherits an intricately maneuvered model whereby it can't be unraveled and undetected by regulators and investigators.

3 Dissect and Understanding of the Ponzi Scheme Infrastructure

The perpetrators start a company with or without registration. Sometimes, the name of the company gets registered but hardly registers or associates with other regulatory bodies such as the Securities and Exchange Board of India (SEBI) and the Securities and Exchange Commission (SEC).

In the above hypothetical scenario (Fig. 1), the perpetrators' head office located in Chennai has branches across the country. The Head office communicates with branches through Server and Client model ERP application whereby branch records are fed into the head office. The branch collects and records investors' details such as name, identity, mobile number, email ID, and banking details. Branch may send all data to head office or selective records such as banking details, name, and mobile number. The Server/Client model ERP software is made with the front end in Dot Net and the Backend database in MySQL. The ERP software provider and support are from the Bangalore location and the application and database are hosted in a cloud server located in Dubai.

The selected records from branch data are fed to the head office. The data from different branches gets consolidated and further trimmed and send to the main database server. The main database server is hosted in a cloud server located in California which is an MS SQL Database with the front end of Python. The main

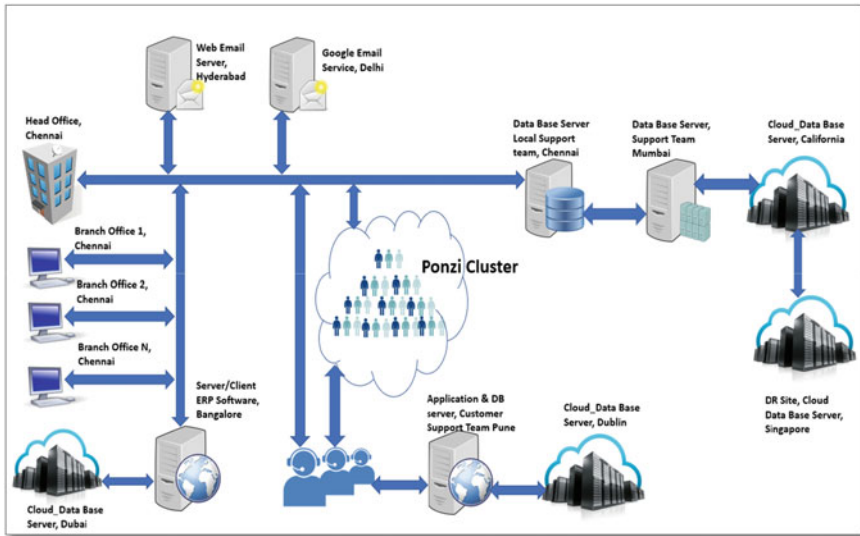


Fig. 1 Schematic diagram of Ponzi scheme IT infrastructure

database of consolidated records is hosted in a cloud server located in California and its DR (Disaster Recovery) site is located in Singapore.

The consolidated main database and applications have been taken care of and supported by a local vendor located in Chennai and the local vendor gets support from a database and cloud management service provider which is located in Mumbai each location has a copy of the data of the main database hosted in cloud server, California. The main database copy at each location may not be the same. The integrity of data should be checked to affirm the databases are the same.

Customer support team has taken care of providing services such as reaching new customers, notifications, and alerts to new customers, and answering customer queries. The front-end application is PHP and the backend database is Postgres DB. The team has been placed in Chennai and the customer support application and database are managed by a vendor located in Pune the application and database are hosted in a cloud server located in Dublin.

The perpetrators have used webmail and been hosted on a server located in Hyderabad. Google email service has also been used and the server is located in Delhi.

The perpetrators have also used NAS storage, laptops, mobile phones, printers, POS, and other digital gadgets which may have crucial evidence.

4 Digital Forensic Investigation of Ponzi Schemes

The digital forensic investigation deployed on the Ponzi scheme was either detected by the monitoring system after notification of several red flags or by complaints given by the victims. The intelligence should be gathered prior to the action against the perpetrator who runs the Ponzi scheme. The intelligence should be sophisticated enough to counter the people and technology involved in the scam. The vagueness in the intelligence would seriously affect the course of the investigation and might lose the plot of the whole magnitude involved in the scam.

The perpetrators who run the Ponzi scheme be definitely predetermined, scripted, and made their system more complex and not to get detected and caught. Understanding and dissecting their infrastructure is important to unravel the real magnitude involved in the Ponzi scheme. The search and seizure operation should be carried out with a high degree of caution without leaving the critical premises and persons inclusive of all gadgets associated with the scam.

The identification of systems and gadgets present in the head office and branches are easier when compare to the servers managed and located in the different parts of the jurisdiction. The time-lapse in the action among the target location may lead to a breaking in the chain of investigation as well as loss of evidence.

The confiscation of digital gadgets associated with perpetrators may provide lead to miss may be missed by pre-intelligence. Forensic imaging and analysis of collected digital devices and gadgets have to be performed.

The cloud server and the servers located in the remote location have to be imaged. Webmail and other email server data have to be forensically imaged or copied.

The analysis of all the data from the individual's mobile, pen drive, or laptop to the organization's application, database, and email data.

Cloud server logins and IPs associated with cloud service provider needs to be identified and probed further to check if any other instances or databases are running. Collect cloud logs to identify any deletion activity that took place before or after the action. Identify any column or table or record that has been deleted, removed, or dropped in the database. If found to be deleted or suspected to be deleted or tampered prompt for the backup database from the respective vendors.

The database and other data associated with the Ponzi scheme are subjected to a credibility check. Since the Ponzi scheme involves huge money and people the victims should not be impersonated as perpetrators and perpetrators should not be impersonated as victims.

The digital sweep (screening of all electronically stored information (ESI) associated with the entity) should be deployed to find any "Smart Ponzi scheme" also running under the regular parent scheme. If found deploy proper detection tools to identify such smart Ponzi schemes with blockchain technology.

The complexity of data redundancy and fragility in different servers and databases analysis so difficult. The database used here is MSSQL, MySQL, and Postgres DB and the application used here is Dot.NET, Python, and PHP so the knowledge to interpret the data and evaluate the overall scam is tedious. The perpetrator might

delete an important table or column even the database if any of the premises gets missed or can be accomplished remotely or accomplished in lapse time during action.

The database analysis should provide maximum pristine data whereby actual consolidated figures involved in the scam can be figured out. The payouts have already been made and the payouts have to be made need to be assessed in the pool of investors who have received more than what they have invested and the pool of investors who have not received part or full from their investments.

The proper database analysis with other corroborative evidence would give proper figures where victims can be benefitted without any hitches.

5 Hurdles in the Forensic Investigation

5.1 People

The Ponzi scheme is a predetermined scam thus people who are involved as brain children will definitely flee or abscond thus getting investments out of such scams would be difficult to trace and confiscate.

5.2 Infrastructure

By nature, the scheme would fall sooner or later so made the infrastructure systematically difficult to trace and detect.

5.3 Technology

Technological feasibility is exploited here by accomplishing complex infrastructure. Technological versatility needs to investigate such complex technological systems. “Media”—laptop, desktop, server, and mobiles to “Data” Database and Email to “Information” Documents and Spreadsheets to “evidence” Final reporting File.

5.4 Jurisdiction

Policy and law always have a problem when handling jurisdictional tyranny.

6 Trends and Challenges

- Smart Ponzi schemes have come and detection of such would be easier but the identification of source code is challenging [7].
- Tracking of investments particularly in cryptocurrencies is difficult as it's decentralized and peer-to-peer in nature.
- Smart contracts are also hard to track so investment in such platforms by perpetrators would certainly be hard to detect.
- Regular Ponzi schemes and smart Ponzi schemes together made a hybrid model which involves combined traditional investments and crypto investments.

7 Conclusion

The Ponzi scheme is not new eventually identified way back in the 1920s executed by the scamster “Charles Ponzi”. The contemporary form of the Ponzi scheme will float always with time to time. Now, technological advancement provides new avenues to accomplish the scam by the perpetrators.

The technology will also help the investigation team, especially digital forensics investigations to detect Ponzi schemes but the nature of the scheme itself scam and the fleeing nature of perpetrators would be made investigations less smooth. The multi-jurisdiction and extra-terrestrial investigations need a lot of time, money, and manpower.

Pre-intelligence prior to Search and Seizure action is so important, especially the scams like the Ponzi scheme.

Proactive is better than reactive or detect. Policy and laws should be amended with respect to the technological phase, crimes, and frauds.

The regulatory bodies have to be vigilant and any red flags raised by the monitoring system should be addressed and investigated and if the absence of red flags with the apparent scams, then the monitoring system should be updated to compete with a phase of the technology, crime, and perpetrator.

References

1. https://www.researchgate.net/publication/324509423_Detecting_Ponzi_Schemes_on_Ethereum_Towards_Healthier_Blockchain_Technology#pf2
2. <https://internationalbanker.com/history-of-financial-crises/charles-ponzi-1920/>
3. https://en.wikipedia.org/wiki/Ponzi_scheme
4. <https://www.investopedia.com/terms/b/bernard-madoff.asp>
5. <https://www.hindustantimes.com/india-news/84-firms-involved-in-8-ponzi-scheme-cases-in-3-years-centre-in-parliament-101649161707682.html>
6. Source: Hindustan Times Published on Apr 05, 2022, 05:58 PM IST
7. <https://ieeexplore.ieee.org/document/9407946>

Holistic Cyber Threat Hunting Using Network Traffic Intrusion Detection Analysis for Ransomware Attacks



Kanti Singh Sangher, Arti Noor, and V. K. Sharma

Abstract In recent times, cybercriminals have penetrated diverse areas or sectors of the human business enterprise to initiate ransomware attacks against information technology infrastructure. They demand for money called ransom from organizations and individuals to save valuable data. There are varieties of ransomware attacks floating worldwide using intelligent algorithms and with the usage of different setup vulnerabilities. In our research work, we are exploring the latest trends in terms of sector-wise infiltration, captured the most popular among available and also the distribution of the number of attacks using the location information available at the country level. To achieve the correlation between the sectors and locations along with the parametric analysis, we have utilized artificial intelligence techniques. Accuracy of the prediction of attack based on the sector level analysis we have implemented Random Forest and XGBoost algorithm. This research work focuses primarily on two aspects, first is to explore the different aspects of ransomware attacks using intelligent machine learning algorithms. The method used insights to severity of spread of ransomware attacks, second research outcome is to forensically evidence finding of the attack traces using traffic analysis. The challenge is to learn from the previous weaknesses available in the infrastructure and at the same time to prepare the organization and countries' own prevention methods based on the lessons learnt, our exploratory analysis using the latest set of data implementing with AI will give a positive dimension in this area. Also, the proactive approach for managing the data safely is based on the finding of digital forensic analysis of infected ransomware traffic.

K. S. Sangher (✉) · A. Noor · V. K. Sharma
School of IT, Centre for Development of Advanced Computing, Noida 201307, India
e-mail: kantisingh@cdac.in

A. Noor
e-mail: artinoor@cdac.in

V. K. Sharma
e-mail: vksharma@cdac.in

Keywords Ransomware attack · Cyber threat intelligence · EternalBlue · SMB Random Forest · XGBoost · KNN REvil

1 Introduction

Ransomware is one of the most sophisticated online threats, dominating security exploitation at both the individual and organizational levels. Data loss is unaffordable because of the sensitivity tied to it and variants are becoming more harmful as time goes on. Therefore, it is imperative to conduct intelligence analysis in order to filter the most recent attacks [1]. If we can manage that, it will enable us to improve the infrastructure, fortify the environment, and, most critically, be well prepared to thwart future attacks.

The security dangers are changing as well as our dependence on digital technology grows in both our personal and professional lives. One of the prominent threats is malware heavily damaging the cyberspace. Ransomware and its impact have changed the reach of malware in worldwide, once infected the resource device restricts the access of files and folders till the ransom raised by the cybercriminal paid, mostly nowadays in digital currency such as Bitcoin to get the data back. Recent trends show that ransomware is not limited to a particular domain but penetrates different sectors such as education, health, information technology, business, and research. Nature of the attack and its consequences are tricky in the case of ransomware as the damage is mostly irreversible even after the removal of the malware that caused the attack. Hence, cyber security becomes a critical concern for researchers and organizations to find the solution to overcome ransomware attacks or to be prepared with preventive solutions. Recently, ransomware has matured in intricacy, difficulty, and diversity to turn into the most vicious among the existing malware trends. In addition to this, Cisco's annual security reported that ransomware is rising at a yearly rate of over 300%. The method used gives insights to severity of spread of ransomware attacks, and the second research outcome is to forensically finding the evidence of the attack traces using traffic analysis [2].

2 Literature Survey

Critical infrastructure is severely impacted by malware, sometimes known as malicious software. The goal of these is to harm the victim's computer or service networks. Malware comes in many different forms, including viruses, ransomware, and spyware. Malware known as ransomware has been shown to use complex attack methods that have undergone numerous mutations. Many different sectors of the economy, including transportation, telecommunications, finance, public safety, and

health services, have been impacted by ransomware. User data is made unavailable or unreachable using crypto modules incorporated in malware. The organization is required to pay a ransom to regain access once ransomware either locks the equipment or encrypts the files. With code obfuscation, various ransomware display polymorphic and metamorphic tendencies, making it difficult to scan for and identify them with current methods. The market is classified by deployment, application, location, and other factors for the many open-source and commercial anti-ransomware programs. AOKasperskyLab, FireEyeInc, MalwarebytesInc, SophosLtd, SymantecCorporation, SentinelOneInc, ZscalerInc, TrendMicro Incorporated, and more top companies offer ransomware prevention solutions. Using software libraries and sandboxes like CryptoHunt, Cryptosearcher, etc., ransomware is discovered by conducting a crypto module search and analysis.

The new ransomware variants are not recognized by the existing solution, and the effects are only discovered after the attack, despite the fact that there are ongoing upgrades or improvements to the existing anti-ransomware systems. Analysis also demonstrates that no ransomware variant changes after each infection, allowing ransomware authors to stay one step ahead of their victims because the ransomware's associated signatures, domains, and IP addresses become dated and are no longer recognizable by threat intelligence and signature-based security tools. In the critical infrastructure sectors, the requirement for a new paradigm to recognize the new and evolved ransomware is crucial. These ransomware attacks have the potential to cause significant financial harm and substantial losses, making cyber-insurances for all enterprises necessary.

The majority of ransomware detection programs use behavioral detection, often known as "dynamic analysis" [3–6] based on dynamic analysis using an SVM classifier. They initially retrieved a specific ransomware component known as the Application Programming Interface (API) call, after which they used Cuckoo Sandbox to analyze the API call history and its behaviour. Q-gram vectors serve as a representation for the API calls. They employed 312 goodware files and 276 ransomware files. The findings showed that using SVM, ransomware could be detected with an accuracy of 97.48%. Vinayakumar et al.'s [5] new approach suggested collecting the API sequences from a sandbox utilizing dynamic analysis.

3 Present Situation of Ransomware Worldwide and India's Stack

India is emerging as a cyber power in the international community and at the same time by the end of this year, about 60% of the Indian population (840 million), will have access to the internet, claims a report. The flip side of this is, increasing cybercrimes. Over 18 million cases of cyber attacks and threats were recorded within the first three months of 2022 in India, with an average of nearly 200,000 threats

every day, according to the cyber security firm Norton. So, there is a strong need to be more vigilant, proactive, and smart while handling the cybercrimes.

3.1 Finding Recent Heavily Used Set of Ransomware Attacks and Their Behavior

The attackers are using the more complex, destructive, and easy-to-execute ransomware variants [7]. The dataset used in our research work provided by the DSCI sources consists of the worldwide penetration along with the per sector as target. The dataset “Ransomware Attacks” has the following attributes: description, sector, organization size, revenue, cost, ransom cost, data note, ransom paid, YEAR, YEAR code, month, location, Ransomware, no of employees Source Name, URL details, etc. The training environment for the ransomware attack dataset is set up using Anaconda v3. Environment implemented using IPython utilizing Jupyter Notebook from Google Collab. The word cloud created from the dataset indicates that REvil, Cryptowall, and Ryuk are some of the most recent trends. Even a small number of the recent ransomware attacks have unclear sources (Fig. 1).

For different operating systems and gadgets, there are many ways that ransomware actually gets onto the computer. One of the most notable assaults in 2021 was REvil, which successfully hacked more than 1500 supply chain management-based firms and spread unintentionally to all associated clients. A remote IT management service provider, Kaseya MSP, was one of those compromised and used to distribute the ransomware REvil/Sodinokibi. The ransom demand was made by evil threat actors, who demanded between \$50 million and \$70 million in Bitcoin to unlock all the encrypted information. Virtual Systems Administrator (Kaseya VSA) is a remote management and monitoring application for networks and endpoints used by businesses and MSPs. The Kaseya network was breached by the REvil attackers, who then took control of the VSA software update system to deploy the initial payload via the Kaseya agent.

C:\Program Files (x86)\Kaseya\{ID}\AgentMon.exe

Fig. 1 Word cloud created using the ransomware attribute from the dataset

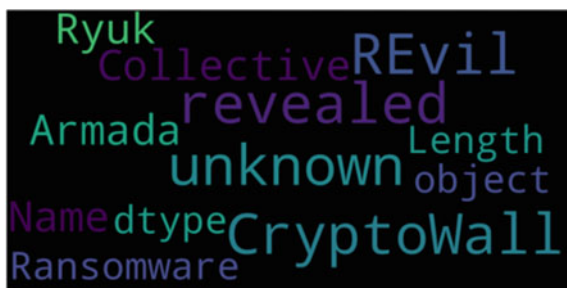
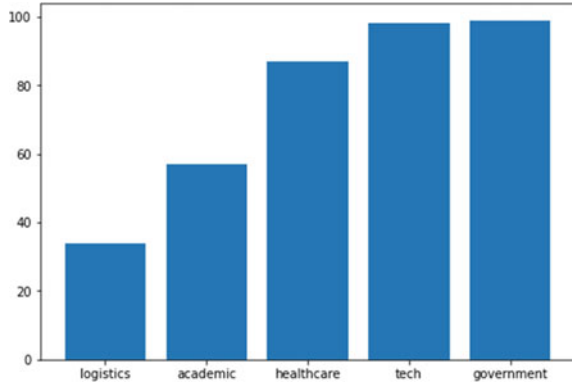


Fig. 2 Existing and emerging targets for ransomware attacks till 2021



The launch method used to intrude initiates payload by following the precaution to get detected. Again to make things perfect it removes the *certutil.exe (cert.exe)* and *agent.crt* transitional binaries too. The first threat intrusion seems genuine, being just an older version of Windows Defender binary (*mshpenc.exe*). The next one is a binary called REvil encryptor dll (*mpsvc.dll*). This facilitates side-loading of *mpsvc.dll* into *mshpenc.exe* to access the data.

Cybercriminals have understood that supply chain attacks are the new area to explore and make profit. Various MSPs or IT infrastructure management platforms are using agent software to provide organizations different services. If the security is breached of these deployments then they will be used to distribute harmful malware. It is a high time to protect the different domains which utilize the network and remote machines to share services the countermeasures and spread within the IT community should be availed [8]. So, based on the recent attacks dataset shared from the Data Security Council of India (DSCI) a premier industry body on data protection in India, setup by NASSCOM. We have used the AI techniques and performed analysis to gather target organizations details for ransomware attacks in last 5 years. Figure 2 shows the outcome from the analysis of the dataset in the form of a graph, which precisely depicts that, in recent time, government organizations are the prime target along with the technology-based industries, and then health care is identified as an attractive spot for cybercriminals where a lot of personal health records and medicinal data are floating further logistics is the upcoming area which is facing lot of variants of ransomware attacks [9].

3.2 Intelligent Discovery of the Attack Trends

Using the experiments on the dataset we tried to visualize the attack reach at the national level with a comparison to worldwide data. It clearly shows that India is one of the major targets and govt. organizations with technology are sharing the target. So, there is an immediate need to strengthen our govt. organizations to protect the

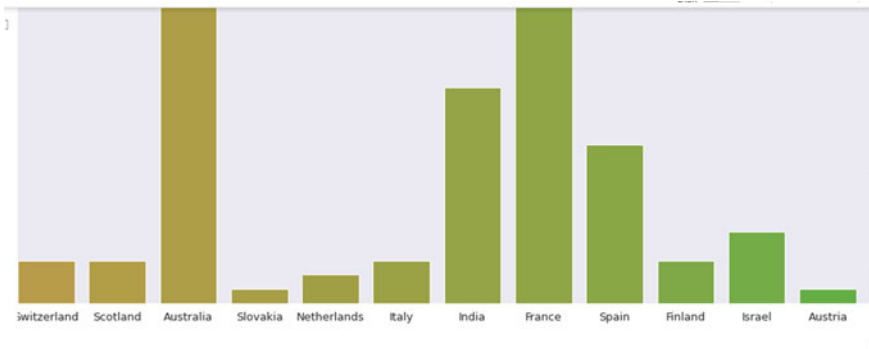


Fig. 3 India’s stack in worldwide statistics in ransomware attacks till 2021

environment to safeguard the data due to its sensitivity and the damage it can cause for our nation (Fig. 3).

One of the significant analyses shows the correlation between the different parameters of the dataset, which gives insights to understand the behavioral pattern after the incident happens as we observed from our intelligent analysis using the AI environment that per year cost of ransom is increasing [10] not only in terms of digital transaction but also due to the inclusion of cryptocurrencies, even is some incident or stories which are considered as case studies to prepare the best case practices also experience that after the ransom amount paid by the victim organization cybercriminals not sharing the decryption key on top of that shared decryption key not able to recover the complete data. Figure 4 depicts the correlation in years with respect to ransom cost.

After finding the latest trends of the attack in terms of most common ransomware attacks and the organization level filtration visualization. The next research work

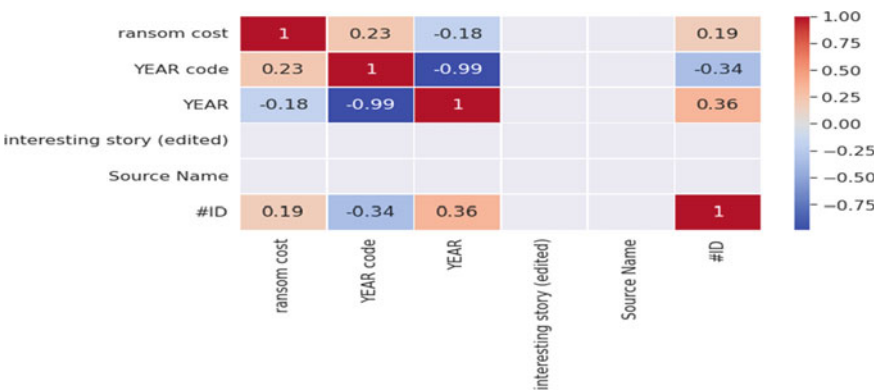


Fig. 4 Correlation which depicts the YEARS and cost of ransom rose

done was to perform the machine learning algorithm implementation for prediction of precise output (in terms of sector wise/organization calibration).

The correlation matrix's cells represent the "correlation coefficient" between the two variables corresponding to the cell's row and column. A strong positive relationship between X and Y is indicated by values that are close to +1, while a strong negative relationship between X and Y is shown by values that are close to -1. Values close to 0 indicate that there is no link at all between X and Y. The ransom price in this case varies without regularly growing or decreasing and is unpredictable. Numerous assaults have occurred recently, and the ransom price is high yet inconsistent. Additionally, the type of attack vector and the victims' response affect the total ransom price for ransomware attacks that is paid on an annual basis. This research work will provide very important inputs to the national-level security application team to check and prepare their environment [11], if the victim is aware, then the prevention will be better. After the training of the dataset, it splits the dataset and categorizes it into training and testing sets. The training set includes the classification information and based on the test data we implemented three algorithms, namely, Random Forest, XGBoost, and KNN. A composite data repository is produced by extracting all the features and combining them with the dataset of ransomware attacks. When the composite dataset was prepared, we ran two phases of tests on the same data repository. To determine the accuracy level in the first stage, we used the Random Forest and XGBoost algorithms, two machine learning methodologies. We also noted how long both algorithms took to process.

XGBoost outperforms the competition because it is forced to learn from its mistakes made in the prior stage (so does every boosting algorithm). The main drawback to using XGBoost is if you have a lot of categorical data. Then, you must perform One Hot Encoding, which results in the creation of more features. The overall line is that you can utilize XGBoost for better outcomes, but be sure to perform the proper preprocessing (Keep an eye on datatypes of features). To avoid overfitting, they employ the random subspace approach and bagging. A random forest can readily handle missing data if they are implemented properly. In the second phase, we used K-Nearest Neighbor (KNN) and assessed its accuracy and processing speed using the same dataset. By making calls to the local repository of Jupyter Notebook, we imported the required packages and libraries. The outcomes of applying various machine learning techniques to the ransomware dataset are displayed in the following screenshots. The results found shown in Fig. 5, indicates that an accuracy of implied machine learning algorithm, i.e., 93%, 92%, and 94%, respectively, for Random Forest, KGBBoost, and KNN model.

For machine learning when a model fits the training set of data too closely, it is said to be overfit, and as a result, it cannot make reliable predictions on the test set of data. This means that the model has only memorized certain patterns and noise in the training data and is not adaptable enough to make predictions on actual data. However, recall was set to 1 for the purposes of our research, and the data reports we received had varying degrees of accuracy. In order to find ransomware, Kharraz et al. [3] used a dynamic analytic method named UNVEIL. In order to find ransomware, the system builds a fake yet realistic execution environment. About

```

from sklearn.metrics import accuracy_score
import xgboost as xgb

xgb=xgb.XGBClassifier()
xgb.fit(X_train,y_train)
preds2=xgb.predict(X_test)
xgb_accuracy=accuracy_score(preds2,y_test)
rf_accuracy=accuracy_score(preds,y_test)
print("Random Forest Model accuracy",rf_accuracy)
print("XGBoost Model accuracy",xgb_accuracy)
knn = KNeighborsClassifier(n_neighbors=1, metric='euclidean')
knn.fit(X_train, y_train)
y_pred = knn.predict(X_test)
knn_accuracy=accuracy_score(y_pred,y_test)
print("KNN Model accuracy",knn_accuracy)

```

Fig. 5 Machine learning implementation

96.3% of the time, this system was accurate. Sequential Pattern Mining was employed as a candidate feature to be used as input to the machine learning algorithms (MLP, Bagging, Random Forest, and J48) for classification purposes in the framework ransomware detection system proposed by Homayoun et al. [6].

As a result, the KNN model produced the greatest results in this investigation. Simply said, KNN only stores a footprint of the training data within the model and doesn't really perform any training. KNN's logic is found in the predict() call of its inference step, which is where it uses previously provided training data to identify the k nearest neighbors for the newly supplied instance and predicts the label. For small- to medium-sized datasets, KNN is probably faster than the majority of other models.

4 Experimental Analysis of Incident

In our research work, the experimental analysis of ransomware packet capture was done using the Wireshark tool. It shows the step-by-step analysis to find out the traces of intrusion injected to perform the attack. A .tar file of WannaCry has been used as a source for analysis, as it depicts the causes or vulnerabilities within the system that allowed the successful execution of the attack; we propose the defense mechanism to protect the asset within the organization's/individual's infrastructure.

4.1 Infected Network Traffic Analysis

The malware's ability to remotely access files is what started the attack, and the attackers used a variety of mechanisms to carry it out. The Server Message Block (SMB) protocol, which is mainly used for file sharing, printing services, and communication between computers on a network, was abused in this case. Ransomware like

WannaCry took use of SMBv1 vulnerabilities by using them. All of these exploits go by the term “Eternal” X. EternalBlue is the one with which most people are familiar. EternalBlue was developed because SMBv1 cannot handle specific packets produced by a remote attacker, which can lead to remote code execution. The WannaCry virus reportedly infected over 230k computers in 2017 and other malware caused over \$1 billion in losses globally.

When WannaCry is installed on the network, the following significant things take place:

- The development of files with the WannaCry document extension specifically for encrypting files.
- Ports TCP 445 and 139 of SMBv1’s outgoing communication.
- The domain’s DNS requests for iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com.
- New entries into Windows registry.

All four of these events are trackable and observable. Finding malware that spreads across networks, like WannaCry, requires monitoring. Before the “.WNCRY” extension is introduced, WannaCry encrypts several distinct types of documents. The ransomware connects to the IPC\$ share on the remote system after the initial SMB handshake, which includes a protocol negotiation request/response and a session setup request/response, is shown in this study paper through analysis of the network data (Fig. 6).

The malware’s ability to connect to a hardcoded local IP is another related component of this attack. The majority of ransomware attacks, as was already noted, use DNS tunneling to create both bidirectional and unidirectional communication between an attacker and the systems on your network. The threat actor can hide until their attack is almost complete if the DNS action is not safe. The system was

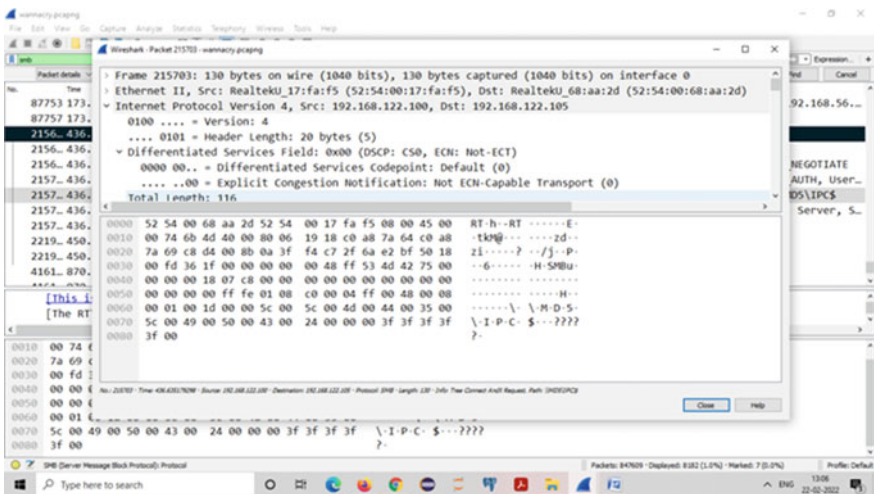


Fig. 6 IPC\$ share on the remote machine

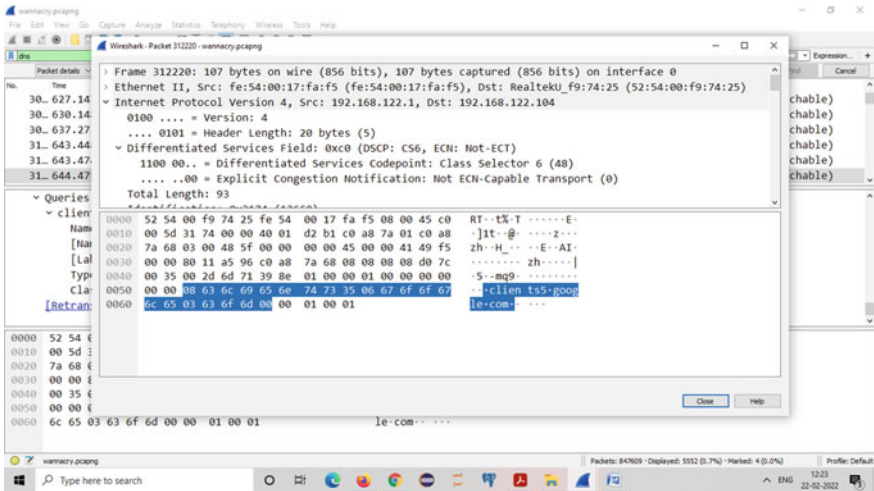


Fig. 7 Malware in traffic

vulnerable to ransomware execution since malware was present, which was discovered during the DNS traffic filtering process. Fig. 7 displays traffic-related malware. A malicious website is `clients5.google.com`.

It then sends a first NT Trans request with a huge payload size and a string of Nops, as seen in Fig. 8. In essence, it relocates the SMB server state machine to the vulnerability’s location so the attacker can make use of it. It then transmits a first NT Trans request with a huge payload size that is made up of a string of NOPs, as seen in Fig. 8.

To enable the attacker to use a specially constructed packet to exploit the vulnerability, it essentially moves the SMB server state machine to the place where it is present. The next step is to check whether the payload has been successfully installed. If it has, then the SMB MULTIPLEX ID = 82 will be found in one of the packets. The same has been done in this experimental analysis using the filter in Wireshark for a stream of packets and shown in SMB MULTIPLEX ID = 82.

The attack was launched utilizing the SRV Driver exploit MS17-010:Buffer EternalBlue’s Overflow. In the worst-case scenario, if an attacker sends specially designed messages to a Microsoft Server Message Block 1.0 (SMBv1) server, they could execute remote code (Fig. 9).

One packet, as seen in the screenshot, signifies that the payload has been installed successfully and that the attacker has run remote code on the victim network. The SMB MultiplexID = 82 field is one of the crucial fingerprints for this attack’s success. The contents of the packets can be seen by right-clicking on the Trans2 packet and choosing to Follow -> TCP Stream. The contents of the payloads that caused the buffer overflow and sent the payload necessary for this exploit are shown here (Fig. 10).

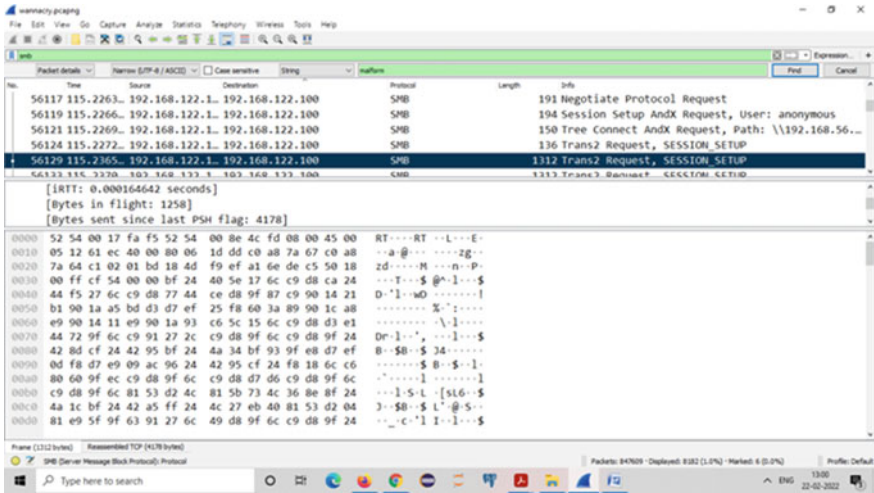


Fig. 8 NT Trans request with a sizable payload

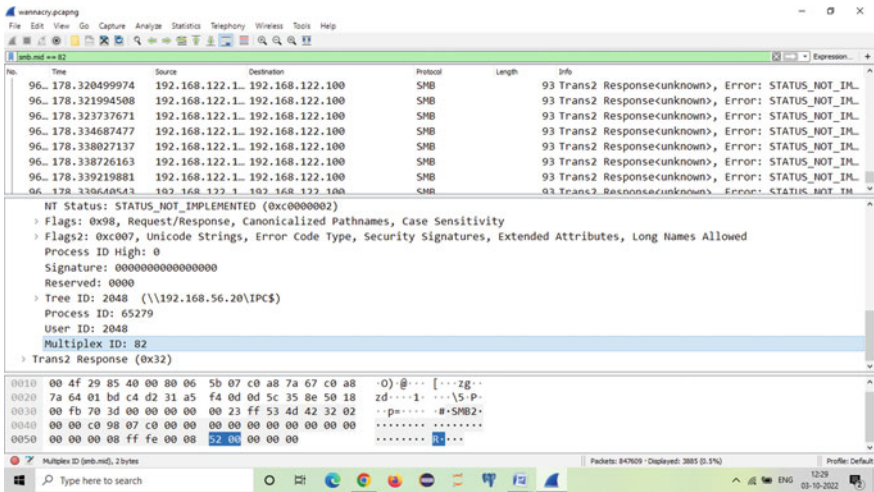


Fig. 9 SMB MULTIPLEX ID = 82 within the selected packet

4.2 Preventive Measures

The user can recognize this occurrence if there is any folder auditing on Windows folders. The following two entries can also be found in the Windows registry:

- HKLM\SOFTWARE\Microsoft\Windows\CurrentVersion\Run\<random string> = "<malware working directory>\tasksche.exe"

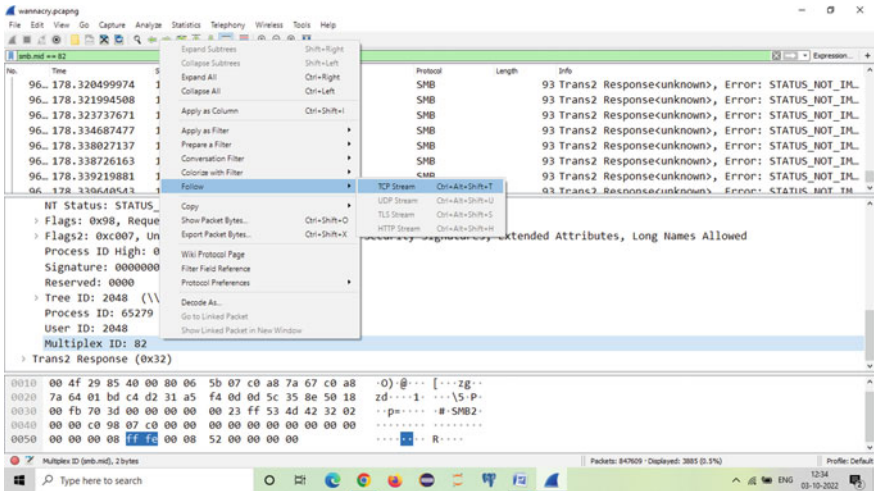


Fig. 10 To view the contents of the packets

- HKLM\SOFTWARE\WanaCrypt0r\wd = “<malware working directory>”.

On the local server, users can manually check for file creation and registry key entries or execute a PowerShell script to check for these occurrences. The other two incidents can be investigated using monitoring tools. Additionally, system files must be examined for incoming TCP requests on ports 139 and 445. A rise in requests on these two SMB-specific ports should raise red flags. DNS monitoring and analysis is the second method for finding WannaCry. To protect the system one more check will be helpful that is instigating the browser’s unique User_agent value. So, the first thing is to secure the system to prevent ransomware attacks, but if the system security is not strong enough, then learning from the entry points of the incident can help to make the secure environment.

5 Conclusion

The paper proposes to intelligently analysis of the ransomware attack data in recent years to visualize the pattern and target of the popular attack, and also how to harden the security to prevent the system from such attacks. This will help tremendously to prepare and harden the organizations’ security infrastructure to protect as well as detect the intrusion. The results present intelligent algorithm solutions to ransomware attack penetration at organization level along with the latest set of attacks floating at the worldwide level [12] and also analyze the infected network traffic to find the traces of the ransomware execution using the tool. The research finding gives insightful directions to be aware of the existing threats and prepare the cyber resilience

environment and platforms where targets are identified and well advanced to enable the systems to fight and protect it from the threat vectors.

6 Future Scope

The future scope of the present work can be the initial root cause analysis or, in other words, finding the vulnerabilities that were the reason behind the success of cybercriminals. Root cause analysis of the attack will definitely help the organizations' security infrastructure handling team to understand vulnerabilities that helps ransomware to enter in their deployments [13]. Post-incident analysis always gives feedback to be prepared for prevention measures as it helps to serve to plan to handle the situation if the incident happens. Recovery phase also needs to be analyzed due to the uncertainty of the data recovery from the ransomware attacks. A few very basic common points in ransomware attacks are browser exploitation, email, etc. Other vulnerability exploration can be improved in future and applying deep learning for intelligent analysis will be a great area to work.

References

1. Connolly LY, Wall DS, Lang M, Oddson B (2020) An empirical study of ransomware attacks on organizations: an assessment of severity and salient factors affecting vulnerability. *J Cybersecur* 6(1):tyaa023. <https://doi.org/10.1093/cybsec/tyaa023>
2. Dasgupta D, Akhtar Z, Sen S (2022) Machine learning in cybersecurity: a comprehensive survey. *J Def Model Simul* 19(1):57–106. <https://doi.org/10.1177/1548512920951275>
3. Lee JK, Chang Y, Kwon HY et al (2020) Reconciliation of privacy with preventive cybersecurity: the bright internet approach. *Inf Syst Front* 22:45–57. <https://doi.org/10.1007/s10796-020-09984-5>
4. Kirda E (2017) Unveil: a large-scale, automated approach to detecting ransomware (keynote). In: 2017 IEEE 24th international conference on software analysis, evolution and reengineering (SANER). IEEE Computer Society, pp 1–1
5. Takeuchi Y, Sakai K, Fukumoto S Detecting ransomware using support vector machines. In: Proceedings of the 47th international conference on parallel processing companion (ICPP '18). Association for Computing Machinery, pp 1–6. <https://doi.org/10.1145/3229710.3229726>
6. Vinayakumar R et al (2017) Evaluating shallow and deep networks for ransomware detection and classification. In: International conference on advances in computing, communications and informatics (ICACCI), pp 259–265
7. Li Y, Liu Q (2021) A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments. *Energy Rep* 7:8176–8186. ISSN 2352-4847. <https://doi.org/10.1016/j.egy.2021.08.126>
8. Bagdatli MEC, Dokuz AS (2021) Vehicle delay estimation at signalized intersections using machine learning algorithms. *Transp Res Rec* 2675(9):110–126. <https://doi.org/10.1177/03611981211036874>
9. Farhat YD, Awan MS (2021) A brief survey on ransomware with the perspective of internet security threat reports. In: 2021 9th international symposium on digital forensics and security (ISDFS), pp 1–6. <https://doi.org/10.1109/ISDFS52919.2021.9486348>

10. Gibson CP, Banik SM (2017) Analyzing the effect of ransomware attacks on different industries. In: 2017 international conference on computational science and computational intelligence (CSCI), pp 121–126. <https://doi.org/10.1109/CSCI.2017.20>
11. Farion-Melnyk, Rozheliuk V, Slipchenko T, Banakh S, Farion M, Bilan O (2021) Ransomware attacks: risks, protection and prevention measures. In: 2021 11th international conference on advanced computer information technologies (ACIT), pp 473–478. <https://doi.org/10.1109/ACIT52158.2021.9548507>
12. Homayoun S, Dehghantanha A, Ahmadzadeh M, Hashemi S, Khayami R (2017) Know abnormal, find evil: frequent pattern mining for ransomware threat hunting and intelligence. *IEEE Trans Emerg Top Comput* 8(2):341–351
13. Jeong D (2020) Artificial intelligence security threat, crime, and forensics: taxonomy and open issues. *IEEE Access* 8:184560–184574. <https://doi.org/10.1109/ACCESS.2020.3029280>

Cyber Security Attack Detection Framework for DODAG Control Message Flooding in an IoT Network



Jerry Miller, Lawrence Egharevba, Yashas Hariprasad, Kumar K. J. Latesh, and Naveen Kumar Chaudhary

Abstract Advancement in the IoT technologies and the futuristic device's usage influences the human life in all the aspects of day-to-day activities. Moreover, human reliance on smart objects makes IoT an important tool. IoT network enables communication among smart devices by embedding software, sensors, etc., which makes an object smart and intelligent. Though it offers many advantages, it is a matter of concern in protecting the privacy, integrity, and availability of the users' data, and these issues need to be addressed in the implementation of the devices before it turns out to be a threat. DDoS is one such security threat which can bring down the resource-constrained IoT network. In this work, we have tried to address the existing availability issues in the IoT communication network and based on the analysis, proposed an attack detection framework for the DDoS traffic generated by IoT devices. The attack detection is done by keeping track of the usage of IoT devices parameter like power consumption and bandwidth, and monitoring the IoT network traffic to oversee the number of messages exchanged between the nodes as part of the RPL DODAG construction so that resources and bandwidth can be used for genuine communication. The proposed work has achieved better bandwidth and our simulation framework obtained good results in identifying DDoS attacks.

Keywords IoT · Integrity · DDoS · RPL · DODAG

J. Miller (✉) · L. Egharevba · Y. Hariprasad · K. K. J. Latesh
Florida International University, Miami, FL 33174, USA
e-mail: millej@fiu.edu

L. Egharevba
e-mail: legha001@fiu.edu

Y. Hariprasad
e-mail: yhari001@fiu.edu

K. K. J. Latesh
e-mail: lkumarkj@fiu.edu

N. K. Chaudhary
National Forensics Sciences University, Gandhinagar, Gujarat, India
e-mail: naveen.chaudhary@nfsu.ac.in

1 Introduction

The Internet of Things (IoT) is a huge arrangement of interconnected heterogeneous gadgets to sense our physical environment and convey the detected data. Every sensor identifies its region and sends the sensed data to the sink node through its neighboring nodes utilizing multi-hop correspondence [1]. The IoT devices and services are quickly becoming pervasive. Threats on the internet are increasing in number alongside the complexities. Digital threats are not new to the IoT internet world; however, IoT has profoundly impacted our lives. Hence, it is vital to venture up and pay attention to digital safeguard. Thus, there is a genuine need to make sure about IoT, which has subsequently brought about a need to thoroughly comprehend the attacks and threats on IoT infrastructure [2]. The useful applications of such a system include smart cities, building automation and industrial, smart grid, smart healthcare, and disaster management. IoT has interconnected billions of smart objects in today's digital world. Nonetheless, the majority of IoT devices are insecure and simple to be contaminated with malware. IoT devices provide wide-range applications and services in the network and the resource are constrained. The attackers easily exploit the vulnerabilities, with heterogeneous threats posed to IoT network [3]. Doubtlessly, smart homes gather huge amounts of private information, and smart offices automate the energy consumption and micromanagement of all the activities of the work environment. Privacy and availability play an important role in securing IoT smart applications.

The very common attacks in IoT networks and devices are physical attacks, encryption attacks, DDoS, and Firmware Hijacking. These attacks are because of the vulnerabilities exploited by the attackers, a few such vulnerabilities are, insufficient validation and authorization, untrusted user interfaces in the IoT applications [4], data leakage due to the integration of the smart device with the user applications, unreliable network, and privacy problems because of untrustworthy end points. With the help of malware bots or the zombies, the attacker gains access to the network and takes the control of the devices in the network. An attacker will start flooding the commands or the messages remotely using the bots [5]. The network layer is the one which is more vulnerable to attacks by constructing the different types of packets and pumped into the layer so that the network is congested and genuine services are denied. The flooding attack is a type of DDoS attack existing in the IoT network, in which the malignant nodes emulate the genuine nodes in all circumstances aside from the paths found by them. A sensor finds its neighbors, builds the topology, and routes the detected information. The node is considered as the local node when a node receives the data packet from another node within its range. There are attack detection methods in the IoT networks to detect the possible threats to devices in the network. Most recent models use machine learning and artificial intelligence to detect the attacks on the live traffic [6–10]. As these models work with the live traffic which requires the usage of the internet and other resources, this degrades the system performance. Machine learning models can be tricked into making incorrect decisions. An attacker node can disrupt and confuse the model's decision-making

process. Machine learning requires a huge data set to make the correct decisions but, if the data set is huge, then preprocessing and decision-making delay the process. This may cause huge damage to the network and to the application users. Few attacks cannot be detected by monitoring the live traffic as the nodes behave like a genuine node. In such cases considering the network traffic over a period of time, applying the statistics to detect the attack would be ideal and helpful. This paper focuses on one of the DDoS attacks called DODAG control message flooding. In general, an IoT system uses the DODAG control message to construct the network topology. All nodes share their configuration details by exchanging the messages and the topology is constructed by choosing the parent node. The protocols require the nodes used for communicating the DODAG control messages or packets for reporting to the neighbors they are existing. This DODAG messages exchange happens periodically to get updated with the changes happening in the network. All the updates should be propagated to the sink node [11, 12]. The DODAG control messages flooding attack is identified as the network layer attack, which targets the routing protocols. The proposed attack detection framework detects the early signs of attack by keeping track of the DODAG message exchange statistics of each node. The packet information is summarized and analyzed to check the signs of attacks by using the indicators of compromise.

The process is automated to collect the network traffic and analyze it as listed below:

1. Capture the IoT device traffic, especially DODAG control messages without disturbing any end points in the network.
2. Redirect the traffic to the dedicated system where it aggregates the data.
3. Based on the packet statistics of the DODAG control messages, the malignant node is identified.

A dedicated checkpoint system must be used which is called an attack detection framework system to aggregate all the data which is received from the packet-capturing tools in the network to analyze the packet data for attack detection. This process is repeated for every pre-defined time interval to make sure the network is secure.

The remaining section of the paper is organized as follows. The next immediate section confers the related work describing peer work associated with the proposed paper with detailed merits and demerits and is identified as Sect. 2. The proposed model and algorithm for the proposed model are discussed in Sect. 3. The simulation setup, results, and discussions of the proposed system are done in Sect. 4. Tools used in the experiment and the model evaluation are discussed in Sect. 5, and Sect. 6 closes with conclusion.

2 Related Works

The work proposed in the paper is a victim-end solution for handling high-rate DDoS that has a dedicated Flooding Attack Detector (FAD). FAD receives the copy of the network traffic destined toward the server on one of its ports. This method works with few constraints in the network, such as a huge packet traffic has to be generated with more zombies in the network within a short duration, source IP of the zombies controlled by the attacker would be spoofed, and DDoS attack traffic is generated with single attack technique [13]. The attack identification module proposed in the paper [14] identifies the malicious flow by designing a long short-term memory model with flow-oriented features such as each fragment length and time duration. Based on these features it marks as malign or benign otherwise, if any flow is suspicious that is again fed to the CNN for the DDoS attack classification [6]. Many states of the art flow-based intrusion detection systems are discussed in the papers [15]. The flow-based attack detection poses some drawbacks as they used generalized network information to identify the attacks in the network. There are different models proposed to detect the threats in IoT using the techniques like deep learning [7], KNN [8], deep CNN [9], record based [16], dynamic path identifier [17], and LS SVM [18] methods.

The paper focuses on security threats on the resource-constrained IoT devices, network layer, and its applications [19]. The network layer attack detection can be performed with the help of performance metrics like CPU usage, data rate, bandwidth, and power consumption [20]. The IoT network topology is constructed by exchanging the Destination Oriented Directed Acyclic Graph (DODAG) control messages among all the devices. All the traffic is routed through the root in the DODAG, and initially, each node announces its presence by sending a DODAG Information Object (DIO). This information is broadcasted to all the nodes in the network, and DODAG is constructed. Whenever a new node is joining the network, it sends a DODAG Information Solicitation (DIS) request, and the node responds back with a DAO Acknowledgment (DAO) confirming the join [11, 12]. In spite of the fact that message security gives privacy and integrity of information communicated among the nodes, an attacker can launch various attacks against the IoT network to bring down the network would affect the availability feature of the security. Attacks on the network layer and routing protocols are generally regular in low-power devices. Different attacks against RPL protocol are selective-forwarding, sink-hole attack, HELLO flood attack, Wormhole attack, and clone ID and Sybil attacks [21]. In this paper, an attack detection framework is proposed to detect the DODAG control messages flooding attack [22]. If malicious nodes act as root node in HELLO flood attack, one way to overcome this attack is a self-healing method in RPL by selecting the different parent [21–23]. If self-healing does not mitigate the attack, other techniques can be used. If a malicious node acts as a sensor node, an attacker uses HELLO flood attack to flood the root node and then eventually spreads to all the nodes in the network through the neighbors. In that case, an attack detection technique would help to detect the existence of the attack and also to find the source of the attack. This

paper proposes an attack detection framework for DODAG control message flooding attack.

A dedicated system with attack detection modules and indicators of compromise will analyze the captured traffic offline for every time interval to detect the signs of attack. The traffic is analyzed with respect to each node and its behavior based on the DODAG message transmission and reception to find the source of attack. This reduces the overhead of bandwidth consumption of online traffic analysis and reduces the overhead of including the attack detection modules in the resource-constrained devices of the IoT network.

3 Proposed Method

The proposed framework constructs the Routing Protocol for Low-power and Lossy Network (RPL) topology using the RPL DODAG tree that contains one root node which is also called a sink node and many sensor nodes [2]. An increased number of DODAG Information Object (DIO) or DODAG Information Solicitation (DIS) messages reflects instability in the network. Based on this, we model the attack detection framework to detect the attacks, if a malicious node enters the network posing as a genuine node in an IoT network, then malicious node starts communicating with the other nodes in the network by continuously sending the DIS messages to its neighbor in the network. Then the node selects its neighbor node as its parent. The malicious node continuously sends the DIS messages and the neighbor node responds by sending the Destination Advertisement Object (DAO) messages. And again, the neighbor node sends the DIS messages to the nodes within its radio frequency range. This continues to propagate through the network along the path till it reaches the sink node. The nodes along the path from the malicious nodes to the root node will be busy exchanging the DIS and DAO messages because of the DIS flooding happening from the malicious node. Henceforth, we can conclude and prove that the device power, network bandwidth, and other resources are not used for genuine communication. In this paper, we propose an attack detection framework by analyzing the network traffic that tracks the number of DODAG messages exchanged over a period of time by each node in the network. When it reaches the threshold, based on the indicators of compromise, we can detect the HELLO flooding attack and the source of the attack by tracking back from the sink node to the attacker node along the path, by considering the power consumption parameter for each node, and by keeping track of the power consumption of each node during the message transmission and reception. If any node is consuming more power than the expected power, then the node can be identified as the source of the attack. As a result of action, block the traffic from the device into the IoT network.

3.1 System Model

Consider the smart home IoT application system equipped with intelligent sensor devices like lighting control, entertainment devices, and appliances [24] are presented in Fig. 1. The information from each and every smart device is transmitted to a single sink node. Assume the sink node in smart home as a smart meter, which keeps track of the power consumption of all electronic appliances with the help of smart sensor nodes. Then. There is a direct communication among the devices within the radio frequency range. The sink node sets up the smart home topology with the help of the DODAG construction algorithm by exchanging the DODAG messages. Then the information is available to the user in a specific user application once it is transmitted through the internet, and the data is processed over the internet.

3.2 Architectural Model

The model contains the IoT communication network among the intelligent devices, the network traffic generated out of it may contain the genuine and malign packets. The same traffic will be fed to the attack detection framework, which has the traffic statistics analyzer. ADF includes analysis phase for analyzing the traffic and detection phase to detect the signs of attack by using the various parameters. ADF uses the defined indicators of compromise to detect the threats against the network traffic. The diagrammatic flow of the proposed model as shown in Fig. 2 involves the IoT network with a variety of intelligent devices communicating over the internet and generates a lot of device traffic. This will be the input to the attack detection framework which intern contains the traffic analyzer, and the ADF that can detect the DODAG Control Message Flooding attack.

In the detection phase, it checks the power consumption details of each node during the traffic generation, and the time spent by each node for sending the DODAG messages. If any node has sent any of the DIO, DIS, or DAO messages continuously over a period of time, the packet and power statistics of that node along the path toward the root node will be summarized and analyzed to check if there are any signs of attack. If the result of the packet statistics for a node crosses the threshold, then that node or device is marked as a malicious node and the traffic will be labeled as malign with red color. If the result of the packet statistics of a node or device is almost

Fig. 1 System model: from smart devices to the user applications

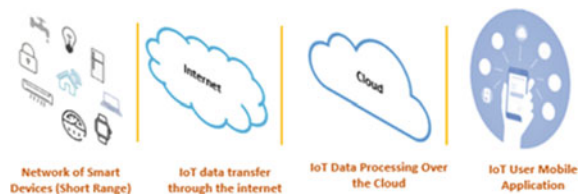
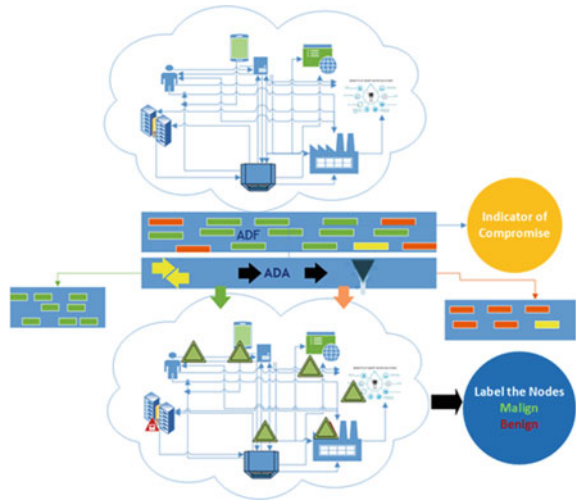


Fig. 2 The proposed model with the implementation of attack detection framework



reached the threshold, then the node will be called suspicious and the traffic will be marked using the orange color. All the traffic which is derived from the suspicious node will be watched continuously to check the signs of the attack, if it crosses the threshold then that node is marked as malign; otherwise, it is marked as benign. If any of the nodes consume more bandwidth for DODAG messages communication, then the network traffic of that node is analyzed and the statistics are applied to get the threshold values and are compared against the indicators of compromise (IOC). In this method, indicators of compromise are defined as (1) Threshold for the packet statistics count of DODAG message exchange; and (2) Bandwidth consumed for the DODAG message exchange. Traffic analyzer and attack detection framework should analyze the traffic using the IOC to derive the values for the below parameters For each node, finding the neighbors:

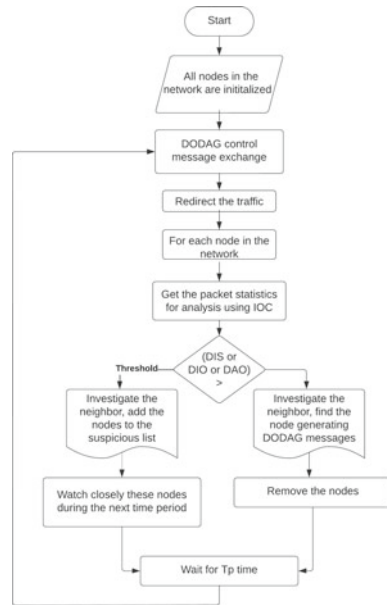
- Collecting the DODAG message exchange packet statistics for the neighbors along the path toward the root node.
- This is to check if the neighbor node is involved in any message exchange.
- Check if the neighbor’s statistics also crosses the threshold.

Periodically, analyze the live traffic to make sure that the network is safe. Apply the Attack detection algorithm mentioned below during the traffic analysis as shown in Fig. 3.

3.3 Attack Detection Algorithm

The algorithm is integrated to run in the attack detection framework, which plays a vital role in the implementation of the proposed model.

Fig. 3 Flow diagram of the attack detection framework process



3.4 Efficiency of Algorithm

Best case: Best-case efficiency of the ADF algorithm can be achieved when the malicious node is near the root. If there are N nodes in the network. And all the nodes fall within the frequency range of the root node. Get the count of DODAG control messages communicated by each node. Count from each node can be used to determine if there is any attack or not. So, the best-case efficiency is $\Omega(N)$.

Average case: Average-case efficiency of the ADF algorithm can be achieved when the malicious node is residing a few nodes away from the root node. If there are N nodes in the network and all the nodes do not fall within the frequency range of the root node then get the count of DODAG control messages communicated by each node. For each node, get the count of DODAG control messages communicated by its neighbor nodes. $1 + 2 + 3 + \dots + N = \sum N = \frac{N(N+1)}{2} = \theta(N^2)$ Count from each node can be used to determine if there is any attack or not.

Algorithm 1 Attack Detection Algorithm

Assume T_p constant value is already defined (T_p – Time period, T_h – Threshold count)

Step 1: Initialize

All nodes, $SN = \{\text{All the nodes}\}$

Suspicious nodes, $SN = \{\phi\}$

Malign nodes, $MN = \{\phi\}$

Step 2: Collect the packet statistics for each node.

For, {each i th node in the network}:

// Collect the Packet statistics including,

Count of DIO messages sent and received, $NiSDIO$, $NiRDIO$

Count of DIS messages sent and received, $NiSDIS$, $NiRDIS$

Count of DAO messages sent and received, $NiSDAO$, $NiRDAO$

Step 3: Identify the nodes which generate too many DIO, DIS, and DAO packets

//Packet statistics count collected for each node in step 2 is compared with T_h count

For {each i th node in AN }:

If ($NiDAO$ sent message count $> T_h$)

// Find the node N_j which generated the DIS messages.

updateservice_chain[x]

$MN = MNU\{N_j\}$

If($NiDIS$ sent count $> T_h$)

Add the nodes to the MN list

$MN = MNU\{N_i\}$

If($NiDIS$ OR $NiDOS$ OR $NiDAO \cong T_h$) // if almost equal

Add the nodes to the SN list:

$SN = SN \cup N_i$

Step 4: if (SN is not empty)

Increment T_p by some value and observe the packet statistics for all node of SN list, repeat the steps 2 to 3

If (MN is not empty)

For {each i th node in MN }:

// Find the source of the packet generation

If (DAO sent message count is more)

// Find the node which generated the DIS messages.

$MN = MN \cup \{N_i\}$

Action: Block the traffic generated by the nodes in the list MN .

END

=0

Worst case: Worst-case of the ADF algorithm can be achieved when the malicious node is farthest from the root node compared to all the other nodes. Worst-case efficiency is the same as the average case. That is, $\theta(N^2)$: If there are more than one malicious node. For each node, get the neighbor nodes. For each neighbor node, get the count of DODAG control messages with their neighbors. Worst-case efficiency would be: $O(N^N)$.

4 Results and Discussion

Section 4 presented the complete process and techniques of the proposed attack detection framework. In this section, we discuss the simulation setup and results of the proposed work. Consider an IoT network forming an RPL topology. The network contains a data transmission from a child node to the root node, forming a DODAG routing topology. To form the DODAG topology, each node exchanges DODAG messages DIO, DIS, and DAO to form the topology with parent/child relationship.

4.1 Simulation Setup

IoT simulation setup includes,

1. Figure 4 shows an IoT network topology with all genuine nodes,
 - One root node is also called sink node.
 - 20 sensor nodes which are randomly placed in the IoT wireless sensor network.
2. Figure 5 shows an IoT network with a malicious node,
 - One root node is also called sink node.
 - Green color node is a sink/root node.
 - 20 sensor nodes which are randomly placed in the IoT wireless sensor network.
 - Yellow color nodes are the sensor nodes.
 - One malicious node which is randomly placed in the network.
 - Pink color node is a malicious node.
3. Each node will run the process inside it to configure the node and network parameters.
 - Node ID for each node is set.
 - CSMA channel is initiated, with the channel parameters CSMA Contiki MAC, channel check rate 8 Hz, radio channel 26.
 - Dynamic link-local IPv6 address is assigned to each node.
 - Sensor process is started.
 - Each node will start listening on the port.
 - Time to live field value is set to 64.

DODAG topology construction will be initiated as presented in Fig. 4. As part of the construction, nodes will send DIS or DIO messages to all the nodes, and the nodes that are in frequency will receive message and acknowledge it by sending DAO messages. Each node sends DIS messages to discover its neighbors using Neighbors Discovery Protocol's (NDP) solicited-node multicast addresses. Here, the solicited-node multicast address is `ff02::1a`. In the simulation, each node is configured to multicast DIS/DIO messages to all the IPv6 devices and all IPv6 routers [25] which are within each device's range. And the devices will respond with their information

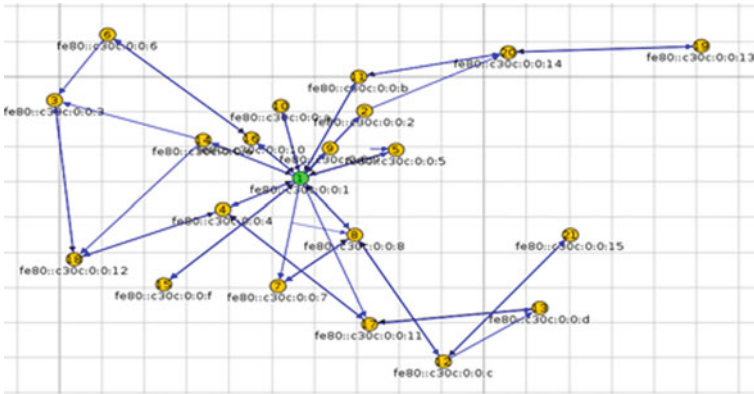


Fig. 4 An IoT network topology with all genuine nodes

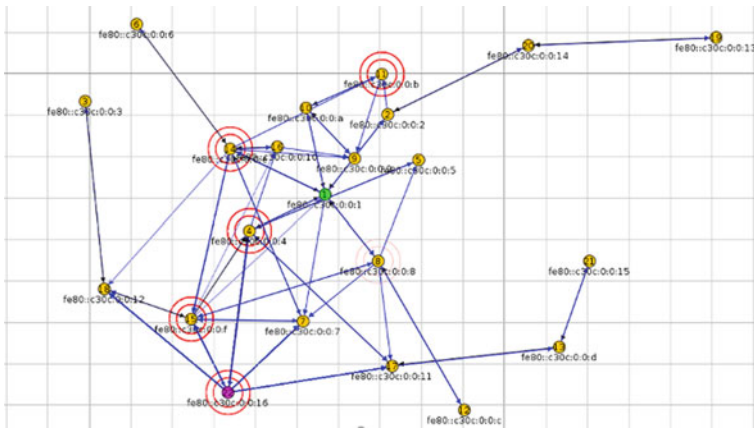


Fig. 5 An IoT network topology with a malicious node

in DAO messages. In our simulation framework, we choose IPv6 protocol for connecting the IoT nodes since few best research works proved that IPv6 with IoT is great deal of strategy in connecting billions of devices in large network. Additionally, we found that the payload and frequency of the large IoT network. In general, the IoT benefits with IPv6 in large-scale deployment [26, 27] and hence to match the reality payload we use IPv6 addresses.

A simulation setup is done with a malicious node inside which is depicted in Fig. 5. The intention of the malicious node is to bring down the network and communication among the nodes so that the genuine nodes are denied by the service access. The attacker uses this property to insert a malicious node in the IoT network to bring down the network to serve genuine requests. The malicious node starts sending the DIS messages to all its neighbors using the solicited-node multicast address ff02::1a. Through this, all the nodes and routers fall within the sender node's network range and

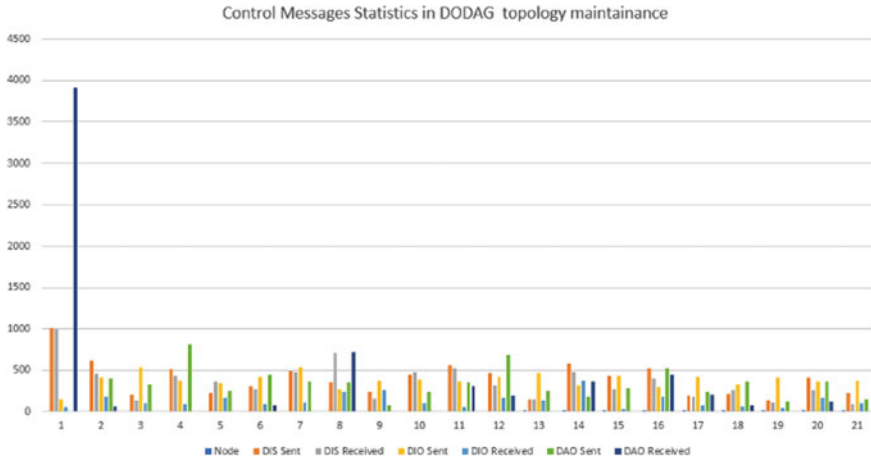


Fig. 6 DODAG control message statistics with all genuine nodes

receive the DIS messages and respond by sending the DAO messages. The malicious node is configured to flood the DIS messages so that all the nodes will be busy in processing the DIS and DAO messages. Most of the network bandwidth will be used to process the control messages than processing the actual data transmission in the network.

The graph in Fig. 6 is established using the statistical data, this clearly indicates that there is a peak in DAO messages received by the root node. This is expected in an attack-free network as the root node/sink node is responsible for maintaining the network topology and sinking all the nodes in the network and keeping the network updated. So, the root node receives the DAO messages from the neighbors periodically. DAO is received for each DIO and DIS message sent by the root node. And also, from the graph, we can clearly make out that root didn't send DAO messages. This is expected because the root node is the parent of all the nodes and is assigned with the highest rank [28].

Malicious node is Node 22 and its neighbor nodes are Node 7, Node 15, Node 14, and Node 18.

$$MN = 22$$

$$MNN = 7, 15, 17, 18$$

The graph in Fig. 7 which is plotted using the statistical data clearly shows that there is a peak in the number of DODAG control messages exchanged by each node in the MN and MNN. The graph shows that Node 22 has sent more DIS messages, and Nodes 7, 15, 17, and 18 have received more DIS messages. Also, it shows that Node 22 has not received any DAO messages, because the malicious node does not want to keep itself busy processing the received messages.

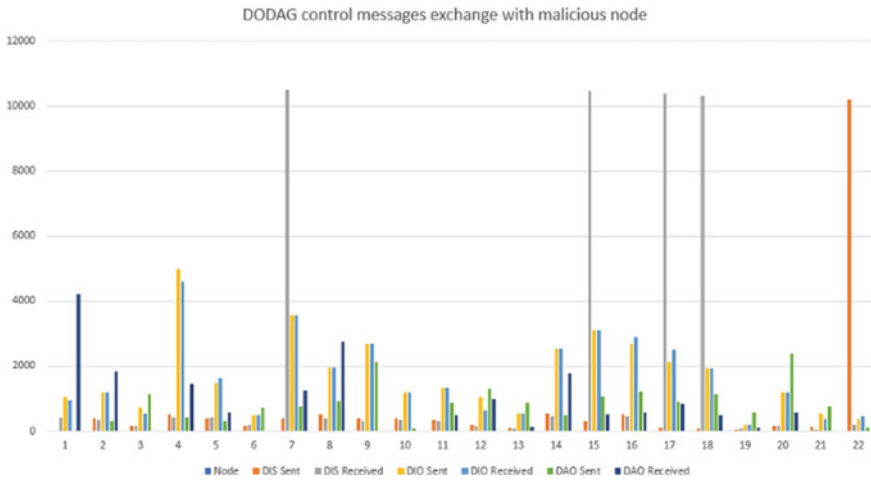


Fig. 7 DODAG control message statistics with a malicious node (Node 22 is malicious)

The intentions of the malicious node:

- Flood the network with DIS messages by sending DIS messages.
- Make its neighbors to be busy in processing the flooded messages.
- Make its neighbors to contribute in flood without their knowledge by making the neighbors to respond with DAO messages.
- Influences its neighbors to propagate this flooding by sending more DIS messages to their neighbors.
- This flooding is destined towards the root node along the path from the malicious node.

Its intention is to flood the network with DIS and DIO control messages and make other nodes to be busy with processing the flooded messages and also make its neighbor to flood with the DAO messages. So, most of the bandwidth, power consumption, and other resources will be invested in exchanging DODAG control messages. From the above graph, it clearly indicates that even the root node has sent a greater number of DIO messages and received more DAO messages. This is required, as the root node ensures that the DODAG topology is up-to-date and this is the node which sinks all the nodes in the network. Figure 8 is the power tracking information of all the genuine nodes in the network. Each node’s power consumption to receive and send the packets is very less as the information required to transmit or receive is only for exchanging the DODAG control messages.

- Radio On: Power required to keep the receiving or transmitting machinery active all the time.
- Radio Tx: Power required for the chip to transmit the PHY layer packets.
- Radio Rx: Power required for the chip to receive the PHY layer packets.

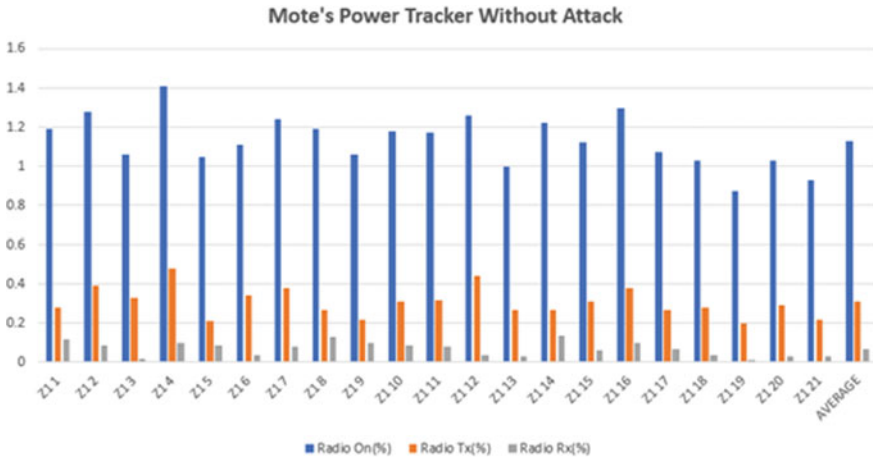


Fig. 8 Power tracking for radio On, Tx, and Rx without a malicious node

In this simulation work, we assumed no additional device inference and power over Ethernet (PoE) activities that consume power and has no connection with the IoT node power consumption. All IoT sensor nodes are refreshed over a preset timer in order to understand there is no internal inference in communication and power channels. This work computed 7 different simulation types, out of which the best is illustrated in the table with RX, TX, and base parameter ON (Table 1).

Figure 9 shows that power consumption by the malicious Node 22 is very high, especially for transmitting the packets, and the radio on is very high as the malicious node transmits more DODAG control messages to keep all the nodes in the network busy by processing the unwanted messages. Power consumption [29] for the reception is less as the malicious node does not keep itself busy by processing the DAO messages. The power consumption is high for Nodes 7, 15, 17, and 18, as these are the direct neighbors of Node 22. There is little more power consumption by other nodes as compared to the IoT network without the malicious node as they are the neighbors of Nodes 7, 15, 17, and 18.

5 Tool Support

Simulation is performed using the Contiki Cooja simulator. The Cooja simulator is a network simulator designed, especially for Wireless and IoT network simulation. The simulator facilitates to add 'N' number of devices with different network configurations, different network topologies based on the user requirements and the nodes within the specific radio range can communicate to perform the specific task. The simulator also provides option to change the parameters such as logging and capturing the packets that can be further used for analysis purposes. This helps to test

Table 1 Power tracking for radio On, Tx, and Rx without a malicious node

Mote	Radio On (%)	Radio Tx (%)	Radio Rx (%)
Z1 1	1.19	0.28	0.12
Z1 2	1.28	0.39	0.09
Z1 3	1.06	0.33	0.02
Z1 4	1.41	0.48	0.1
Z1 5	1.05	0.21	0.09
Z1 6	1.11	0.34	0.04
Z1 7	1.24	0.38	0.08
Z1 8	1.19	0.27	0.13
Z1 9	1.06	0.22	0.1
Z1 10	1.18	0.31	0.09
Z1 11	1.17	0.32	0.08
Z1 12	1.26	0.44	0.04
Z1 13	1.00	0.27	0.03
Z1 14	1.22	0.27	0.14
Z1 15	1.12	0.31	0.06
Z1 16	1.3	0.38	0.1
Z1 17	1.07	0.27	0.07
Z1 18	1.03	0.28	0.04
Z1 19	0.87	0.2	0.01
Z1 20	1.03	0.29	0.03
Z1 21	0.93	0.22	0.03

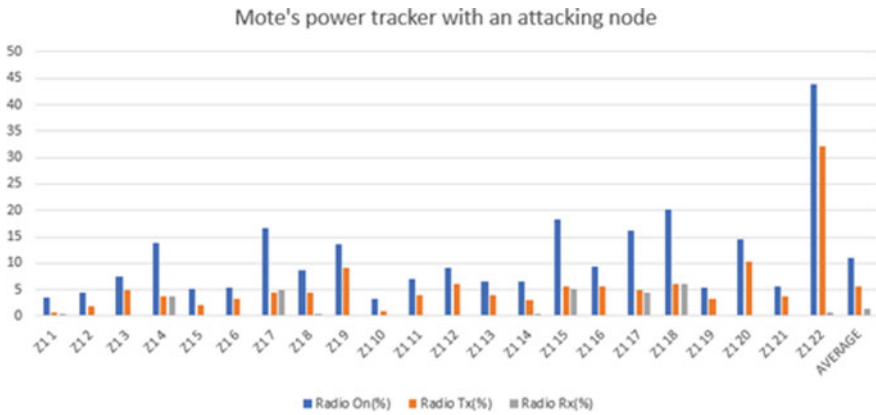


Fig. 9 Power tracking for radio On, Tx, and Rx with a malicious Node 22

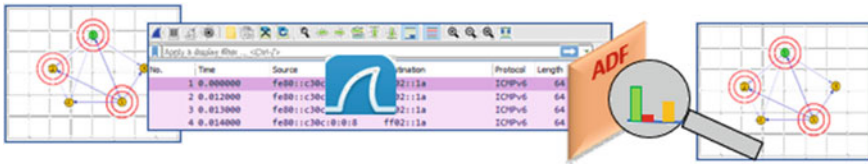


Fig. 10 Tools used in the implementation of attack detection framework

and debug the different protocols of the IoT network stack, and also to test the IoT device properties like battery performance, power consumption, bandwidth usage, and memory usage [30]. This can be used to test and debug different attack scenarios using the threat modeling.

As shown in Fig. 10, the IoT network traffic is captured and redirected to the Wireshark tool which is a network traffic and packet analysis tool. This filtered traffic from the Wireshark is used as an input to the attack detection framework for further analysis. The ADF analyzes the traffic deeply to check if there is a malicious packet traffic and the source of attack.

6 Conclusion

This paper introduced an algorithm with an attack detection framework to detect the threats posed in the IoT sensor network. As the devices in the IoT network are resource constrained, the DDoS attacking technique targets to drain the device resources to be used for non-legitimate activities. The IoT network topology construction uses the DODAG control message exchange which an attacker can use to flood the messages in the network posing as a legitimate node in the network. The DODAG message exchange packet traffic is captured, summarized, and analyzed over a period of time. This attack detection framework focused to discover the indicators of compromise and to check whether the node is malicious in the network. The work is primarily focused to discover the power consumption and network bandwidth usage throughout the simulation to check the signs of attack. The process is repeated for every predefined interval time to ensure network security.

Acknowledgements Research was sponsored by the Army Research Office and was accomplished under Grant Number W911NF-21-1-0264 and W911NF-21-1-0201. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation herein.

References

1. Kuo YW, Li CL, Jhang JH, Lin S (2018) Design of a wireless sensor network-based IoT platform for wide area and heterogeneous applications. *IEEE Sens J* 18(12):5187–5197. <https://doi.org/10.1109/JSEN.2018.2832664>
2. Khan AR, Kashif M, Jhaveri RH, Raut R, Saba T, Bahaj SA (2022) Deep learning for intrusion detection and security of Internet of things (IoT): current analysis, challenges, and possible solutions. *Secur Commun Netw* 2022
3. Akram J, Akram A, Jhaveri RH, Alazab M, Chi H (2022) BC-IoDT: blockchain-based framework for authentication in internet of drone things. In: *Proceedings of the 5th International ACM Mobicom workshop on drone assisted wireless communications for 5G and beyond*, pp 115–120
4. Trnka M, Cerny T, Stickney N (2018) Survey of authentication and authorization for the internet of things. *Secur Commun Netw* 2018. <https://doi.org/10.1155/2018/4351603>
5. Lohachab A, Karambir B (2018) Critical analysis of DDoS—an emerging security threat over IoT networks. *J Commun Inf Netw* 3(3). <https://doi.org/10.1007/s41650-018-0022-5>
6. Shaaban AR, Abd-Elwanis E, Hussein M (2019) DDoS attack detection and classification via Convolutional Neural Network (CNN). In: *Proceedings of the 2019 IEEE 9th International conference intelligent computing information system (ICICIS 2019)*, pp 233–238. <https://doi.org/10.1109/ICICIS46948.2019.9014826>
7. Srinivas TAS, Manivannan SS (2020) Prevention of hello flood attack in IoT using combination of deep learning with improved rider optimization algorithm. *Comput Commun* 163:162–175. <https://doi.org/10.1016/j.comcom.2020.03.031>
8. Luong NT, Vo TT, Hoang D (2019) FAPRP: a machine learning approach to flooding attacks prevention routing protocol in mobile ad hoc networks. *Wirel Commun Mob Comput* 2019. <https://doi.org/10.1155/2019/6869307>
9. Anbarasan M et al (2020) Detection of flood disaster system based on IoT, big data and convolutional deep neural network. *Comput Commun* 150:150–157. <https://doi.org/10.1016/j.comcom.2019.11.022>
10. Alsamiri J, Alsubhi K (2019) Internet of things cyber attacks detection using machine learning. *Int J Adv Comput Sci Appl* 10(12):627–634. <https://doi.org/10.14569/ijacsa.2019.0101280>
11. Kalinowska-Górska K, Solano Donado F (2014) Constructing fair destination-oriented directed acyclic graphs for multipath routing. *J Appl Math* 2014. <https://doi.org/10.1155/2014/948521>
12. Tian H, Qian Z, Wang X, Liang X (2017) QoI-aware DODAG construction in RPL-based event detection wireless sensor networks. *J Sensors* 2017. <https://doi.org/10.1155/2017/1603713>
13. Boro D, Bhattacharyya DK (2017) DyProSD: a dynamic protocol specific defense for high-rate DDoS flooding attacks. *Microsyst Technol* 23(3):593–611. <https://doi.org/10.1007/s00542-016-2978-0>
14. Jia Y, Zhong F, Alrawais A, Gong B, Cheng X (2020) FlowGuard: an intelligent edge defense mechanism against IoT DDoS attacks. *IEEE Internet Things J* 7(10):9552–9562. <https://doi.org/10.1109/JIOT.2020.2993782>
15. Umer MF, Sher M, Bi Y (2017) Flow-based intrusion detection: techniques and challenges. *Comput Secur* 70:238–254. <https://doi.org/10.1016/j.cose.2017.05.009>
16. Pham TND, Yeo CK, Yanai N, Fujiwara T (2018) Detecting flooding attack and accommodating burst traffic in delay-tolerant networks. *IEEE Trans Veh Technol* 67(1):795–808. <https://doi.org/10.1109/TVT.2017.2748345>
17. Luo H, Chen Z, Li J, Vasilakos AV (2017) Preventing distributed denial-of-service flooding attacks with dynamic path identifiers. *IEEE Trans Inf Forensics Secur* 12(8):1801–1815. <https://doi.org/10.1109/TIFS.2017.2688414>
18. Sahi A, Lai D, Li Y, Diykh M (2017) An efficient DDoS TCP flood attack detection and prevention system in a cloud environment. *IEEE Access* 5(c):6036–6048. <https://doi.org/10.1109/ACCESS.2017.2688460>

19. Fang X, Yang M, Wu W (2018) Security cost aware data communication in low-power IoT sensors with energy harvesting. *Sensors (Switzerland)* 18(12):1–18. <https://doi.org/10.3390/s18124400>
20. Adina P, Venkatnarayan RH, Shahzad M (2018) Impacts & detection of network layer attacks on IoT networks. In: *Proceedings of the 1st ACM MobiHoc workshop on mobile IoT sensing, security privacy, mobile IoT SSP 2018*. <https://doi.org/10.1145/3215466.3215469>
21. Wallgren L, Raza S, Voigt T (2013) Routing attacks and countermeasures in the RPL-based internet of things. *Int J Distrib Sens Netw* 2013. <https://doi.org/10.1155/2013/794326>
22. Verma A, Ranga V (2020) Security of RPL based 6LoWPAN networks in the internet of things: a review. *IEEE Sens J* 20(11):5666–5690. <https://doi.org/10.1109/JSEN.2020.2973677>
23. Kühn F, Hellbrück H, Fischer S (2018) A model-based approach for self-healing IoT systems position paper. In: *SENSORNETS 2018—Proceedings of the 7th International conference sensors networks*, vol 2018, *Sensornets*, pp 135–140. <https://doi.org/10.5220/0006639401350140>. https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html#
24. Yang H, Lee W, Lee H (2018) IoT smart home adoption: the importance of proper level automation. *J Sensors* 2018. <https://doi.org/10.1155/2018/6464036>
25. Wong KS, Wan TC (2019) Current state of multicast routing protocols for disruption tolerant networks: survey and open issues. *Electron* 8(2). <https://doi.org/10.3390/electronics8020162>
26. Ziegler S, Crettaz C, Ladid L (2013) IoT6—moving to an IPv6-based future IoT. Part of the *Springer Lecture Notes in Computer Science* book series (LNCS, vol 7858). *The Future Internet Assembly, The Future Internet* pp 161–172
27. Savolainen T, Soininen J, Silverajan B (2013), IPv6 addressing strategies for IoT. *IEEE Sensors J* 13(10):3511–3519, INSPEC Accession Number: 13736133, <https://doi.org/10.1109/JSEN.2013.2259691>
28. Tutunović M, Wuttidittachotti P (2019) Discovery of suitable node number for wireless sensor networks based on energy consumption using Cooja. In: *International conference advances communication technology ICACT*, vol 2019, pp 168–172. <https://doi.org/10.23919/ICACT.2019.8702021>
29. Collotta M, Ferrero R, Rebaudengo M (2019) A fuzzy approach for reducing power consumption in wireless sensor networks: a testbed with IEEE 802.15.4 and wireless HART. *IEEE Access* 7:64866–64877. <https://doi.org/10.1109/ACCESS.2019.2917783>
30. Mahmud A, Hossain F, Juhin F, Choity TA (2019) Merging the communication protocols 6LoWPAN-CoAP and RPL-CoAP: simulation and performance analysis using Cooja simulator. In: *1st International conference advances science engineering robotic technology 2019 (ICASERT 2019)* vol 2019, pp 1–6. <https://doi.org/10.1109/ICASERT.2019.8934540>

Application of Digital Forensic Evidence in Hit and Run: A Comparative Study with Special Reference to § 304 Part II of IPC



Hiral Thakar and Manav Kothary

Abstract The developments of the new era in technology have reflected change in the existing system of investigating and judicial system with regard to evidence admissibility and reliance by the judiciary across jurisdictions. The criminal-jurisprudence has witnessed notable transformation with the use of scientific tools, procedures, and methodologies, which has become increasingly pivotal in delivering justice to the aggrieved. India accounts for 11% of all fatal traffic accident deaths worldwide despite owning only 1% of all vehicles and even fewer than 1% conviction rate, due to lack of proper evidence. How the digital forensic evidence in helping the legal system to encase allegations made under § 304A of the IPC to under § 304 Part II of the IPC is conveyed in this paper with the help of hit and run case instances. In addition to it, this paper also aims to establish the veracity and admissibility of the digital forensic evidence in the court of law and how the examination of its role has a different impact on two mostly identical cases with different investigation outcomes. In one instance, the accused was convicted due to the evaluation of digital evidence while in other he was acquitted due to lack of any proper forensic evidence. The authors have highlighted the challenges before the judges in deciding a criminal case or settle a legal dispute. Relying solely on witness testimony, who are seldom tutors or could not be relied upon might jeopardize the entire case proceedings. Thus, the authors argue for using forensics evidences to reconstruct the crime screen and corroborate the witness/victims testimony.

Keywords Hit and run · Digital forensic evidence · Closed-circuit television footage · Validating technique · Criminal-jurisprudence

H. Thakar (✉)

School of Law, Forensic Justice and Policy Studies, National Forensic Sciences University, Gandhinagar, Gujarat, India

e-mail: 101flmlcc2122015@nfsu.ac.in

M. Kothary

United World School of Law, Karnavati University, Gandhinagar, Gujarat, India

1 Introduction

Technology has rapidly evolved and advanced throughout time. As a result, it is crucial to adapt and alter in line with technological advancements. Courts and legislature have made it quite clear that they see the need for a change. For example, a crime has been committed in the darkness of night which leaves no room for any man to witness such incidence but the culprit and the perpetrator. When there is no one available to confirm the accuser's identity, circumstantial evidence may be used to demonstrate that a crime was committed. Nevertheless, video monitoring will allow the investigators to see the entire occurrence. The film includes descriptions of the event schedule, the criminal's tactics, and the attacker's entry and exit points. The CCTV footage can be used to disprove other types of evidence, such witness testimony, even though there are many reasons why this is not feasible. When a suspect is recognized, for example, or when the criminal comes into contact with a substance from which forensic evidence may be obtained, for example, the video can assist detectives in identifying those who were engaged in the incident directly or indirectly. Investigators can still use the recordings to determine the criminal's timeframe by using them. Investigators will utilize the CCTV footage to examine the veracity of the witnesses' and suspects' statements. The traditional method of the criminal justice system based on eyewitness testimony has made effective criminal prosecutions nearly impossible; this form of prosecution undermines the criminal justice system. Judges cannot decide a criminal case or settle a legal dispute based solely on the testimony of potential spinners of yarn or unreliable witnesses. The witnesses failed to appear at the scheduled hearings or refused to submit to the legal process. Even after carefully considering the main question and the counter-questions, the judges are unable to reach a firm verdict regarding the incident. Many times, despite their belief or having seen the evidence, the witnesses no longer appear before the court to give testimony because they fear being attacked by suspects or other criminals, which could be fatal. The public ministry spends a substantial amount on criminal court proceedings. Due to a lack of conviction or strong evidence, many horrific criminals are found guilty or let off the hook based on even the most remote suspicion. As a result, traditional legal actions typically result in the loss of the vast majority of public resources and the exoneration of the accused on the basis of presumed innocence. Today's crimes are technologically perpetuated, and crimes based on science are developing that can only be solved with forensic technology.

Our ancestors were alive, they appear to have used scientific methods in some degree when conducting criminal investigations. *Kautilya's "Arthashastra"* [1] contains a comprehensive reference on this subject and was written approximately 2300 years ago. Since the 1930s, law enforcement, courts, and several anti-crime commissions throughout the world have all agreed to rely more on science to fight crime and enforce the law. The studies of the "*President's Crime Commission Task Force reports on Police and Science and Technology*" [2], which were published in the 1960s, called for a greater use of scientific evidence in criminal investigations and legal proceedings. The use of digital technology in criminal cases has resulted in

significant advancements in the decision-making process. Forensic Science Laboratories use cutting-edge scientific techniques to assess and dissect physical evidence. The increasing use of forensic evidence in court may be attributed to a variety of factors, including the spate of fugitive crime cases in India where many suspects frequently dropped their charges due to a lack of scientific support.

The Information Technology Act [3] (hereinafter referred to as IT act) has recognized and defined electronic records as “*data, record or data generated, image or sound stored, received or sent in an electronic form or micro film or computer-generated micro fiche.*” The Indian Evidence Act allows the use of electronic records as evidence [4]. This paper offers solutions to a number of significant issues by drawing on two closely related legal traditions, including:

- i. Is digital evidence considered to be substantial or corroborated evidence in legal proceedings?
- ii. In what circumstances did the court take into account the digital-legal evidence?
- iii. Has exoneration been confirmed after the acceptance of digital evidence result in a verdict?

Regression analyses demonstrated that digital evidence played a consistent and robust role in case-processing decisions. This paper examined the role and impact of digital evidence on case-processing outcomes by comparing two famous case laws related to hit and run, namely, *Vismay Amitbhai Shah v. State of Gujarat* [5] and *Salman Salim Khan v. State of Maharashtra* [6]. Wherein one case digital evidence paved the way of the wrongdoer to imprisonment and while in other lack of digital evidence set him free.

2 NCRB Report

The eighth place among the causes of death on a global scale is occupied by traffic accidents. It is concerning how frequently accidental deaths occur in India. According to the National Crime Records Bureau gathered information on “Traffic Accidents” in 2021 [7], traffic accidents resulted in 373,884 injuries and 173,860 fatalities. In India, traffic accidents cost 1.55 lakhs lives in 2021; the average number of deaths are 18 per hour, or 426 per day [8]. According to the 2019 World Bank report, “*Guide for Road Safety Opportunities and Challenges: Low- and Middle-Income Countries Country Profiles,*” India accounts for 450,000 accidents per year that claim the lives of 150,000 people which is 11% of all fatal traffic accident deaths worldwide despite owning only 1% of all vehicles, it the highest percentage in the world, there are 53 traffic accidents nationwide every hour, resulting in one fatality every four minutes [9]. Fewer than 1% of these cases result in a conviction, while other cases build up over years, there can be n no. of reasons for this. It takes two to three years for the cases involving fatal accidents to be brought before the courts. The police are also lackadaisical with filing their report; it takes between three and six months to file an accusation [10]. More than 80% of highway accident cases result in acquittals.

Obtaining a conviction in the 20% that remain is also a mammoth task since our judicial system is overburdened with fewer hands to aid it [11]. One of the main causes of the delay in accident cases is the lack of eyewitness testimony for a variety of reasons, including the fact that people seldom show up for court appearances out of fear of a drawn-out trial. It is pertinent to note that, NCRB or the World Bank has not incorporated in their statistics, the aspect, where any sort of digital and/or forensic evidence was found in the crime scene and whether further investigation by the police was done in the line of digital or forensic investigation and if yes, does it have given better result than relying just upon eyewitness testimony. It is also interesting to note that the inclusion of forensic evidence in criminal investigation and trial for both lawyers and police is not something common and this lack of connection between these two subjects is also a reason for low conviction and due to poor investigation.

3 CCTV: Electronic Witness

Human witnesses can change from their stand before police and magistrate respectively but digital witness stand by the same assertive testimony as given in the first place. The prevalence of video surveillance cameras as a deterrent to crime is becoming more and more common in society. For those who view the increased use of CCTV cameras as invasive and undermining privacy, this may be reason for alarm. Particularly in metropolitan areas, there has been a rise in the usage of CCTV cameras. The CCTV camera can and does record everything in between, even though it primarily records crimes as they are being committed or in the moments before or after such an event. Whatever one may think about the usage of CCTV cameras, the reality is that they have innumerable benefits for determining the conditions. There may have been a CCTV camera around the accident site that recorded what truly transpired in the moments before and during the collision. Additionally, there may be traffic control cameras in place that will perform the same function as above if the collision happened close to a junction. Recently, in a car accident occurred due to over-speeding the former Tata Sons chairman, Cyrus Mistry, died. According to the CCTV footage, as obtained by The Palghar police in Maharashtra of the Mercedes car in which he was traveling from Ahmedabad to Mumbai. A 20 km journey took only 9 min, according to video surveillance photographs, indicating that the luxury vehicle was traveling at a speed of 180–190 km per hour [12].

4 CCTV as Electronic Evidence

Any admission of guilt must be supported by evidence that establishes the guilt of the accused beyond a reasonable doubt. The current status of the CCTV footage can be explained by one of two separate scenarios:

1. When video surveillance recordings are the only evidence available, can they be used as foundational evidence to establish the accuser's guilt at the accuser's request?
2. What will happen to the CCTV video, if the recording and eyewitness testimony do not corroborate each other?

It will be sufficient to demonstrate the *actus-reus* at the instance of the accused in both instances, if the CCTV video is appropriate and clear and the origin of the CCTV is proven beyond a reasonable doubt. Also on a worldwide scale, the value and relevance of using CCTV as evidence in Ld. court have been noted. In "*Gubinas and Radavicius v. HM Advocate, the High Court in Aberdeen*," [13] it has been noted that, even if every witness claims one thing while the video surveillance sequence reveals something else, the electronic evidence will be cited rather than the eyewitness testimony. The Court's comment demonstrates that the content of the video surveillance sequence is deemed sufficient to establish that a crime was committed and identify the offender. In the case of *Tomaso Bruno and Associates, v State of U.P.* [14], a murder in Varanasi was committed and two other Italian citizens were found guilty of the crime. The Supreme Court has noted that video surveillance photographs are a potent piece of evidence that might have been used to prove the crime, and that the refusal of the accusation to provide the video surveillance footage casts serious doubt on their case.

5 Procedure for Admissibility of CCTV Footage as Evidence

Primary evidence is considered to be CCTV video that has been created automatically and stored on a CD, DVD, hard disk, flash memory, or USB key without the assistance of a human, that DVR will be regarded as primary evidence under § 62 of the IEA and it will not be necessary to adhere to the specifics of § 65B(4) of the IEA because the original document is being examined by the court. However, it is impossible to present the complete system to the Court if there are several cameras placed and the data is digitally stored on sizable servers. In such a circumstance, the only alternative to provide it is necessary to transfer the data from the enormous server to a CD or USB for the court, which is not feasible without human interaction and increases the possibility that the data may be falsified even somewhat. The duplicated or retransmitted evidence will be viewed as supporting evidence in such a case. If any secondary electronic evidence, such as a USB or CD, is presented as evidence, it must strictly adhere to the requirements of § 65B(4) of the Indian Evidence Law, which calls for the acquisition of a certificate from the owner of the surveillance camera attesting that the material taken from its DVR has not been altered. You must get a certificate from the server's administrator. The main goal of the certificate is to demonstrate the proper operation of the computer from which the electronic record is generated in front of the court for examination in order to show that the

content has not been altered or falsified as a result of a computer malfunction. The certification is not necessary to demonstrate the veracity of the information contained in the computer-generated dossier. The main goal of § 65B is to accept secondary evidence as evidence. According to a panel of three judges in the “*Arjun Panditrao Khotkar case v. Kailash Kushanrao Gorantyal*” [15], the certification required by § 65B(4) of IEA is a pre-requisite for the acceptance of evidence presented through an electronic record. The Court has further emphasized that the certificate required by § 64B(4) IEA is not necessary, if the original document is submitted to the Court for review.

6 § 304 of IPC

According to the gravity and seriousness of the offence, culpable homicide not amounting to murder is divided into two categories under § 304 of IPC, in order to determine the appropriate punishment. § 304 of IPC does not create new offences but instead establishes a range of possible sentences depending on the gravity and nature of the offence, namely,

- (a) When the conduct would have constituted a murder, it would not have been because it falls within one of the exceptions to the § 300 of IPC. Alternatively, the purpose serves as the foundation for determining responsibility under the provisions of Part I of § 300 of IPC, “*when these bodily injuries are caused with the **intent** to cause death or with the intent to inflict such bodily injury that is likely to cause death.*” The provisions of part I provide for either perpetual imprisonment or temporary imprisonment for one or the other description for a period of up to ten years and fine.
- (b) The foundation of the punishment in Part II of sec. 300 of IPC is **knowledge**. It applies to anybody who dies after carrying out an act knowing that it would cause death, as specified in cl. (3) of § 299 of the IPC, as well as to anyone, who is not covered by cl. 4 of § 300 of IPC but is covered by one or more exceptions to § 300 of IPC. The second part of this § specifies a punishment that might last up to 10 years or a fine.

7 § 304A of IPC

Initially, the penal law did not foresee punishment in cases when someone would negligently cause the death of another person, and responsibility for having caused the death only applied in cases of murder and culpable homicide, which do not constitute murder. The IPC was amended by adding § 304A of IPC in 1870 through The IPC Amendment No. 27 of 1870, which covers cases in which someone kills another person due to reckless or negligent behavior without intending to do so and

without knowing that doing so would result in death. The § 304A of IPC addresses homicide through careless and momentary acts.

The two principles of criminal liability in cases of negligence [16] are

1. The circumstances must be such that the accused's negligence went beyond a straightforward discussion of indemnification between the parties and demonstrated such disregard for the lives and safety of others that it was elevated to the level of an offence against the State and behavior subject to punishment.
2. The phrase "imprudence" is frequently used to describe a lack of concern for a risk, even though the accused may have known about the risk and intended to avoid it while demonstrating enough negligence to warrant a penalty for the methods used to do so.

Essential Ingredients According to IPC § 304A, the following requirements must be met in order to prosecute a murder case:

- (1) The victim must die;
- (2) The act done by the accused must have caused this death; and
- (3) That the accused's alleged act was careless or episodic and did not amount to a culpable homicide.

Additionally, it has been decided that it must be the *causa causans*, meaning that the immediate or operational reason is insufficient to qualify as the *causa sine qua non*, or a necessary or unavoidable cause [17].

8 Rash and Negligent Act

The act of taking a chance that unfavorable consequences would follow but yet hoping that they won't materialize is known as Rashness. Contrarily, negligence is the failure to fulfill a legal responsibility [18]. Criminal negligence is serious and punishable, as is negligence or the failure to take reasonable and necessary precautions to prevent harm from occurring to anyone, whether it be the general public or a specific individual who, given the circumstances surrounding the accusation, was required to be adopted. In the *State of Karnataka v. Sharanappa Basanagouda Aregoudar* [19] four people had perished in an accident determined that, given the gravity of the case, Balakrishnan J, speaking on behalf of the Court expressed that:

The courts need to be on guard for the accused if it is determined that their behavior was reckless and negligent so that they don't very easily evade the law. The punishment imposed by the courts must deter potential offenders and be proportionate to the seriousness of the offense. Certainly, judges have the discretionary power to weigh a wide range of potentially relevant facts when determining the appropriateness of a sentence, but this power must be exercised with due consideration for the interests of the society as a whole. It is unnecessary to add that the criminal justice system's conviction of offenders is perhaps its most visible face.

The difference between IPC § 304A and § 304 is that the former completely excludes the culpable homicide, but the latter, whether in Part I or Part II, asserts that such homicide does not constitute murder. In other words, the § 304 states that even though the conduct is not murder, it must still be a culpable homicide, whereas the § 304A states that anything that is not a culpable homicide and is committed in a certain way as specified in the § is punishable in accordance with the § 304A of the IPC.

9 Salman Salim Khan v. State of Maharashtra [6]

9.1 The Case of the Prosecution

On the night of 27th September 2002 at around 9:30 pm, Salman Khan was accompanied by his friend Kamaal Khan, his bodyguard Ravindra Patil and Sohaib Khan and his bodyguard Sohail Khan and his bodyguard to the Rain Bar and allegedly consumed alcohol. Thereafter around 1:30 am, Salman Khan, who was accompanied by his friend Kamaal Khan and bodyguard Ravindra Patil to J.W. Marriot hotel, from where Salman Khan was on wheels. At around 2:15 am they left J.W. Marriot hotel for Galaxy Apartment. At around 2:45 am, Salman Khan was on the influence of alcohol, when he was on wheels and then dashed the huge white Toyota land cruiser on American Express Bakery heading from St. Andrews Road, killing one person and injured four.

9.2 But According to Salman Khan's Version

On the night of 27th September 2002 at around 9:30 pm, Salman Khan was accompanied by his friend Kamaal Khan, his bodyguard Ravindra Patil his driver Altaf who was driving his car, and Sohaib Khan and his bodyguard went to Rain Bar. Thereafter, Altaf drove Salman Khan, his friend Kamaal Khan, his bodyguard Ravindra Patil to J.W. Marriot hotel. And there because Altaf was feeling ill, Salman called driver Ashok Singh and around 2:15 am, they left J.W. Marriot hotel for Galaxy hospital. At around 2:45 am, during their way, from Manuel Gonsalves Rd, the left-frontside tire burst while negotiating a right turn. Hence, the car breaks into American Express Bakery and because the left front side was badly damaged, Salman Khan had to come outside from the driver's gate.

Initially charges were framed under S. 304A for negligent driving and proceedings was initiated before Metropolitan Magistrate. But when the trial commenced and witnesses including police eyewitness were examined, charges were framed under S. 304, Part II by Metropolitan Court. But because the prosecution failed to establish:

1. Any sort of digital or forensic evidence in this case.
2. The papers that vanished: Only nine original copies of the 63 witness declarations were made available to the court as of July 25, 2014. Following this audience, the accusation repeatedly requested adjournments while taking its time to locate the materials. The allegation presented all but one of the case's "vanished" papers to the sessions court on September 12. According to Special Prosecutor Pradeep Gharat, the journals were discovered in the office of the Administrative officer for the position of Bandra Police [20]. Gharat stated during the submission of papers that the only missing document was the declaration of one of the witnesses, which had already been filed.
3. Enchantment of blood: Another key witness in the case in December 2014 was medical expert witness Dattaray Bhalshankar, a chemical analyzer who testified in court that Khan's pre-incident blood alcohol content test came back positive for 62 mg of alcohol. It was said that this exceeded the Motor Vehicle Act's 30 mg per 100 ml blood limit by more than twice. However, during his counter-interrogation, the witness was unable to recall how he had used the sample method of analysis. Additionally, the defense stated that he was not an expert and had not taken the necessary precautions.
4. Was the laboratory up to date? The State of Kalina's medical-legal laboratory, where Khan's blood samples were sent, does not have the necessary authorization. Would that involve proving it and chemically analyzing blood emulsions? Typically, even if credit is advised and preferred, the absence of credit hasn't prevented people from accepting or appreciating its results, unless the chain of custody, manipulation, packaging, and delivery of the sample was deemed inappropriate. In the case of Khan, the defense attorney Shrikant Shivade stated that the laboratory was not accredited and maintained that the sample had not been transferred from the hospital to the laboratory with the proper manipulation and ingredients. However, the judge of the accelerated sessions had noted in her judgement of criminal responsibility against the actor Shiney Ahuja in 2011 that "*the Kalina laboratory does not have accreditation even though the process of obtaining accreditation is ongoing*" [21]. The defense further argued that the laboratory must be accredited by the National Advisory Council for Testing and Analysis Laboratories (NABL). The accreditation certifies that the laboratory complies with international standards for competence and quality. The incident was resolved by the corps and bodyguard R. Patil.
5. Even in death, a testifier is staunch in his testimony: Salman Khan's bodyguard, former police agent Ravindra Patil, was instrumental in getting the actor convicted in the 2002 fire escape case. After the accident, Patil dialed 911 and is now the one filing charges against the actor. It played a crucial role. He didn't just make the accusation; he also testified against Khan before the lower court. Khan was drunk at the time of the collision, according to Patil's testimony in court. The lead investigator, retired ACP Kishan Sehgal, stated that he had claimed to have asked the actor to drive more slowly. However, the actor had not complied. "We have a number of witnesses to tell us what happened after the collision in this case. But we were unaware of the sequence of events before to the disaster. Patil

was in the car with Khan after they left the hotel, and she helped us tie everything together, according to Shengal.

6. The man: Ravindra Patil, a police officer, was a member of the 1998 gang. The dying days of Patil were spent in the shadow of his illness and his hopelessness. Following his testimony before the court, Patil abruptly vanished without ever being given an explanation. The police detained him in March 2006 for failing to show up for work, and later, in November, he was relieved of his duties. After learning of his tuberculosis diagnosis, his family abandoned him since his health had deteriorated over the previous year [22]. In 2007, a friend found her in the streets near Sewri and brought her to the hospital. He passed away on October 3, 2007. Before his passing, Patil told a number of people that he had never changed his account of the incident and had told the police and the court the whole truth. ‘Salman Khan was driving,’ said Kamaal Khan: Singer, Three or four days after the 2002 event, Kamaal Khan made a statement that was captured on video in Marathi in which he revealed that Salman Khan was the one who had been in charge of the erratic night. At the time of the decision against Salman in May 2015, Kamaal Khan was ordered to be in London. According to unconfirmed evidence, he could even reside in Zurich, Switzerland [23].” People were sleeping on the trottoir, Salman and Kamaal Khan left the area, according to agent Patil. Salman was so distraught that he fell twice and then absconded the place, according to one of the injured workers, who made a statement after the tragedy. This statement did not affect Salman Khan and Kamaal Khan’s relationship because the latter continued to sing for the superstar Salman Khan both after the 2002 statement and even after 2007, when Kamaal had “faded away” from Mumbai and India, as can be seen in the filmography of Kamaal Khan. In fact, Kamaal appears to have sung in just one movie without Salman.

10 Vismay Amitbhai Shah v. State of Gujarat [5]

On the night of February 25, 2013, Vismay Amitbhai Shah drove a BMW from Pakwan Cross Road to Premchand Nagar Road, close to Judges’ Bungalows, a distance of approximately 1.5 km, at an excessive speed of more than 110 km per hour, imprudently, and carelessly. He collided with a motorcycle, sending the driver and passenger flying 15 feet into the air. He then struck his vehicle, causing the tree to bend and causing significant damage to the vehicle as well. The victim died instantly, and after two days the passenger succumbed to their injuries. There have also been allegations that he left the scene of the collision without stopping, trying to help the injured, intimidating the police, or stopping to help the injured. Closed-circuit television: The court heavily relied upon a silent and digital witness, who can never go off its statement. There were several cameras, but the main camera was Lakshmi Gathiya Rath’s camera no. 03, which recorded the whole event and helped in noting the vehicle’s speed.

10.1 *Speed-Test Validating Technique by FSL, Gandhinagar* [32]

Now, the pertinent question which is raised here, is how can a CCTV camera calculate the speed of cars or motorcycles? It can be done by calculating the time it took the vehicle to go between two spots in the video, and if one knows the physical distance between them. As a result, the vehicle travels X meters in Y seconds, which may then be converted to kilometers per hour (km/h). In connection with the incidence a sealed parcel was given to the Directorate of Forensic Science, Gandhinagar by the Police inspector in due compliance with § 65B of the evidence act. The said sealed parcel contained a CD-R which whose maker was FRONtECH and its capacity was 700 MB. The CD-R containing the video from 00-09-56 am to 00-30-59 am (20:24 min) of 25th February 2013 of camera no. 03 of Lakshmi Gathiya Rath. The first thing which was checked by the FSL was whether there was any editing/alteration/modification in the said CD-R or not but it was found that there were no editing/alteration/modification in the CD-R file and with proper hash value. A credibility of a digital evidence is through its hash value, if its hash value is same throughout the procedure then it shall be well appreciated by the court but if any sort of change in the value is found then there is a possibility of human tampering with it. MIDAS system and Cognitech software whose build was: 6.0105 was used by FSL team. Their technique was to examine the video, frame by frame. Then the Meta Data of the file was examined and by applying mathematics as explained above, the speed of the car was established. The same technique was used in Vismay Shah's case, details:

10.2 *Timeline Analysis of the Video*

From the moment the car arrived till it exited a total distance of 15 m was captured in the impugned camera and they were marked as Mark-1 and Mark-2.

Distance between Mark-1 and Mark-2:

$$d = 15 \text{ m}$$

> Frame number when suspect car at Mark-1 = 10

> Frame number when suspect car at Mark-2 = 22

► Number of frames between Mark-1 and Mark-2 = $22 - 10 = > 12$ frames

Frame per second rate of the Video = 25 fps

Time for suspect car to reach from Mark-1 to Mark-2

$$t = (12 \times 1) / 25 = 0.48 \text{ s}$$

Velocity of suspect car from Mark-1 to Mark-2

$$v = \text{Distance} / \text{Time} = 15 / 0.48$$

$$\approx 31.25 \text{ m/s}$$

$$(31.25 \times 3600) / 1000 \text{ km/h}$$

$$\approx 112 \text{ km/h}$$

Hence, the FSL has determined a specific speed of 112 kmp.

As per the procedure, the investigating agencies have demanded a speed report from BMW dealership, which should state the speed of the car after reviewing the damage inflicted upon it. The BMW dealership confirmed that the speed of the vehicle being driven by Vismay was approximately 110 kmph. To further validate the method applied to establish the speed of the car, on 08th April 2013, the Director of Forensic Science, FSL, Mr. Hitesh Trivedi sir and his team went to the crime scene with the Investigating Officer Manoj Sharma, Police Inspector, Vastrapur. There they asked the police vehicle gypsy to cover a distance of approximately 1.5 km, crossing the same cam-03 of Lakshmi Gathiya Rath, three times but with different speeds, first with the speed of 50 kmph then with the speed of 60 kmph and then with a speed of 70 kmph.

From the moment the car arrived till it exited a total distance of 15 m was captured in the impugned camera and they were marked as Mark-1 and Mark-2.

Date and Time: 08.04.2013 from 12:44:54 to 13:00:59.

Distance between Mark-1 and Mark-2:

$d = 15 \text{ m}$

► Number of frames between Mark-1 and Mark-2 = 27 frames.

Frame per second rate of the Video = 25 fps

Time for police vehicle gypsy to reach from Mark-1 to Mark-2

$t = (27 \times 1)/25 = 1.08 \text{ s}$

Velocity of suspect car from Mark-1 to Mark-2

$v = \text{Distance/Time} = 15/1.08$

$\approx 13.88 \text{ m/s}$

$(13.88 \times 3600)/1000 \text{ km/h}$

$\approx 50 \text{ km/h}$

After the whole activity was performed by the police vehicle gypsy on the field. The FSL team has then and there, at the same time, took the routine backup in their pen drive. The pen drive was then brought and transferred to the MIDAS system, which has inbuilt Cognitech software, and a hash value was created then the frames were made. By applying the same technique, the speed of the car, when it was at 60 kmph and 70 kmph respectively, was calculated and the results also became the same. Hence, this was the validation of the technique done by the FSL which has given a significant contribution toward the decision-making.

10.3 Judgment

It was established during the subsequent examination that the driver was aware of the announcement of the police commissioner of Ahmedabad, regarding the requirement to drive a vehicle within a specific speed limit and that he was fully aware that there is significant traffic in the area, even at night. The complainant and other eyewitnesses to the event determined that the vehicle was traveling at an excessive speed. The

FSL's observation even shows that no brake has been affected on either tire. An accident was caused because of the driver's lapse in concentration at the intersection of the three routes (T-junction) close to Laxmi Ganthiya Rath. The BMW car swerved to the left of the motorcycle and came to a stop after colliding with the tree. Even the RTO estimated that the car's speed was higher than the Notification of the Police Commissioner and forensic report, which, while being hotly contested, showed that the speed was 112 km/h. In these circumstances, when the accident occurred, the individual decided to flee and clearly reneged on taking the necessary actions for the medical treatment of the injured or intimidating the police. The police had filed the F.I.R. in accordance with IPC § 279, 338, and 304A as well as for violations of §§ 177, 184, and 134(1)(b) of the Motor Vehicles Act. But on July 13, 2015, the Id. judge of the second additional session in Ahmedabad (rural) found him guilty of the offences punishable under § 304 (II), 279, and 427 of the IPC as well as § 177 and 184 read together with § 134(1)(b) of the Motor Vehicles Act, and he was sentenced to five years in rigorous prison with fine. And in the appeal before the Hon'ble High Court of Gujarat, the conviction was upheld.

11 Conclusion

This research paper entails that judiciary must create its own paradigms. Every case involves some form of evidence in electronic form, thus a judge must be ready to wear technocratic hat every time such a case comes—rather than disregard the e-evidence. Applying tech to achieve a desired result is one aspect, the other is to give legal value to e-evidence. One may lose relevant evidence not because of “lack of technology” but because of “lack of appreciation of technology.” So when e-evidences are being produced before the courts, then instead of rejecting the same, the judges must ask—Whether the investigators/litigants took care in gathering the evidence? Could the evidence be faked? In *Mohammed Ajmal Amir Kasab v. State of Maharashtra* (2012) it was laid that “one must appreciate the courts which relied on electronic evidences, whether in the form of CCTV footage, mobile devices, memory cards, data storage devices, intercepted communications over VoIP, IP Addresses, etc. while delivering the judgement.” Due to lack of infrastructural resources like computers, mobiles, internet connectivity in various parts of the country, there have been arguments raised that such an Act will divide the country into “digital haves” and “digital have-nots.” However, such arguments are not well founded, as the digital core of the country is expanding every day, with newer people being added to the digital framework. In light of COVID one must appreciate, that all the services were being transferred to the general public digitally, from medical to school, to shopping and payments, to transfer of benefits through PDS and Aadhar services. The Courts were also functioning to protect and safeguard the citizen's right to privacy and personal liberty, during the times of COVID.

§ 304 part (II) of IPC is the least used § while charging a person. According to the statistics given by Delhi police, just three incidents under the provisions of § 304 Part

(II) were reported out of the 3,584 cases that were recorded in 2014. 1,277 cases were reported between January 1 and May 15, 2015, and just one of those cases resulted in the application of § 304 Part (II). According to police statistics, Gurgaon had 127 fatal accidents and 164 non-fatal accidents in 2014, however none of them were covered under § 304 (II) of IPC. This § was not used once during the 53 fatal accidents and 75 non-fatal occurrences that occurred in Gurgaon between January 1 and May 15, 2015. In Noida, there are 835 cases that have been recorded, but no complaints have been made in accordance with § 304(II) of IPC [24]. Likewise, in the instant case laws as discussed above also both the accused were charged under § 304A which grants lesser punishment but the trial pronounced them sentence under § 304 part (II). In the appeal before the high court, Salman Khan got acquittal but Vismay Shah was incarcerated, the clear and unequivocal reason is the strong, scientific, and responsible investigation by the investigating authorities and their heavy reliance on the digital evidence placed on record. However, in case of Salman Khan, the prosecution was failed to place on record, single admissible forensic evidence, which may have changed the verdict. Hence, it won't be incorrect to say that digital Evidence had helped and will help in framing of the charge from § 304A to 304 Part (II) of IPC.

Since evidence is a vital part of all investigations, it is essential that investigators are aware of the numerous legal definitions of evidence, the different types of evidence, and how evidence is assessed and valued by the court. The fundamental building elements of the inquiry are the evidence, which must be accepted, obtained, documented, safe-guarded, vetted, analyzed, shared, and presented in a form that the court would find acceptable if it were the foundation for a ruling. Evidence will remain a crucial aspect to take into account while creating the appropriate research techniques as we proceed through this book. During an issue hearing, the court has the power to accept or reject any evidence that is submitted. To determine whether it will be accepted or rejected, all evidence is evaluated. Eyewitness reports of what happened to a suspect as well as forensic evidence found at the crime scene are examples of the types of evidence that may be accepted or excluded. The exclusionary defense has the potential to invalidate any piece of evidence, thus investigators should be mindful of this possibility. Depending on a set of guidelines and the kind of evidence that was submitted, the court will decide whether certain evidence must be dismissed if there is a disagreement. If a witness is called to testify, the court will first decide if they are competent and reliable to do so. The general rule is that a credible witness will most likely be a competent witness (R v Schell, 2004). Competent means that you are legally authorized to testify, whereas constrainable means that you are legally qualified to testify. In this book's part on witness management, a variety of criteria that will be covered in more detail will both affect a witness' competency and objectionability. The court will hear a witness' testimony if one is deemed to be both competent and unreliable, determine the witness' credibility afterward, and then weigh the relevance of the facts given. If it is established that a witness is unreliable or inconsistent, their evidence won't be admitted in court. When deciding whether to accept material evidence into evidence, the court considers a number of considerations in the same manner that it does witness testimony. The following are some

of these elements, although we will go into more detail in our chapter on handling crime scenes:

- (a) If the evidence was lawfully collected;
- (b) How it was gathered, tagged, and stored, and;
- (c) If it was tainted in any way, as well as;
- (d) Whether the evidence's continuity chain was correctly maintained, are all important considerations;
- (e) The exclusion of evidence at trial may occur if one of these criteria is not met.

Additionally, the accused's violations of the Charter of Rights and Freedoms [25] may result in evidence being gathered, which the court may altogether exclude. There were several instances in which these rights and liberties were not upheld:

- (a) Inaccurate disclosure of all evidence prior to the trial will allow the accuser to fully refute the allegations.
- (b) Failure to provide the necessary warning and care required by § 10 of the Charter when obtaining a suspect's statement.
- (c) Improper or unlawful searching of a person or of a person's property.
- (d) Denial of the right to consult with an attorney after being arrested or detained.

An investigator can prevent errors that can lead to the removal of evidence from a trial by being aware of the standards for collecting, managing, and conserving evidence. An investigator may be able to prevent the entire exclusion of important pieces of evidence from use in the trial due to a breach of the Charter by adhering to the standards for defining violations of the Charter. Law enforcement agencies now utilize computers to both commit and prevent crime as a result of the ongoing growth of the discipline of digital evidence forensic science. Information that has been preserved or transferred in binary format and is admissible as evidence in a court of law is known as digital evidence. A portable phone and a computer's hard disk are two more locations where it could be located. Electronic crime, or "e-crime," including child pornography and credit card fraud, is frequently associated with digital proof. Today, however, digital evidence is used in the prosecution of all offenses, not simply those involving computers. For instance, the emails and phone records of the suspects might provide vital information about their motive, location, and connections to other suspects, as well as their whereabouts at the time of the crime. For instance, a disquette in 1995 helped police track down the BTK serial murderer, who had been evading capture since 1974 and killed at least 10 deaths. The police were unable to apprehend him for thirty years. In the parking lot of a neighboring Home Depot in January 2005, Rader hid a package for Wichita's KAKE-TV station inside a cereal box he had parked behind a truck. The owner, nevertheless, saw it as a command and responded accordingly. Rader later got in touch with the station to ask if they had gotten the gift. After he was found, the police were able to see security camera footage that showed a person driving a black Jeep Cherokee dumping the item.

Data that has been preserved or transferred in binary format is referred to as digital evidence and can be presented as proof in a legal proceeding. It may also be

found on a portable phone and a computer's hard disk in addition to other places. Digital evidence is frequently connected to electronic crime, or e-crime, such as child pornography and credit card fraud. However, today, all types of crimes are being prosecuted utilizing digital evidence, not only electronic crimes. Emails or phone records, for instance, may include vital information about the suspects' motive, location at the moment of the crime, and connections to other suspects [26]. Law enforcement agencies have included the acquisition and processing of electronic information, also known as criminalistic information, into their infrastructure in order to combat cyber-crime and collect digital evidence pertinent to all crimes. Law enforcement agencies must train their staff to obtain digital evidence and keep up of the swift development of technology, such as computer operating systems. Our nation's criminal justice system is at a crossroads because even horrible offenders regularly evade the reach of the law, and dependable, competent, and trustworthy witnesses to crimes seldom come forward to testify in court. Due to intimidation, fear, and a number of other causes, even the accusation's trustworthy witnesses become hostile. As a result, the research organization must seek out extra strategies to raise the caliber of the investigation, which can only be done by compiling scientific evidence. In the era of science, we need to have solid legal foundations in both science and law. People today feel that new and innovative approaches must replace long-standing conventions and guiding ideas if we are to rescue our criminal justice system. Due to the emergence of new types of crimes and the complexity of those crimes, which make conventional methods and equipment obsolete, forensic science has to be strengthened for crime detection. Oral testimony is affected by the ability to see, to be humbled, to be swayed by other pressures, to forget, etc., while medical evidence is not affected by these same things. These scientific documents must be able to be handled and understood by the legal system. It would be advantageous for judges to often engage with scientists and engineers since this would increase their understanding of how to manage scientific evidence and successfully handle criminal cases based on scientific evidence. However, we highlight the need to encourage scientific evidence as a means of detecting and establishing crimes above and above other types of evidence rather than assuming that it will always be the most trustworthy. Ratio and the recent conclusions of the Court Apex in the *Dharam Deo Yadav v. State of Uttar Pradesh* [27] case call for a reference to this phase, where the crime scene must be handled accurately and scientifically. In criminal prosecutions, especially those based on circumstantial evidence, the forensic profession is extremely important. The key components of the crime could be established, the suspect could be located, and the accused's guilt or innocence could be established. Searching thoroughly for any evidence that could be used as evidence in the case is one of the investigator's main duties while on the site of the crime. The investigator may be shielded from any material evidence contamination that may emerge at the crime scene throughout the gathering, packing, and shipping of evidence. The appropriate measures must be taken to protect the evidence from tampering, contamination, and damage while still preserving it. The corpus of scientific data includes both soft and hard sciences, including economics, psychology, and sociology, as well as the hard sciences of physics, chemistry, mathematics, and biology. The opinions are obtained

from people with technical, scientific, or other information whose knowledge, skills, experience, education, training, or other qualifications would be able to help the court comprehend the evidence or identify the causal component. When the court is compelled to evaluate circumstantial evidence, scientific and technological evidence usually plays a crucial role. The nation's criminal justice system is frequently at a crossroads because even horrible offenders routinely evade the reach of the law, and competent eyewitnesses to crimes who have gained the public's confidence and are trustworthy witnesses seldom appear in court. Intimidation, fear, and a number of other circumstances cause even the accusation's reliable witnesses to become hostile. The involvement of forensic science is vital in these circumstances since circumstantial evidence is all that is available. She can support the establishment of the crime's components, help locate the offender, and help evaluate if the accusation is true or not. Searching thoroughly for any prospective evidence that could be useful in establishing the crime is one of the investigator's key duties while on the site of the crime. In order to improve the quality of the research, the organization conducting it must thus seek for additional strategies. This can only be done by assembling scientific evidence. We must lay down solid legal foundations in both science and law in the scientific age. With the implementation of the IT Act of 202 and the Indian Evidence Act's allowance of digital proof as credible evidence, the Indian legislature does not short for paper. Due to the emergence of new types of crimes and their increasing sophistication, which makes existing techniques and instruments obsolete, there is a need to strengthen digital evidence for crime detection. Traditions and guiding concepts from the past must be replaced with fresh and innovative approaches. Digital evidence is not susceptible to the defects that might affect spoken testimony, such as the observer's capacity, outside influences, amnesia, etc. As was plainly stated above, Vismay Shah received a prison term whereas Salman Khan was declared innocent due to a lack of digital or reliable forensic evidence. The FSL in Gandhinagar uses a relatively straightforward method for confirming speed tests with CCTV that involves performing basic time and distance calculations. But FSL, Gandhinagar deserves praise for using this straightforward maths in this situation. It brings up a number of fresh opportunities for the use of CCTV cameras in the area of digital evidence, with proving speed being one of them. The constant discussions between judges and scientists, engineers, and other professionals would be helpful in elucidating how to handle scientific evidence and successfully manage criminal cases supported by scientific evidence. It is crucial to promote scientific evidence as an extra tool for detecting and proving crimes above and beyond other types of evidence since it is possible that scientific evidence may not always serve as the only sign of guilt, which is why it may be incorrect to presume.

“No comprehensive understanding of the cosmos is given to us by science. Instead, it serves as a means of refining and developing theoretical accounts of reality that are then subjected to a more thorough examination” [28]. Science and technological developments have facilitated new discoveries and innovations while also simplifying research procedures. Both the items recovered from the crime scene and the digital information stored on technological devices are useful in solving crimes. The examining agencies may find these documents useful. Since it is possible for the

substance of the digital data to be altered, the analyst must carefully and prudently modify, examine, and manipulate the digital evidence. Due to the potential use of phone calls, location monitoring, and other digital data in criminal investigations. The data obtained on social media platforms, in texts, emails, and other important digital data may offer a trail for detectives, assisting in apprehending the offender and proving his or her guilt in court. The substance of electronic evidence is acceptable before courts, but courts must follow certain procedures before recognizing it as admissible due to the heightened risk of tampering and it is thus urged that digital evidence be handled cautiously and rationally position of digital. It is important to consider the evidence used in the charge's framing under IPC §§ 304A and 304 Part (II). When charging a person, § 304 part (II) of the IPC is the least frequently invoked. Just three incidences falling under the purview of § 304 Part (II) were reported out of the 3,584 cases that were registered in 2014 [24], according to the information provided by the Delhi police, as was previously stated and reiterated. Between January 1 and May 15, 2015, there were 1,277 recorded cases; however, only one of those incidents led to the application of § 304 Part (II). In 2014, Gurgaon saw 164 non-fatal accidents and 127 fatal accidents, although none of them fell under § 304(II) of the IPC, according to police figures. In the 53 fatal accidents and 75 non-fatal incidents that took place in Gurgaon between January 1 and May 15, 2015, this § was never used. There have been 835 cases reported in Noida, however no complaints have been filed in line with § 304(II) of the IPC. Similar to the prior case laws, both of the accused in the present instance were prosecuted under § 304A, which has a reduced penalty, but the trial determined that they should be sentenced by § 304 part (II). The strong, responsible, and scientific investigation conducted by the investigating authorities and their heaviest reliance on the digital evidence entered into the record are the obvious and undeniable reasons why Vismay Shah was imprisoned in the appeal before the high court while Salman Khan was acquitted. However, the prosecution failed to provide even one piece of acceptable forensic evidence in the Salman Khan case, which may have altered the outcome. Therefore, it is accurate to state that digital evidence assisted in and will continue to assist in the drafting of the charge from § 304A to 304 Part (II) of the IPC.

References

1. Tewari RK, Ravikumar KV (2000) Development of forensic science in India: a historical account. *JIAFS* 46
2. Criminal Justice and Criminalistics, California State University, 5151 State University Drive, Los Angeles, CA 90032
3. Section 2(t) IT Act 2000
4. Section 65A IE Act 1872
5. Criminal Appeal No. 864 of 2015
6. 2016 (1) ABR (Cri) 343.F

7. As per the data released by the National Crime Record Bureau under ‘*Accidental Deaths and Suicides rates*’. National crime records bureau report 2020 on traffic accidents. https://ncrb.gov.in/sites/default/files/ADSI-2021/adsi2021_Chapter-1A-Traffic-Accidents.pdf. Last visited 09 Sept 2022
8. Guha S Road accidents claim 18 lives every hour across India, says Government data. <https://www.timesnownews.com/auto/road-accidents-claim-18-lives-every-hour-across-india-says-government-data-Section-94036016>. Last visited 09 Sept 2022
9. India tops the world with 11% of global death in road accidents: World Bank report. The Economic Times. <https://m.economictimes.com/news/politics-and-nation/india-tops-the-world-with-11-of-global-death-in-road-accidents-world-bank-report/Sectionshow/80906857.cms>. Last visited 10 Sept 2022. See also Niaz S India tops world in road deaths, injuries. Anadolu Agency. <https://www.aa.com.tr/en/asia-pacific/india-tops-world-in-road-deaths-injuries/2425908>. Last visited 10 Sept 2022
10. Rohit TK Convictions in accident cases low. The Hindu. <https://www.thehindu.com/news/cities/chennai/Convictions-in-accident-cases-low/Section14416142.ece>. Last visited 09 Sept 2022
11. Suresh VS, a lawyer at the Madras High Court, dealing with motor accidents and claims cases, Sindhu Kannan, Less than 1% of road accident cases result in conviction in Tamil Nadu. Times of India. <https://timesofindia.indiatimes.com/city/chennai/less-than-1-of-road-accident-cases-result-in-conviction-in-tamil-nadu/Sectionshow/85119122.cms>. Last visited 09 Sept 2022
12. Police obtain CCTV footage of Cyrus Mistry’s car shortly before it crashed. Tribune India. <https://www.tribuneindia.com/news/nation/police-obtain-cctv-footage-of-cyrus-mistrys-car-shortly-before-it-crashed-429072>. Last visited 8 September 2022
13. Gubinas and Radavicius v HM Advocate, High Court at Aberdeen 2017 HCJAC 59, Scotland, in para 59
14. Criminal Appeal No. 142 of 2015
15. Civil Appeal Nos. 20825-20826 of 2017
16. Bateman: Lord Hewart CJ in R v Bateman (1925) All ER Rep 45; (1925) Cr App Rep 8
17. Kurban Hussain Mohammadeli Bangwalal v. State of Maharashtra, AIR 1965 SC 1616
18. Id
19. Appeal (crl.) 407 of 2002
20. Salman Khan hit-and-run-case: all that happened during the day. Times of India. <https://timesofindia.indiatimes.com/entertainment/hindi/bollywood/photo-features/salman-khan-hit-and-run-case-all-that-happened-during-the-day/photostory/47200703.cms>. Last visited 19 Aug 2022
21. Pacheco S (2015) Kamaal Khan, the man who was with Salman Khan during the night of accident. Indian Express, May 12. <https://indianexpress.com/article/entertainment/bollywood/who-is-kamaal-khan-salman-khan-hit-and-run-case/>. Last visited 16 Aug 2022
22. Id
23. Salman Khan 2002—hit-and-run case: HC suspends Salman’s five-year jail sentence. Times of India. May 8, 2015. <https://timesofindia.indiatimes.com/entertainment/hindi/bollywood/news/Salman-Khan-2002-hit-and-run-case-HC-suspends-Salmans-five-year-jail-sentence/articleshow/47199985.cms?imageid=47171584#slide9>. Last visited 17 Aug 2022
24. Bhatia S (2015) It’s true: ‘ordinary people’ hardly ever see Sec 304 (II) applied in accidents. Times of India, June 5. <https://timesofindia.indiatimes.com/city/noida/its-true-ordinary-people-hardly-ever-see-sec-304-ii-applied-in-accidents/Sectionshow/47559479.cms>. Last visited 11 Sept 2022
25. The Canadian Charter of Rights and Freedoms. Government du Canada. <https://www.justice.gc.ca/eng/csj-sjc/rfc-dlc/ccrf-cedl/>
26. Smith BH (2018) How a floppy disk exposed Dennis Rader as The BTK Killer. Oxygen True Crime, Aug 29. <https://www.oxygen.com/snapped/crime-time/floppy-disk-exposed-dennis-rader-btk-killer>. Last visited 2 Sept 2022
27. (2014) 5 SCC 509
28. Daubert v. Merrell Dow Pharmaceuticals, Inc. 509 U.S. 579 (1993)

29. Gaur KD (2020) IPC, 7th edn. Lexis Nexis
30. Storm KJ, Hickman MJ (2015) Forensic science and the administration of justice critical issues and directions. Sage
31. (2019) Ratanlal & Dhirajlal The IPC, 36th edn. Generic

Vismay Shah Case Documents

32. To view the official documents and attested copies of the case, visit at https://drive.google.com/drive/folders/1rYL6_DM1ZXrX-tG5I7tYq60UQE2bcd8u?usp=sharing

A Forensic Video Upscaling Colorizing and Denoising Framework for Crime Scene Investigation



S. Prema and S. Anita

Abstract Digital videos have been widely used as key evidence sources in Forensic crime scene investigations. Resolution is one of the most dominating parameter which affects the overall quality of the video. The main goal of this paper is to find an efficient forensic video analysis framework to assist the forensic crime scene investigation. A forensic video analysis framework (FVAF) that employs an efficient video enhancing deep learning model for increasing resolution of the low quality videos is used. The low resolution video is fed as input to the model. First, the video is pre-processed using fastai deep learning library. Large videos are cropped to manage runtime efficiently. Second, the video is rescaled for increasing the resolution by Spatial Resolution method. The framework successfully increases the resolution of the video from SD-standard definition Resolution type of 480p with Aspect Ratio 4:3 of Pixel size 640×480 to Full Ultra HD Resolution type of 8K or 4320p with Aspect Ratio 16:9 of Pixel Size 7680×4320 . The rescaled videos are submitted for colorization process. DeOldify deep learning model using Self-Attention Generative Adversarial Network and Two Time-Scale Update Rule is adopted by FVAF framework for colorizing the videos. Also, the colorized videos are trained and tested by various video enhance AI models model Gaia High Quality 4K rendering and Theia fine Tune detail. 4K not rendered and Theia Fine Tune Fidelity: 4K not rendered and video denoise AI models model Standard, clear, lowlight, severe noise and Raw. The upscaled and colorized video is also trained and tested using denoise video enhance AI and video denoise AI models. The results of each model are stored for comparison. From the stored results best video enhance AI model and the best video denoise AI models is selected. Lowligh AI model and Gaia high quality 4K rendering are used

Electronic supplementary material The online version of this chapter (http://doi.org/10.1007/978-981-99-5091-1_18) contains supplementary material, which is available to authorized users.

S. Prema

Arulmigu Arthanareeswarar Arts and Science College, Thiruchengodu, Tamil Nadu, India

S. Anita (✉)

Bhusanayana Mukundadas Sreenivasaiah College for Women, Bengaluru, Karnataka, India

e-mail: anita.sigamani@gmail.com

in this FVAF to produce high standard video for Forensic Analysis. We run this model using GPU to efficiently pre-process the video. By this framework, we increase the resolution of the video footages to further assist the forensic crime investigation.

Keywords Forensic · Video enhancement · Spatial resolution · Colorization · Deep learning

1 Introduction

Forensic Video Analysis plays an important role in the evaluation of video. It is often necessary for a forensic technician, analyst, or video forensic expert to perform a digital video evidence recovery in order to secure the Digital Media Evidence and establish a chain of custody. LEVA defined Digital Media Evidence as “Information of probative value stored in binary form” [1]. CCTV surveillance video recordings are the most common type of digital media evidence (DME). Once the recordings have been secured, an accurate chain of custody can be presented to the trier of fact. In addition, the forensic expert that acquired the video evidence can ensure that the highest quality versions of the recording are obtained so that a successful forensic video enhancement or forensic image comparison can be performed. The objective of Forensic Video Enhancement is to clarify or enhance the events as they occurred. This is done using nondestructive techniques to preserve the video evidence integrity, and pixel quality. Clarifying or enhancing the events as they occurred assists the Trier of Fact to weigh the video evidence and its relevance to the litigation. Video forensic experts are often asked to enhance CCTV Surveillance video recordings and to provide video image enhancement for identification purposes in the court. Popular enhancement techniques are applied to an investigation are Scaling [2], Pixel Interpolation, sharpening, warp stabilization, Shadow and highlight adjustments, frame averaging, pixel aspect ratio calibration, Color Correction, Reverse Projection, Photogrammetry, Motion Tracking, Demonstrative Video Exhibits. Digital videos are used as key evidence sources in Forensic crime scene investigations. Resolution is one of the most dominating parameter which affects the overall quality of the video. The objective of this paper is to develop Advanced Forensic Video Restoration Framework to perform accurate investigation.

1.1 Forensics

Forensic science, also known as criminalistics, is the application of science to criminal and civil laws, mainly on the criminal side during criminal investigation, as governed by the legal standards of admissible evidence and criminal procedure. Forensic science is a broad field that includes; DNA analysis, fingerprint analysis, blood stain pattern analysis, firearms examination and ballistics, tool mark analysis,

serology, toxicology, hair and fiber analysis, entomology, questioned documents, anthropology, odontology, pathology, epidemiology, footwear and tire tread analysis, drug chemistry, paint and glass analysis, digital audio video, and photo analysis. Forensic scientists collect, preserve, and analyze scientific evidence during the course of an investigation. While some forensic scientists travel to the scene of the crime to collect the evidence themselves, others occupy a laboratory role, performing analysis on objects brought to them by other individuals. Still, others are involved in the analysis of financial, banking, or other numerical data for use in financial crime investigation, and can be employed as consultants from private firms, academia, or as government employees. Computers are used for committing crimes, and, thanks to the burgeoning science of digital evidence forensics, law enforcement now uses computers to fight crime. Digital evidence is information stored or transmitted in binary form that may be relied on in court. It can be found on a computer hard drive, a mobile phone, among other places. Forensic video analysis is the scientific examination, comparison, and/or evaluation of video in legal matters. The video forensic process must be performed in a forensic lab that is equipped with the appropriate tools and follows best practice protocols in order to process the video recording with integrity and accuracy.

2 Literature Review

2.1 Survey on Deep Learning Algorithms

Deep learning has evolved over the past five years, and deep learning algorithms have become widely popular in many industries. It is a type of machine learning that works based on the structure and function of the human brain. Industries such as health care, eCommerce, entertainment, and advertising commonly use deep learning (Table 1).

2.2 Architectures Design of Deep Learning Models

See Figs. 1, 2, 3, 4, and 5.

3 Video Super-Resolution Methods Based on Deep Learning: A Survey

See Table 2.

Table 1 Survey on deep learning algorithms

Deep learning algorithm	Developed by	Year	Components and operations	Output	Uses
Recurrent neural networks (RNNs) [3, 4]	David Rumelhart's	1986	<ul style="list-style-type: none"> • Training • Gradient descent • Global optimization method • Fully recurrent • Elman and Jordan networks 	• (Batchsize, units)	• One vector per timestamp per sample
LeNet/ convolutional neural networks (CNNs) [5–7]	Yann LeCun	1988	<ul style="list-style-type: none"> • Multiple layers • Convolution layer • Pooling layer • ReLU activation function • Downsampling • Flattening 	<ul style="list-style-type: none"> • Feature map • Linear vector 	<ul style="list-style-type: none"> • Image processing and object detection • Identify satellite images
Long short term memory networks (LSTMs) [8, 9]	Juergen Schmidhuber's	1995	<ul style="list-style-type: none"> • Chain-like structure • Four interacting layers • Forget irrelevant parts • Update the cell-state values 	<ul style="list-style-type: none"> • 4 different sets of results • Dault: last hidden state • Reurn_ sequences = true: all hidden states 	<ul style="list-style-type: none"> • Speech recognition • Music composition • Pharmaceutical development
Generative adversarial networks (GANs) [10–12]	Ian Goodfellow	2014	<ul style="list-style-type: none"> • A generator • A discriminator • Initial training • A zero-sum game 	<ul style="list-style-type: none"> • Inception score • Inception-v3 • Fréchet inception distance (FID) 	• Improve astronomical images

4 Problem Statement

Resolution is one of the most dominating parameter which affects the overall quality of the video. Resolution greatly influences the viewing experience. One of the key aspects of delivering a good quality video is understanding and determining the correct video resolution of an on-demand video or live stream. Comparing with the frame rate, Video quality feature is the most important to deal with. For performing effective analysis of crime cases in Forensic field, video enhancement plays a vital

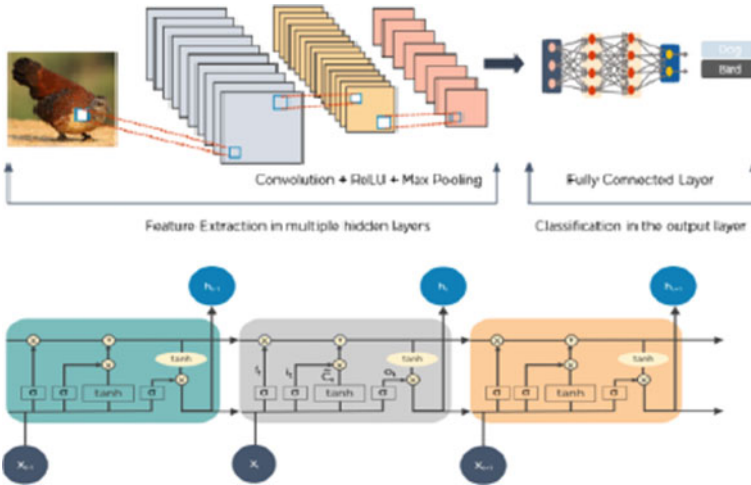


Fig. 1 CNN [7] and LSTMs flow of operation [8, 9]

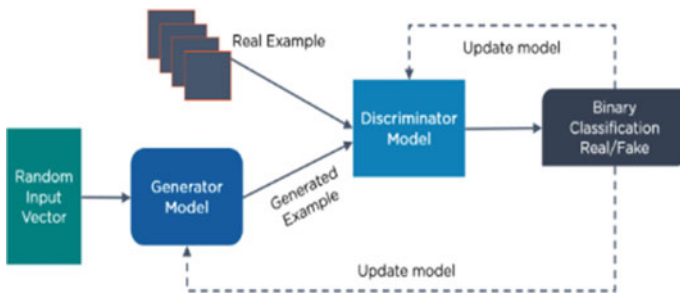


Fig. 2 Unfolded RNN [13] and a working process of GAN [12]

role. This paper is created with the objective of enhancing and colorizing the old footage videos for helping in efficient Forensic crime scene investigation.

5 Proposed Methodology

In this framework, we first pre-process the video. Towards pre-processing the video, we crop large videos to manage runtime efficiently. Next, we rescale the video for increasing the resolution. In our FVAF model, we use Spatial Resolution method for increasing the resolution of the low quality video. Our framework successfully increases the resolution of the video from SD-standard definition Resolution type of 480p with Aspect Ratio 4:3 of Pixel size 640×480 to Full Ultra HD Resolution type of 8K or 4320p with Aspect Ratio 16:9 of Pixel Size 7680×4320 .

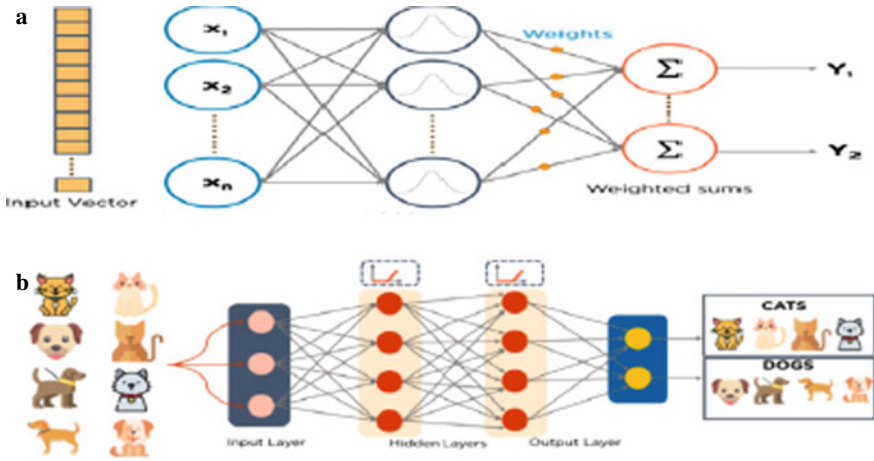


Fig. 3 a RBFN model [14] and b MLP process [15]

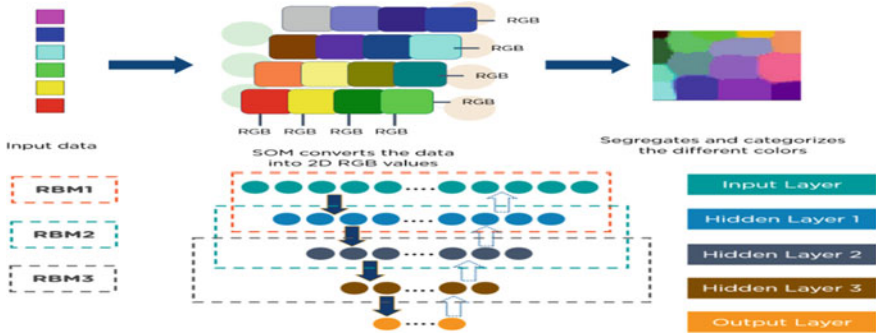


Fig. 4 SOMs process [16] and DBN architecture [17]

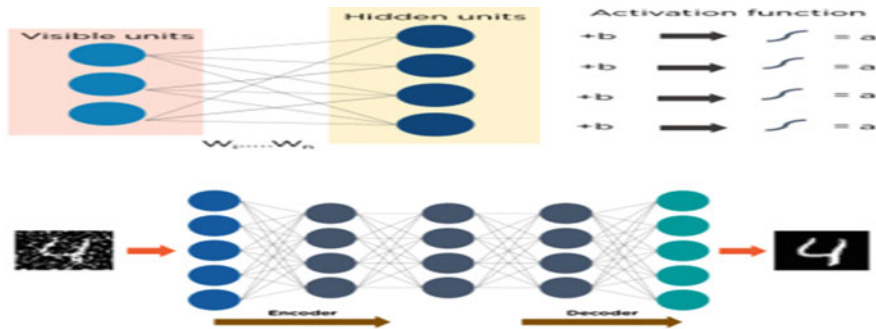


Fig. 5 RBMs function [18] and autoencoders operation flow [19]

Table 2 Existing video super-resolution methods based on deep learning

Model	Loss function	PSNR	SSIM	Ref.
Frame recurrent video super resolution (FRVSR) 2018	Loss with total variation regularization	26.17	0.798	[9]
Super-resolution optical flow for video super-resolution (SOFVSR) 2019	MSE loss and MC loss	NA	NA	[18]
Motion estimation and motion compensation network (MEMC-Net) 2021	Charbonnier (Cb) loss	34.95	0.9679	[20]
Dual subnet and multistage communicated upsampling (DSMC) 2021a	Cb loss; perceptual loss; the dual loss	27.56	0.8934	[21]
Video restoration based on deep learning: comprehensive survey 2022	Reconstruction loss	NA	NA	[22]
Video super-resolution via dense non-local spatial-temporal convolutional network (DNSTNet) 2020	1-norm loss	NA	NA	[23]
Space-time-aware multi-resolution network (STARnet) 2020	Three losses	NA	NA	[1]
BasicVSR++ (VideoLDV dataset) 2022	Charbonnier loss [24]	31.63	NA	[9]
CDVSR (Video LDV dataset) 2022	Charbonnier loss [24]	23.03	NA	[21]

Also, this framework aims in colorizing black and white videos by using DeOldify deep learning model using Self-Attention Generative Adversarial Network and Two Time-Scale Update Rule. We run this model using GPU to efficiently pre-process the video. By this framework, we increase the resolution of the video footages to further assist the forensic crime investigation. The upscaled frame is shown in Fig. 6a and Colorized frame is shown in Fig. 6b. In our FVRF model we use Spatial Resolution method for increasing the resolution of the low quality video. Spatial resolution is the height and width of the frame, which is measured in pixels. It's the total number of pixels in every individual frame. DeOldify Model includes Self Attention GANs, Progressive Growing of GANs, and Two time scale update rule. In this proposed framework DeOldify deep learning model is used to colorize the video. The DeOldify model uses design specification of self-attention used in the "Self-attention Generative Adversarial Networks" [25]. "Traditional convolutional GANs generate high resolution details as a function of only spatially local points in lower resolution feature maps [25]. In SAGAN, details can be generated using cues from all feature locations [12]. The discriminator can check that highly detailed features in distant portions of the frames are consistent with each other." Deoldify using self-attention guarantees maximal continuity, consistency, and completeness in colorization [26]. GANs: Generative adversarial networks (GANs) are deep neural net architectures [24]. GANs are comprised of two nets namely Generator and Discriminator, trained

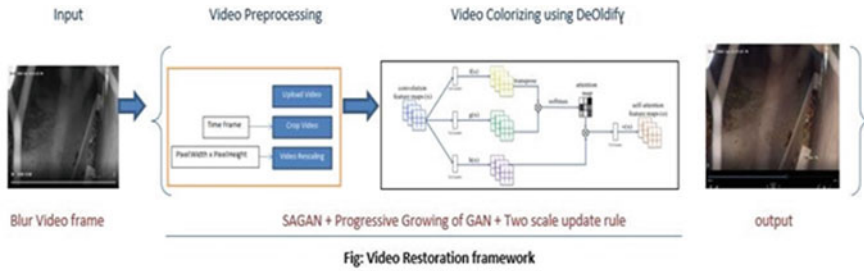


Fig. 6 Video restoration and colorization framework

one against the other. Typically, the generator is of main interest. The discriminator is an adaptive loss function that gets discarded once the generator has been trained. Generative model tries to fool the discriminator while discriminator acts as a detective trying to catch the forgers. In Progressive Growing of GANs the key takeaway is to grow both Critic and Generator model’s layers progressively starting from a low resolution, and to add new layers that model increasingly fine details as the training progresses [27]. Two time scale update rule: It is choosing different learning rates for the generator and discriminator. DCGAN [10] and WGAN-GP [28] using different learning rates have achieved state-of-the-art results. ‘Using different learning rates’—Is choosing a higher learning rate for the discriminator and a lower one for the generator. ConvBlock is used to build the Unet Block which in turn is used to build the Unet Model [29]. U-Net based generators are used in DeOldify. The Unet block uses Conv2D, activation layer, and Batch normalization modules.

5.1 Notation

The notations used are: x : Real data, z : Latent vector, $G(z)$: Fake data, $D(x)$: Discriminator’s evaluation of real data, $D(G(z))$: Discriminator’s evaluation of fake data, $Error(a,b)$: Error between a and b [9].

5.2 *The Discriminator*

The goal of the discriminator is to correctly label generated images as false and empirical data points as true [9]. Therefore, we might consider the following to be the loss function of the discriminator:

$$LD = \text{Error}(D(x), 1) + \text{Error}(D(G(z)), 0) \quad (1)$$

5.3 *The Generator*

We can go ahead and do the same for the generator. The goal of the generator is to confuse the discriminator as much as possible such that it mislabels generated images as being true.

$$LG = \text{Error}(D(G(z)), 1) \quad (2)$$

5.4 *Binary Cross Entropy*

A common loss function that is used in binary classification problems is binary cross entropy. As a quick review, let's remind ourselves of what the formula for cross entropy looks like:

$$H(p, q) = \mathbb{E}_{x \sim p(x)}[-\log q(x)] \quad (3)$$

5.5 *Training the Discriminator*

When training a GAN, we typically train one model at a time. In other words, when training the discriminator, the generator is assumed as fixed. In min-max, the quantity of interest can be defined as a function of GG and DD. Let's call this the value function:

$$V(G, D) = \mathbb{E}_{x \sim p_{\text{data}}}[\log(D(x))] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))] \quad (4)$$

6 Results

6.1 Video Upscaling and Sharpening

The upscaled and colorized video is further trained with AI models [11] for producing better results. The video is upscaled with theia models in theia-Detail + theia fine tune-Fidelity, Artemis Aliased, and Moire and Gaia models in Gaia High Quality 4K rendering. A more enhanced clarity of the sky and the vehicles moving on the road are produced using theia fine tune-Fidelity: 4K. The results obtained are as follows (Figs. 7, 8, and 9).



Fig. 7 a Original scaled 50%, b Gaia high quality 4K rendering, c Theia fine tune fidelity: 4K not rendered, d Artemis aliased and Moire zoomed at 36%



Fig. 8 a Original scaled 50%, b Theia fine tune fidelity: 4K rendering, c Theia fine tune fidelity: 4K not rendered, d Artemis high quality: 4K not rendered, zoomed at 36%. Theia fine tune fidelity: 4K rendering is found to be best compared with Artemis high quality: 4K not rendered. The objects in Fig. 8b are seen with clear edges than in the output of Artemis high quality model



Fig. 9 Results produced by DAIN and Artemis Aliased and Moire

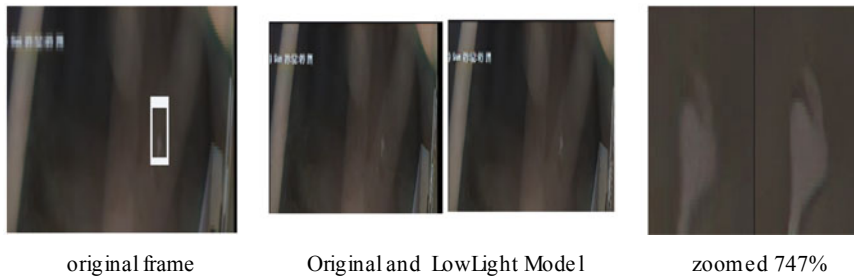


Fig. 10 The original video denoised with remove noise: 26 and enhance sharpness: 15. By setting removing noise to 26 and enhance sharpness to 15 and zooming at 747% the blur image has been sharpened and the noise level is reduced and the person wearing white shirt with pale red color ribbon on round his head is seen

6.2 Video Denoising

See Figs. 10 and 11.

The Artemis processed video is passed as input to the Video Enhance AI models Gaia High Quality 4K rendering, Theia fine Tune detail 4K not rendered, and Theia Fine Tune Fidelity: 4K not rendered at Zoom percent 59. Comparatively shows a clear view of the person walking with less artifacts.

6.3 Video Colorization

See Figs. 12, 13, 14, and 15.

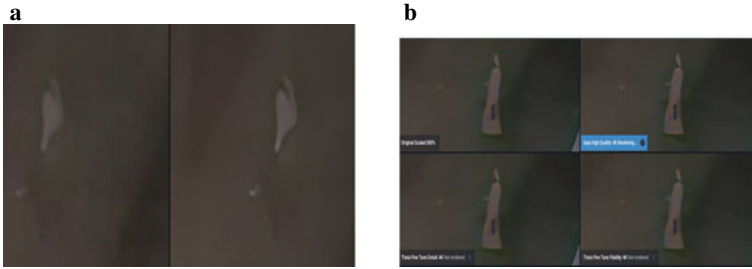


Fig. 11 The original video denoised with remove noise: 16 and enhance sharpness: 24. By setting removing noise to 16 and 3 enhance sharpness to 24 and zooming at 200% the person’s shirt edges are seen clear and the image sharpness is increased. The video frame is extracted as an image and tested with denoise AI models standard, clear, lowlight, severe noise, and raw. Comparatively lowlight model has shown improved sharpness with reduced noise. **a** Original scaled 50%, **b** Gaia high quality 4K rendering, **c** Theia fine tune detail 4K not rendered, **d** Theia fine tune fidelity: 4K not rendered zoomed at 59%



Fig. 12 Video resolution: SD—standard definition resolution type of 480 pixels resolution with aspect ratio 4:3 of pixel size 640 × 480



Fig. 13 Video resolution: full ultra HD resolution type of 8K or 4320p with aspect ratio 16:9 of pixel size 7680 × 4320



Fig. 14 Upscaled video



Fig. 15 Video colorization using Deoldify

6.4 Future Research Direction and Challenges

1. **Combine Ultra-high resolution with high FPS:**
High-speed imaging captures 1000 frames per second so that fast-moving subjects can be portrayed smoothly and fast motions can be analyzed with the constraint of lower resolution and several seconds of recording time.
2. **Extended Reality (XR): Immersive media and high-speed imaging:**
Immersive media demands resolution above 8K. Technologies developed for “High-Speed Imaging” will be useful as more pixels are used.
3. **Handling distracting stutters and dropped frames is a challenge:**
In video editing applications, slowing down footage from a 24 fps clip would result in distracting stutters and dropped frames. One can use chromos to create beautiful slow motion results regardless of the source frame rate.
4. **Maximizing FOV (Field of View) > 8K resolution is a challenge:**
Although Maximizing FOV is more relevant to broadcasting; in the near future it will find its way to cinema as well in camera manufacturers and in RED V-Raptor firmware updates.

7 Conclusion

The black and white video frames are converted to colorized video frames. Also, the colorized videos are trained and tested by various video enhance AI models models Gaia High Quality 4K rendering, Theia fine Tune detail 4K not rendered and Theia Fine Tune Fidelity: 4K not rendered and video denoise AI models models Standard, clear, lowlight, severe noise and Raw. The upscaled and colorized video is also trained and tested using denoise video enhance AI and video denoise AI models. The results of each model are stored for comparison. From the stored results best video enhance AI model and the best video denoise AI models is selected. Lowlight AI model and Gaia high quality 4K rendering are selected and used in FVAF to produce high standard video for Forensic Analysis. We run this model using GPU to efficiently pre-process the video. By this framework, we increase the resolution, colorize, upscale, and denoise the video footages to successful investigations in the forensic field.

References

1. Li S, Li W, Cook C, Zhu C, Gao Y (2018) Independently recurrent neural network (IndRNN): building a longer and deeper RNN. [arXiv:1803.04831](https://arxiv.org/abs/1803.04831) [cs.CV]
2. Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR 2016

3. Kohonen T (1982) Self-organized formation of topologically correct feature maps. *Biol Cybern* 43(1):5969. <https://doi.org/10.1007/bf00337288>. S2CID 206775459
4. <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8>
5. Bao W (2019) Depth-aware video frame interpolation
6. Chu M, Xie Y, Leal-Taixé L, Thurey N Temporally coherent GANs for video super-resolution (TecoGAN)
7. Ronneberger O, Fischer P, Brox T U-Net: convolutional networks for biomedical image segmentation
8. Sun W, Sun J, Zhu Y, Zhang Y (2020) Video super-resolution via dense non-local spatial-temporal convolutional network (DENSENet). *Neurocomputing*
9. <https://www.topazlabs.com/learn/introducing-video-enhance-ai-v2-3-frame-rate-conversion-and-slow-motion/>
10. Karras T, Aila T, Laine S, Lehtinen J Progressive growing of GANs for improved quality, stability, and variation
11. <https://github.com/LongguangWang/SOF-VSR>
12. Zhang H, Goodfellow I, Metaxas D, Odena A Self-attention generative adversarial networks. Last revised 14 Jun 2019 (version, v2)
13. Bianchi FM, Maiorino E, Kampffmeyer MC, Rizzi A, Jenssen R (2017) An overview and comparative analysis of recurrent neural networks for short term load forecasting
14. Ulf E (2002) Radial basis function networks (RBFN) January 2002
15. Popescu MC, Balas VE, Perescu-Popescu L, Mastorakis N (2009) Multilayer perceptron and neural networks. *WSEAS Trans Circ Syst* 8(7):579–588
16. García-Lamont F, Cervantes J, López-Chau A et al (2020) Color image segmentation using saturated RGB colors and decoupling the intensity from the hue. *Multimed Tools Appl* 79:1555–1584. <https://doi.org/10.1007/s11042-019-08278-6>
17. Li C, Wang Y, Zhang X, Gao H, Yang Y, Wang J (2019) Deep belief network for spectral–spatial classification of hyperspectral remote sensor data. *Sensors* 19:204. <https://doi.org/10.3390/s19010204>
18. Broomhead DS, Lowe D (1988) Radial basis functions, multi-variable functional interpolation and adaptive networks. Technical report RSRE 4148. Archived from the original on April 9, 2013
19. Michelucci U (2022) An Introduction to Autoencoders. <https://doi.org/10.48550/arXiv.2201.03898>
20. Bao W, Lai W, Zhang X, Gao Z, Yang M (2021) MEMC-Net: motion estimation and motion compensation driven neural network for video interpolation and enhancement. *IEEE Trans Pattern Anal Mach Intell* 43(3):933–948. <https://doi.org/10.1109/TPAMI.2019.2941941>
21. Hinton G (2009) Deep belief networks. *Scholarpedia* 4(5):5947. Bibcode: 2009SchpJ...4.5947H. <https://doi.org/10.4249/scholarpedia.5947>
22. Rota C, Buzzelli M, Bianco S, Schettini R (2023) Video restoration based on deep learning: a comprehensive survey. *Artif Intell Rev* 56(6):5317–5364. <https://doi.org/10.1007/s10462-022-10302-5>
23. Sun W, Sun J, Zhu Y, Zhang Y (2020) Video super-resolution via dense non-local spatial-temporal convolutional network. *Neurocomputing* 403:1–12. <https://doi.org/10.1016/j.neucom.2020.04.039>
24. <https://www.topazlabs.com/topaz-video-ai>
25. Yang R, Timofte R, Zheng M, Xing Q NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: dataset, methods and results
26. Zhang H, Goodfellow I, Brain G, Odena A Self-attention generative adversarial networks
27. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA Generative adversarial networks: an overview. *IEEE*
28. Radford A, Metz L, Chintala S Unsupervised representation learning with deep convolutional generative adversarial networks
29. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A Improved training of Wasserstein GANs

30. <https://www.topazlabs.com/VideoEnhanceAI/VideoQualitySoftware/TopazLabs>
31. Vincent P, Larochelle H (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res. <https://imagine-4d.com/multimmersion/>
32. Rota C (2022) Video restoration based on deep learning: comprehensive survey. Springer
33. Schmidhuber J (1993) Habilitation thesis: system modeling and optimization. Page 150 ff demonstrates credit assignment across the equivalent of 1,200 layers in an unfolded RNN

A Review of Face Detection Anti Spoofing Techniques on Varied Data Sets



Pratiksha K. Patel and Jignesh B. Patel

Abstract In the recent scenario, and also during the pandemic situation everything is processed as well as transferred digitally. Nowadays from kids to an Adult, every age group relies on digital platform which may result in cybercrime. Nowadays cybercrimes are on its peak. E.g. user's photo can simply be found on social media, these photos can be spoofed by facial recognition software (FRS). This digital face identity theft can be used to attempt varied activities related to money which can lead to banking fraud. Spoofing is a type of scam in which criminals attempt to obtain someone's personal information by pretending to be a legitimate business, a neighbor, or some other innocent party. To intercept these problems of recognizing real faces against fake faces, various researchers determine Face Anti-Spoofing techniques on varied data sets. Existing research still faces difficulties to solve spoofing attacks in the real world, as datasets are limited in both quantity and quality. The key aim of this paper is to contribute a detail study of Face Anti-spoofing techniques and evaluation of varied Datasets. Finally, we achieved from the study that many researchers have found truthful methods which solve spoofing threats. But, existing work requires a more proficient face anti-spoofing algorithm by which cyber crimes can be reduced.

Keywords Biometrics · Security · Anti-spoofing · Face

1 Introduction

In the recent scenario and also during the pandemic situation, everything is processed as well as transferred digitally. Nowadays from kids to an Adult, every age group rely on digital platforms which may result in cybercrime. Nowadays cybercrimes are on

P. K. Patel (✉)

Shri Shambhubhai V. Patel College of Computer Science and Business Management, Surat, India
e-mail: wait4pratiksha@gmail.com

J. B. Patel

Department of Computer Science, H.N.G. University, Patan, India
e-mail: jbpatel@ngu.ac.in

their peak. E.g. users' photos can simply be found on the social media, these photos can be spoofed by facial recognition software (FRS). Spoofing is a type of scam in which criminals attempt to obtain someone's personal information by pretending to be a legitimate business, a neighbor, or some other innocent party.

Contribution and Motivation

During the pandemic situation, it was found that many people rely on digital platforms where face recognition is used, to prevent face spoofing attacks where intruders use face mask to gain access of personal accounts. From this, it was realized that, work should be done on Face Anti-spoofing techniques which can prevent intruders even with face mask. Many researchers have worked in diverse directions, such as to develop efficient face anti-spoofing algorithms on limited data sets and to design 2D face Presentation attack detection system. The researchers have very less clue about face mask datasets. We desire to achieve high accuracy with these types of datasets, and to do so this paper presents varied datasets on which face Anti-Spoofing techniques work effectively.

1.1 Face Spoofing Attacks

Face spoofing attacks can be broadly categorized into two types—one is Impersonation and other is obfuscation [1]. Impersonation consists of (1) Photo attack using paper photographs, screenshots, or taking photos from digital camera etc. (2) video attack or replay attack in which video is replayed from mobile or tablet, or laptop. (3) 3D Mask attack in which attacker creates a 3D model of face or sometimes mask is used as a choice for spoofing [1]. Obfuscation includes partial attacks such as glasses on the face or tattoo; make up is also a part of whole obfuscation attack type [1]. To resolve this issue of spoofing attack on face, researchers design a secure face recognition system (Fig. 1) [1].

Face recognition is the process of identifying and confirming an individual using their face identity. It is a category of Biometric Security and is used in various industries such as surveillance, Banking, Internet search Engines, Self Payment System, Airport and custom offices, smart homes, etc. This large scale deployment of FRS has



Fig. 1 Face spoofing attacks

attracted intensive attention to the reliability of face biometrics against spoof attacks [2]. To solve these issues, various face anti-spoofing techniques have been proposed, but the issue has still not been solved due to the difficulty in finding discriminative and computationally inexpensive features and methods for spoof attacks.

1.2 Face Anti-spoofing Techniques

To perceive spoofing attacks, various Anti-spoofing techniques are developed, which include sensor level technique, feature level technique, and score level technique. Existing literatures state that sensor level techniques are also referred as hardware-based techniques [3]. Numerous research works with this technique have been held since 2005 to till date. We know that sensor level techniques use hardware device and sensor, to sense signals of the living body e.g. blood pressure, intrinsic property, challenge response, and multimodality. Drawback of sensor level technique is; it required higher level of cooperation, detection rate with this technique is much slow, as well as it is expensive (Fig. 2) [3].

Due to above mentioned difficulty with sensor level techniques; researchers have been found focusing more on feature level techniques [3]. Feature level techniques are also known as software based techniques which comprise of two methods static and dynamic. Dynamic feature level technique leads to loss of accuracy against video attacks. It is comparatively slow and it cannot be used for single image scenario.

Due to cons in dynamic feature level technique, researchers work with static feature level technique which is faster, transparent to the user, and provides less cooperation from user as well as it is less intrusive; and this method is also being improved day by day. In Anti-spoofing techniques, higher fake detection rate is achieved with sensor level technique; while feature level techniques are less expensive, less intrusive, and more user friendly, and also its implementation is hidden from the user [4]. The score level technique present much lower performance compared to sensor and feature level techniques. Therefore it is used in support sensor and feature level techniques [4].

Some researchers work with manual feature expression, which is further classified into image texture feature, Human Computer interaction, Life information, Image

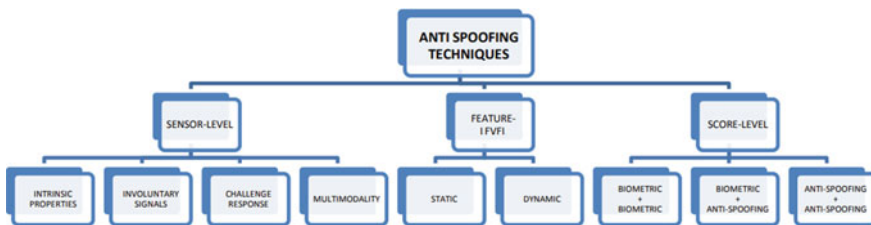


Fig. 2 Face anti-spoofing techniques

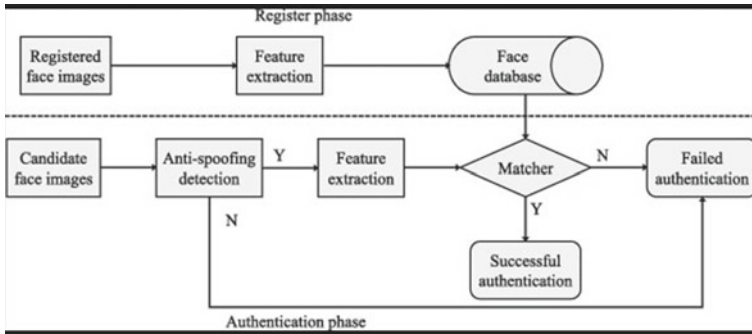


Fig. 3 Step by step process of face anti-spoofing techniques

Quality, and Depth Information [5]. In today's scenario to dig out some essential and abundant face features, Deep learning techniques were introduced, which additionally include Network Updating, Transfer Learning, Feature Integration based on Texture and Depth Information, and Spatio-Temporal Information [5]. Figure 3 demonstrates step by step process of face anti-spoofing techniques which comprise Feature extraction and classification techniques; by which spoofing attack is easily identified. Table 1 summarizes a comparative study of anti-spoofing techniques.

2 Face Anti-spoofing Datasets

Face Anti-spoofing techniques achieved high accuracy to distinguish between real and fake faces based on diverse datasets. Many researchers worked with varied datasets but still, it has some loop holes which need further improvement with existing datasets.

Table 2 summarizes public face anti-spoofing datasets (FAS) with different environmental conditions, for different subjects with varied capturing devices. It includes a number of video clips, training, and testing sample of subjects. It can be seen that most datasets [10, 12–16] include only photo attacks such as print and replay attack under simple illumination conditions. From last 5 years, researchers have been working with large scale datasets such as celebaspoof and CASIA-SURF HiFi Mask [20, 26] which include more than 60,000 images and 50,000 videos respectively. Apart from Traditional datasets which capture images from RGB camera, latest work has been carried out by Thermal [20], infrared [20] camera. All above mentioned diversities in dataset have led to a challenge for researchers to work with it and enlighten research work to make secure Face Recognition system.

Table 1 Comparison of various face anti-spoofing techniques based on few important attributes like technique used, image/video type, feature extraction, classifier, dataset

Anti-spoofing techniques	Type M/D	Year	Image/video type	Pre-processing	Feature extraction	Classifier	Dataset
Color texture analysis [6]	M	2015	RGB image	RGB + HSV + gray scale	Color local LBP descriptor	SVM	CASIA-Replay Attack
Vision based [7]	M	2016	Gray scale	Viola-Jones and Gaussian filter	HOG, Gabor, wavelet, SURF	SVM	Replay Attack
Multi cue integration framework [8]	D	2016	RGB/video	Viola-Jones face detector	Shearlet and dense optical flow	Softmax	Replay Attack, CASIA-FASD, 3D MAD
Textual + depth information [9]	M/D	2017	RGB	LBP from Kinect Camera	Viola-Jones face detector, CaffeNet CNN	SVM	CASIA
Ultra DNN [10]	D	2017	RGB video seq	ResNet 50 model	LSTM approach	Softmax	CASIA-FASD, Replay Attack
Texture based + filtering [11]	M	2019	RGB	Contrast + dynamic FE by DOG filter	LBPV	SVM	NUAA
Handcrafted + deep network feature [12]	M	2019	RGB image	LBP, open face detector	FASNet, VGG-16	VGG-16 binary classification	Replay Attack, Replay Mobile, 3D MAD
Deep color FASD [13]	D	2019	RGB video seq	RGB, HSV, YCbCr (3 channel)	Deep FASD (combine three channels)	Softmax + voting	CASIA-FASD
Score level + challenge response [14]	M	2020	RGB image	Weight score level fusion for face landmark	I blink detection method, challenge response	Mouth + speech features pass to ML	CASIA-MFSD, Replay Attack, MSU-MFSD
Spatial + temporal [15]	M	2020	RGB video	LBP	DCT + DWT	SVM	CASIA-FASD, Replay Attack

(continued)

Table 1 (continued)

Anti-spoofing techniques	Type M/D	Year	Image/video type	Pre-processing	Feature extraction	Classifier	Dataset
Texture based [16]	M	2020	RGB image	RGB + HSV + YCbCr + ED-LBP	Spatial pyramid from HE concatenation	SVM	CASIA-FASD, Replay Attack, Replay Mobile, OULU-NPU
Dynamic color texture [17]	M	2020	Gray/RGB image	RGB + HSV + YCbCr	LDNP, TOP	Pro CRC	CASIA-FASD, Replay Attack, UVAD
Generalization of color texture [18]	M	2020	RGB	HSV + YCbCr, LBP	RI-LBP, LPQ, CoALBP, RIC-LBP	SURF-FV	CASIA-FASD, Replay Attack, MSU-MFSD
CNN [19]	D	2021	RGB video stream	Single shot multi box detector	HE + data augmentation + CNN	Softmax	CASIA-SURF
Multi-stream using 3D information [20]	M	2022	3D information from 3D smart camera	Spherical, clipping + face pose unification	Depth map + SNM	FeatherNet + BCE pretrained	WMCA, Self Generated Anti-3D dataset
Masked face PAD + CNN [21]	M/D	2022	RGB face with mask	RBP + CpqD	Inception + N/W, FASNet pretrained arch	VGG-16	Comparison with 7 state of art methods
Texture analysis + optical flow + CNN [22]	M/D	2022	RGB image/videos	Optical flow for movement + amplitude of face	Motion textures cues + attention module	Softmax	Replay Attack, OULU-NPU, HKBU-MARs V1
Image quality assessment + CNN [23]	M/D	2022	RGB image	Handcrafted feature of RGB image	GD + AGD + GSD (extract 21 FE)	SVM	CASIA, Replay Attack, QULU-NPU, UVAD, Siw

(continued)

Table 1 (continued)

Anti-spoofing techniques	Type M/D	Year	Image/video type	Pre-processing	Feature extraction	Classifier	Dataset
Handcrafted + DNN [24]	M/D	2022	RGB image	LBP + DNN	MP + HFM module	ResNet50 pretrained	MSU-MFSD, Replay Attack, CASIA-FASD, QULU-NPU
Temporal sequence sampling + CNN [25]	M/D	2022	RGB videos	TSS of every frame + 2D face affine motion	Single RGB image + CNN	Softmax + BiLSTM	MSU-MFSD, CASIA-FASD, OULU-NPU

The summary of Table 1 elaborates on research in face anti-spoofing techniques from 2015 to the current year, which include starting from face anti-spoofing technique name, followed by its type whether it is Manual (M) or Deep learning (D) base, followed by year, Image or Video Type data. Subsequent columns emphasize on Face Anti-spoofing steps (Pre-processing, Feature Extraction, Classifier). The last column indicates name of dataset on which researchers carried out the work

Table 2 Face anti-spoofing datasets

Data set	Video clips	Attack type (P/V)	Subjects/users	Environmental conditions	Capture device	Total papers
Replay Attack [12, 14–16]	1300	P/V	50	Different lightning conditions	Built-in Webcam of laptop or video in front of camera	81
MSU-MFSD [14]	280	V	35	(720 × 480) and (640 × 480) resolutions	Laptop + Andriod + iphone	52
CASIA-MFSD [14]	600	12 video	50	Different resolutions and lightning conditions	Laptop (replay, wrap, print, cut print attack)	40
CASIA-FASD [10, 13, 15, 16]	Limited	P	50	Diverse attacks	Normal webcam	20
Replay Mobile [12, 19]	1190	P + V both	40	Different lightning conditions	iPADMini2(10S), LQ-G4 Smartphone	16
CASIA-SURF [20]	Videoclips	V	Subject + modalities	Different resolutions	RGB, depth camera	9
HQ-WMCA [20]	2904	V	555-bonafied 2349-PA videos	Data from different channels	Color depth camera, thermal, infrared	6
CASIA-SURF HiFi Mask [20, 26]	54,600	V	75	225 realistic mask × 7 types of sensors	Sensors	5
MLFP [20]	7	P + V	7-3D latex mask 3-2D print attack	Data from different channels	Video, thermal, infrared channels camera	4
OULU-NPU [6, 16, 20, 27]	–	P + V	55	2D face	Print attack, display device attack from web cam	4

Table 2: A summary of existing public datasets for Face Anti-Spoofing. The table list indicates that some of the presentation attacks were captured by RGB camera, Depth camera, multiple sensors, Laptop and Mobile devices. Dataset include photo (P) and video (V) attacks with a varied number of subjects

2.1 Evaluation Parameter

To check the accuracy of different datasets, researchers require some evaluation criteria. APCER, BPCER, ACER, ACC, and HTER to check accuracy on Replay Attack, CASIA-FASD, and Replay Mobile datasets [10, 12, 14–16, 19]. 3D MAD

dataset uses HTER and ACC. Recently more improvement required in cross datasets, Intra Data sets, Inter Data sets with negligible error rate and high accuracy.

3 Summary and Future Directions

- Tables 1 and 2 summarizes that, less testing was done on YALE-RECAPTURED dataset and Celebaspoof dataset, and mobile data set which work in different environments and illumination conditions.
- Face Anti-spoofing techniques work with only systems extracted spatial feature; others extracted only temporal features; which needs further improvement.
- There is a need to prepare a hybrid approach for face Anti-Spoofing detection which extracts spatial-temporal feature.
- As a result, to achieve high accuracy there is a need to use hybrid Face Anti-spoofing Techniques i.e. Manual and Deep Learning Based.

4 Conclusion

This paper is an attempt to give an insight on the several datasets which have been examined for anti-spoofing techniques. This paper also offers light on a detailed discussion of the comparative study of face anti-spoofing techniques. Researchers have recently worked with large datasets, but the face recognition system still needs to be improved. As hacking and spoofing techniques become more common, difficulties in developing more secure and crime-free face recognition software are to be expected.

Acknowledgements I would like to express my sincere gratitude to Dr. Jignesh B. Patel, and my colleagues for their research guidance.

References

1. Yu Z, Qin Y, Li X, Zhao C, Lei Z, Zhao G (2021) Deep learning for face anti-spoofing: a survey. In: Computer vision and pattern recognition. IEEE
2. Akhtar Z, Foresti GL (2016) Face spoof attack recognition using discriminative image patches. *J Electr Comput Eng* 14. <https://doi.org/10.1155/2016/4721849>
3. Galbally J, Marcel S, Fierrez J (2014) Biometric anti spoofing methods: a survey in face recognition. *IEEE Access* 1–24
4. Correya M, Thippeswamy G (2018) Face biometric anti-spoofing. *Int J Adv Res Comput Commun Eng* 7(3)
5. Zhang M, Zeng K, Wang J (2020) A survey on face anti-spoofing algorithms. *J Inf Hiding Priv Prot* 2(1):21–34. <https://doi.org/10.32604/jihpp.2020.010467>

6. Boulkenafet Z, Komulainen J, Hadid A (2015) Face anti-spoofing based on color texture analysis. In: 2015 IEEE international conference on image processing (ICIP). Center for Machine Vision Research, University of Oulu, Finland, pp 2636–2640
7. Masood F, Ur Rehman Z, Zia S (2016) Performance analysis of face anti-spoofing feature extraction techniques. *IEEE Access* 1–12
8. Feng L, Po L, Li Y, Xu X, Yuan F, Cheung TC (2016) Integration of image quality and motion cues for face anti-spoofing: a neural network approach. *J Vis Commun Image Represent* 38:451–460. <https://doi.org/10.1016/j.jvcir.2016.03.019>
9. Wang Y, Nian F, Li T, Meng Z, Wang K (2017) Robust face anti-spoofing with depth information. *J Commun Image Represent* 10(1016):332–337
10. Tu X, Fang Y (2017) Ultra-deep neural network for face anti-spoofing. In: International conference on neural information processing. Springer, pp 686–695. https://doi.org/10.1007/978-3-319-70096-0_70
11. Hasan MR, Hasan Mahmud SM, Li XY (2019) Face anti-spoofing using texture-based techniques and filtering methods. *J Phys: Conf Ser* 1229(1). <https://doi.org/10.1088/1742-6596/1229/1/012044>
12. Das PK, Hu B, Liu C (2019) A new approach for face anti-spoofing using handcrafted and deep network features. In: IEEE international conference on service operations and logistics, and informatics (SOLI), pp 33–38. <https://doi.org/10.1109/SOLI48380.2019.8955089>
13. Larbi K, Ouard W, Drir H, Amor BB, Amar CB (2019) DeepColorFASD: face anti-spoofing solution using a multi-channelled color space CNN. In: IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 4011–4016. <https://doi.org/10.1109/SMC.2018.00680>
14. Chou C-L (2021) Presentation attack detection based on score level fusion and challenge-response technique. *J Supercomput* 77:4681–4697
15. Zhang W, Xiang S (2020) Face anti-spoofing detection based on DWT-LBP-DCT features. *Signal Process: Image Commun* 89. <https://doi.org/10.1016/j.image.2020.115990>
16. Shu X, Tang H, Huang S (2021) Face spoofing detection based on chromatic ED-LBP texture feature. *Multimed Syst* 27:161–176
17. Zhou J, Shu K, Liu P, Xiang J, Xiong S (2020) Face anti-spoofing based on dynamic color texture analysis using local directional number pattern. In: Proceedings—international conference on pattern recognition, pp 4221–4228. <https://doi.org/10.1109/ICPR48806.2021.9412323>
18. Boulkenafet Z, Komulainen J, Hadid A (2018) On the generalization of color texture-based face anti-spoofing. *Image Vis Comput* 77:1–9. <https://doi.org/10.1016/j.imavis.2018.04.007>
19. Yu Z, Qin Y, Li X, Zhao C, Lei Z, Zhao G (2021) Deep learning for face anti-spoofing: a survey, pp 1–25
20. Deng P, Ge C, Qiao C, Wei H (2022) Multi-stream face anti-spoofing system using 3D information. In: Digest of technical papers—IEEE international conference on consumer electronics, vol 2022-Janua, pp 0–5
21. Fang M, Damer N, Kirchbuchner F, Kuijper A (2022) Real masks and spoof faces: on the masked face presentation attack detection. *Pattern Recognit* 123. <https://doi.org/10.1016/j.patcog.2021.108398>
22. Li L, Xia Z, Wu J, Yang L, Han H (2022) Face presentation attack detection based on optical flow and texture analysis. *J King Saud Univ - Comput Inform Sci* 34(4):1455–1467. <https://doi.org/10.1016/j.jksuci.2022.02.019>
23. Chang HH, Yeh CH (2022) Face anti-spoofing detection based on multi-scale image quality assessment. *Image Vis Comput* 121:104428. <https://doi.org/10.1016/j.imavis.2022.104428>
24. Cai R, Li Z, Wan R, Li H, Hu Y, Kot AC (2022) Learning meta pattern for face anti-spoofing. *IEEE Trans Inf Forensics Secur* 17:1201–1213. <https://doi.org/10.1109/TIFS.2022.3158551>
25. Muhammad U, Yu Z, Komulainen J (2022) Self-supervised 2D face presentation attack detection via temporal sequence sampling. *Pattern Recogn Lett* 156:15–22. <https://doi.org/10.1016/j.patrec.2022.03.001>

26. Rattani A, Derakhshani R (2018) A survey of mobile face biometrics. *Comput Electr Eng* 72(2018):39–52
27. Belli D, Major B A personalized benchmark for face anti-spoofing, pp 338–348

Ethereum Blockchain-Based Medicine Supply Chain



Jigna J. Hathaliya, Priyanka Sharma, and Sudeep Tanwar

Abstract The medicine supply chain is a process of transferring medicine across various stakeholders. The manual process of the supply chain needs help to locate the source of the medicine accurately. In addition, many people consume the medicine without validating it, which generates a risk for patients. To address this, in this paper, we have designed an Ethereum-based smart contract for tracking medicine in real-time accurately. We have also developed a smart contract to verify the medicine and achieve medicine compliance. Initially, the BCN verifies the credentials of each stakeholder and adds them to the BCN. Afterward, Admin can assign the roles to each stakeholder. The supplier can initiate collecting raw forms and send them to the manufacturer with the raw form ID. The manufacturer can verify the raw forms of medicine with its ID and manufacture the medicine accordingly. Further, this medicine verifies by the medical authority nodes and added medicine in the BCN. Later on, this medicine was transferred across other stakeholders, and locate the source of the medicine accurately. We have also validated our smart contract with vulnerability analysis. We have used the Oyente tool to verify our smart contract validity. Thus, the proposed scheme achieves real-time traceability, transparency, immutability, and medicine compliance.

Keywords Medicine · Blockchain · Ethereum · Supply chain · Vulnerability analysis · Smart contract

J. J. Hathaliya (✉) · S. Tanwar
Department of Computer Science, Institute of Technology, Nirma University, Ahmedabad,
Gujarat 382481, India
e-mail: 19ftphde36@nirmauni.ac.in

S. Tanwar
e-mail: sudeep.tanwar@nirmauni.ac.in

P. Sharma
Samyak Infotech Pvt. Ltd, Ahmedabad, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
S. J. Patel et al. (eds.), *Information Security, Privacy and Digital Forensics*,
Lecture Notes in Electrical Engineering 1075,
https://doi.org/10.1007/978-981-99-5091-1_20

279

1 Introduction

In the entire world, every year number of deaths occur due to the consumption of counterfeit medicine, which impacts the healthcare industry [1]. It is a prime concern of the healthcare industry to deliver authentic medicine to the patients [2]. Medicine supply chain used for the delivery of the medicine, in which medicine transfers from one asset to a different asset. Thus, it requires the authentication of medicine at each stage. But if any unauthorized user can take the place of an authorized user and perform certain operations to misuse the medicine in between the logistics process. In the logistics process, if people are not getting the medicine or there is a delay in medicine, that impacts their recovery. In addition, the manual supply chain does not authenticate each user. Thus, any user can transfer the products and result in unintended activities [3]. Many authors have introduced the digital supply chain to address this concern, which moves medicine from one asset to another with the RFID technology [4]. Sometimes, it does not detect the tag easily.

To solve the issues mentioned above of the supply chain process, BCN provides a promising solution to track and trace the medicine at each stage in the entire supply chain process. The medicine supply chain comprises various stakeholders, including suppliers, manufacturers, warehouses, distributors, shippers, pharmacy shops, and end consumers. BCN authenticates each user and allows them to participate in the medicine supply chain process. Each stakeholder is associated with different roles in this process and performs the operations from the product's origin to the end consumers. It initializes with the user registration with its ID and Ethereum wallet address EthAddress. In this process, Admin has the right to add and remove the users, and the Admin can verify the user's ID and EthAddress and include their ID in the process. Admin also adds and removes the medicine from the valid medicine list. If the medicine sells before, then Admin can remove this medicine from the list and update the list. First, the supplier can collect the raw forms and send it to the manufacturer for manufacturing the medicine with a unique hash ID to track the medicine batches at each stage. After that, medicine batches transfer from the warehouse to the distributor and the pharmacy shop. In this process, each stakeholder shares the real-time medicine update status to locate the medicine and know the history of the medicine. This paper addresses the challenges of the medicine supply chain process and provides an Ethereum-based solution. We have created smart contracts that eliminate third-party holders' use and give rights to access the data. We have also designed access control methods to validate each stakeholder and medicine. The proposed scheme achieves with end-to-end visibility, reliability, traceability, compliance, and security of the supply chain process. We also enlighten the flow of this proposed approach using algorithms to know the functionalities of the scheme. The vulnerability analysis of the smart contracts to see the validity of smart contracts and free from various attacks.

2 Related Work

This section presents the current work related to the BCN-based medicine supply chain process. In the earlier stage, many authors have used RFID-based systems to validate a medicine. Authors in this, [9] proposed a BCN-based counterfeit medicine authentication. The system uses an RFID code and detects expired and defective drugs using a smartphone. It ensures the quality of the drug, transaction security, and data privacy, but they cannot track the medicine at each stage, and it does not make their system reliable. Later on, Lau et al. [7] explored the BCN-based supply chain process to achieve traceability. They also provide the medicine regulations to provide anti-counterfeiting solutions. They have deployed their system on a cloud-based platform and provide an efficient solution for anti-counterfeiting medicine.

Konapure et al. [5] proposed the BCN-based pharmaceutical supply chain process. They have designed smart contracts to track efficient medicine at each level. The system achieves immutability and decentralization but cannot detect counterfeit medicine. Later, [6] addresses this issue and identifies the counterfeit medicine from the supply chain process. They have used a QR code for medicine tracking at each stakeholder. The system helps to locate the source of the counterfeit medicine. This system achieves transparency and security of the medicine, but they did not provide the medicine compliance or discuss drug recall. Further, [8] presented the drug supply chain monitoring based on BCN. In this system, they design smart contracts to assign the roles to each stakeholder and achieve the drug's consistency, traceability, and transparency. The system also identified the counterfeit drug but cannot detect the medicine after it enters the supply chain process and does not validate its smart contract. We proposed an Ethereum-based medicine supply chain process to address all aforementioned issues. The proposed scheme designs a smart contract to assign the roles and locate and track the medicine at each stage in the supply chain process. All valid nodes of BCN can validate the medicine, and it can add only valid medicine. Further, this medicine transfers across each stakeholder in the supply chain process. Thus, it achieves medicine compliance. We have also provided the vulnerability analysis of the smart contract to make our system robust. Table 1 shows the comparative analysis of the existing medicine with the proposed scheme considering the parameters such as objectives, BCN, pros, and cons.

2.1 Research Contributions

The contributions of the scheme are presented as follows:

- Presented an Ethereum-based solution for a medicine supply chain that ensures the system's decentralization, reliability, immutability, security, and end-to-end traceability.

Table 1 Comparative analysis of the existing BCN-based medicine supply chain scheme with the proposed scheme

Author	Year	Objective	Blockchain network	Pros	Cons
Alam et al. [9]	2021	Presented BCN-based counterfeit-medicine authentication	Smartphone-based RFID	Detect expired and defective medicines using RFID tag	Not design a smart contract to track the medicine at each stage
Lau et al. [7]	2021	Designed BCN-based anti-counterfeiting solution with following regularity standards of medicine	Ethereum-based smart contract	Follow regularity standards while manufacturing the medicine	Not provide the vulnerability analysis of the smart contract
Konapure et al. [5]	2022	Proposed smart contract-based system architecture for pharmaceutical supply chain process	Ethereum	Provide medicine traceability at each stage	Not able to detect counterfeit medicine
Anjum et al. [6]	2022	Proposed a BCN-based counterfeit-drug identification	Ethereum-based decentralized App	Identify the counterfeit drug, and achieve traceability using QR code	Does not provide medicine compliance, medicine recall
Dave et al. [8]	2022	Developed BCN-based pharmaceutical supply chain process to monitor the medicine	Ethereum-based smart contracts	Achieves the transparency, traceability, and consistency	Only validate the medicine during the manufacturing stage
The proposed scheme	2022	Ethereum-based Medicine supply chain	Ethereum-based smart contract creation and assign the roles to each stakeholder	Smart-contract vulnerability analysis to make the system reliable, Medicine Compliance, traceability, end-to-end visibility	–

- We develop Ethereum-based smart contracts for each stakeholder to perform the supply chain operations securely and update the medicine status to maintain the traceability of the medicine at every stage.
- Perform vulnerability analysis of the smart contracts to provide a robust solution that secures against various vulnerabilities and attacks.

The remaining paper is as follows. Section 3 describes the proposed approach of the Ethereum BCN-based approach. The results are presented in Sect. 4. Finally, Sect. 5 provides the conclusion.

3 Proposed Architecture

Figure 1 shows the proposed architecture of the medicine supply chain process. This process is associated with different stakeholders performing their roles. Initially, Admin registered with AID and its EthAddress. Admin has rights to verify the stakeholders such as User_ID = SID, SHID, MID, WID, DID, PID, and their EthAddresses included in the BCN. Next, Admin can verify the medicine's compliance by confirming the approval of the medicine by the government, then it includes in the valid medicine list (Approved_Medicine_List), and Admin can include the medicine in the supply chain process and initiates the attribute of medicine status. Admin also verifies the medicine recall status; if it is 0, it proves that it is a valid medicine and has not been sold before to any consumer. Later on, this medicine includes in the valid medicine list in the BCN. Further, Admin can verify the SID and include their ID in the BCN. Suppliers can collect raw materials, make a raw material packet, and transfer it to the manufacturer (MID) for the manufacturing process and update the medicine status. Later, Admin can verify the shipper and include their ID in the BCN. The shipper gets the raw packets with its RM_ID, receiver ID, H_ID, and transporter type and loads the shipment. Moreover, Admin can verify the manufacturer (MID) and include their ID in the BCN. MID verifies the raw packet with RM_ID, SHID, H_ID, and SID, creates new medicine batches, and transfers the shipment with specific MED_ID, H_ID, transporter type, medicine receiver ID, and updates the medicine status. Afterward, Admin verifies and includes WID in the BCN. Next, Warehouse (WID) verifies the medicine batches with MED_ID, MID, SHID, and H_ID and do ready for shipment to the distributor with MED_ID, H_ID, medicine receiver ID and updates the medicine status. In the next stage, Admin verifies the distributor and includes their ID in the BCN. The distributor (DID) verifies the medicine batches with specific credentials and activates the shipper with a specific transporter

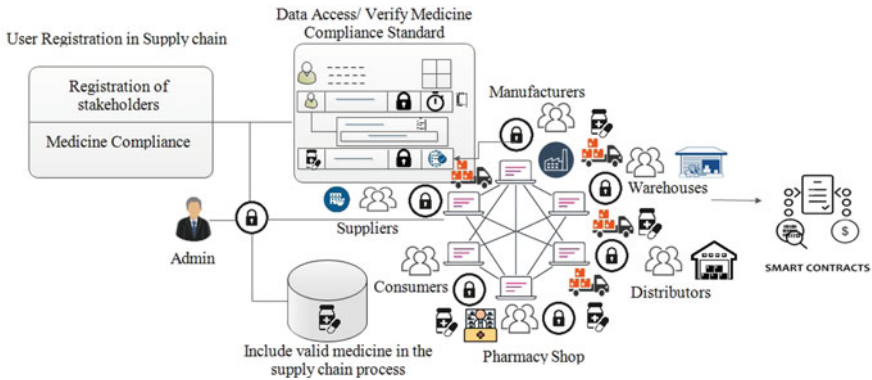


Fig. 1 Proposed architecture

type for transferring medicine from the distributor to the pharmacy shop (PID). Next, Admin can verify the pharmacy shop (PID) and include their ID in the BCN. Next, the Pharmacy shop verifies and receives the medicine batches and updates the medicine status. In addition, the Pharmacy shop shared a sell medicine status, medicine stock information to record the history of the medicine. Thus, This architecture achieves end-to-end visibility, real-time traceability, security, and transparency of the supply chain process.

3.1 Proposed Algorithm

Admin Working Algorithm 1 shows the working process of admin. Initially, Admin (AID) registered with their Ethereum Address (EthAddress), and AID grants access to all stakeholders in the supply chain process. AID can verify the UserID and include the UserID with its EthAddress in the BCN network. Admin also verifies the medicine (MED_ID) and adds it to the valid medicine list. Admin can verify the medicine recall status, and if it is 0, then medicine is not previously sold, prove that MED_ID is valid, and add medicine to the supply chain process. The supply chain initializes at the supplier stage, transfers medicine to the other stakeholders, and updates the medicine status accordingly. This feature tracks the real-time status of the medicine and achieves traceability. The complexity of Algorithm 1 is $O(n)$.

Algorithm 1 Algorithm for Admin working

START**Input:** Admin ID (AID) Registration with EthAddress**Output:** Grant_Access to all stakeholders $S_E, M_E, SH_E, W_E, D_E,$ and P_E transactions \in BCN**Initialization:** Admin should be valid node. Admin can Read / Write / Update / Include / Withdraw Nodes ($S_{ID}, M_{ID}, SH_{ID}, W_{ID}, D_{ID}, P_{ID}$)

```

1: procedure ADMIN(AID)
2:   while (True) do
3:     if (User(ID) IS VALID) then
4:       if (UserID  $\notin$  BCN) then
5:         Include_User to the BCN
6:         Include_User(ID), User_EthAddress, BCN
7:       else
8:         Update_User_List(ID), in BCN
9:       end if
10:    else
11:      Remove_User(ID)
12:    end if
13:    if (MED_ID  $\in$  VALID_MED_LIST) then
14:      if MED_ID  $\notin$  BCN) then
15:        Include_Medicine to the BCN
16:        Include_Medicine (MED_ID, BCN)
17:      else
18:        if (Medicine_Recall_Status = 0) then
19:          MED_ID is Valid
20:          Medicine_Update_Status = 0
21:        else
22:          Remove_MED_ID from VALID_MED_LIST
23:        end if
24:      end if
25:    else
26:      Remove_MED_ID from VALID_MED_LIST
27:    end if
28:  end while
29: end procedure

```

Medicine Supply Chain Process The Algorithm 2 shows the flow of the supply chain process. Admin can initiate the supply chain process, verify the supplier (SID), assign the role, and include their ID in the BCN. The supplier can make raw packets and update the medicine status. Next, the Admin can verify the shipper and add their ID to the BCN. The shipper can check the transporter type and load the shipment with RM_ID or MED_ID, H_ID, and MedicineReceiver ID and its EthAddress. Afterward, Admin can verify the manufacturer MID and include their ID in the BCN. The manufacturer can verify the raw material with the RM_ID, SID, H_ID, and SH_ID. Next, MID can make medicine batches with medicine ID MED_ID and update the medicine status. After that, the procedure shipper activates with transporter type “2” and loads the shipment with specific MED_ID, H_ID, and MedicineReceiver ID and its EthAddress. Admin can verify the warehouse (WID) and include their ID in the BCN. WID verifies the medicine batches with specific MED_ID, M_ID, SH_ID, and H_ID, updates the status and calls the shipper procedure with transporter type “3,” and loads the shipment with MED_ID, H_ID, and MedicineReceiver ID and its EthAddress. Similarly, Admin can verify distributor (DID) and Pharmacy shop (PID) and include them into BCN. DID and PID verify the shipment with MED_ID,

Algorithm 2 Process Flow of the Medicine Supply Chain Process

START Input: Admin can initiate the supply chain process

Output: Assign the specific role to each stackholder

Initialization: A_E should be valid node Include_UserID / Withdraw_UserID / Assign_Roles / Revoke_Role / Verify_UserID / Verify_UserEthAddress / Verify_MED_ID / Include_MED_ID

```

1: procedure SUPPLIER((SID))
2:   while (True) do
3:     if (SID IS VALID) then
4:       Made_Raw_packet (RM_ID)
5:       Update_Medicine_Status
6:     else
7:       Withdraw_SID
8:     end if
9:   end while
10: end procedure
11: procedure SHIPPER((SHID))
12:   while (True) do
13:     if (SHID IS VALID) then
14:       if (Transporter_type = 1) then
15:         Shipment_Load (RM_ID, S_ID, H_ID, M_ID)
16:         Update_Medicine_Status
17:       else if (Transporter_type = 2) then
18:         Shipment_Load (MED_ID, M_ID, H_ID, W_ID)
19:         Update_Medicine_Status
20:       else if (Transporter_type = 3) then
21:         Shipment_Load (MED_ID, W_ID, H_ID, D_ID)
22:         Update_Medicine_Status
23:       else if (Transporter_type = 4) then
24:         Shipment_Load (MED_ID, D_ID, H_ID, P_ID)
25:         Update_Medicine_Status
26:       end if
27:     else
28:       Withdraw_SHID
29:     end if
30:   end while
31: end procedure
32: procedure MANUFACTURER(MID)
33:   while True do
34:     if (MID IS VALID) then
35:       Verify_RawMaterial(RM_ID, S_ID, H_ID, SH_ID)
36:       Receive_RawMaterial(RM_ID)
37:       Manufacture_Medicine_Batches(MED_ID)
38:       Update_Medicine_Status
39:       Procedure(Shipper)(SHID, Transporter_type = '2')
40:     else
41:       Withdraw_MID
42:     end if
43:   end while
44: end procedure
45: procedure WAREHOUSE(WID)
46:   while True do
47:     if (WID IS VALID) then
48:       Verify_Medicine_Batches(MED_ID, M_ID, H_ID, SH_ID)
49:       Receive_Medicine_Batches(MED_ID)
50:       Procedure(Shipper)(SHID, Transporter_type = '3')
51:       Update_Medicine_Status
52:     else
53:       Withdraw_WID
54:     end if
55:   end while
56: end procedure
57: procedure DISTRIBUTOR(DID)
58:   while True do
59:     if (DID IS VALID) then
60:       Verify_Medicine_Batches(MED_ID, W_ID, H_ID, SH_ID)
61:       Receive_Medicine_Batches(MED_ID)
62:       Procedure(Shipper)(SHID, Transporter_type = '4')
63:       Update_Medicine_Status
64:     else
65:       Withdraw_DID
66:     end if
67:   end while
68: end procedure

```

MedicineSender ID, SH_ID, and H_ID, receive the medicine batches accordingly, and update the medicine status. If the Admin finds any malicious activity, withdraw the User ID from the BCN. The complexity of an algorithm is $O(n)$.

4 Results

This section presents the result of the proposed scheme. In this scheme, we have developed smart contracts for each stakeholder. Each stakeholder is associated with their role in this medicine supply chain process. Admin can verify the users with their ID and EthAddress, and only valid nodes add to the BCN.

Figure 2 shows the user registration in which only the Admin can add or remove the users. After that, the supplier can collect the raw forms and send them to the manufacturer for further manufacturing.

Further, Admin also has the access rights to verify the medicine and include it in the BCN.

Figure 3 shows that medicine is included in the BCN. After that, medicine transfers from the manufacturer to the warehouse, distributor, and pharmacy shop accordingly, and similarly update, the medicine status at each stage.

```
{
  "from": "0xd9145CCE52D386F254917e481e844e9943F39138",
  "topic": "0x4f21f3f17ed270f2364768b3b974273cef2ab48718556fb5c9fb67147b31f570",
  "event": "UserRegistration",
  "args": {
    "0": "0x5c680f78f3E7ce0460398d8FABdfD3f9F5021678",
    "1": "0x54686973206973206a69676e612068617468616c697961206a68736468646864",
    "EthAddress": "0x5c680f78f3E7ce0460398d8FABdfD3f9F5021678",
    "Name": "0x54686973206973206a69676e612068617468616c697961206a68736468646864"
  }
}
```

Fig. 2 User registration

```
{
  "address MRA": "0xAb8483F64d9C6d1EcF9b849Ae677d03315835cb2",
  "bytes32 Desn": "0x5468652061646472657373206973206e69726d6120756e69766572736974792c",
  "bytes32 RF": "0x676c75636f73652c20666973686f696c2c206e65656d2c2067696e6765722020",
  "uint256 Quantity": "2",
  "address Medicineshipper": "0x4B209938c481177ec7E8f571ceCaE8A9e22C02db",
  "address MedicineCollector": "0x78731D3Ca6b7E34aC8F824c42a7cC18A495caba8",
  "uint256 MedicineCollectorType": "1"
}
```

Fig. 3 Add medicine

4.1 Vulnerability Analysis

This section describes the vulnerability analysis of the smart contract.

Figure 4 shows the vulnerability analysis of the smart contract. According to the security mentioned above analysis, the Ethereum smart contract developed for medicine traceability explored using efficient tools to discover any vulnerabilities in any code. These tools were applied during development to enhance the smart contract’s reliability. The smart contract is created using the Remix IDE, which offers some code debugging and run-time error alerts. Moreover, building trust in the smart contract is not adequate to achieve the system’s robustness. To address the issue, SmartCheck is used to find code vulnerabilities from various levels of security. The output indicated that the smart code was bug-free following numerous iterations of modification. By comparing the smart contract to its knowledge base and analyzing it, SmartCheck confirmed that it was free from vulnerabilities that may have exposed it to abuse and cyberattacks. The Oyente tool [10] was also used to investigate the security of smart contracts. Oyente, a Linux-based program, thoroughly examines the code to isolate potential security holes. It intends to defend against well-known vulnerabilities like call stack depth attacks and re-entrancy attacks against the Ethereum smart contract. After verification of the smart contract, the Oyente tool generates the report by checking each parameter. Figure 4 shows the presence of the vulnerabilities in the smart contract.

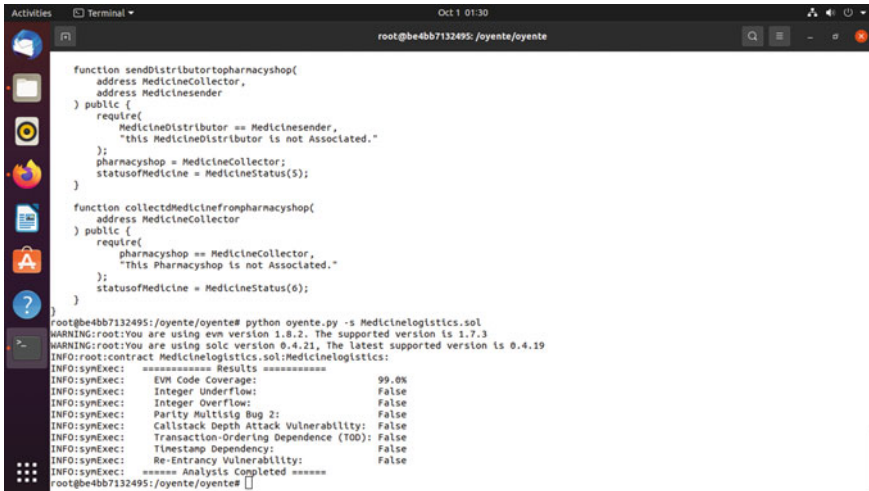


Fig. 4 Vulnerability of smart contract

5 Conclusion

We have designed an Ethereum-based smart contract to assign a role to each stakeholder. Each smart contract performs its assigned roles and verifies and tracks the medicine at each stage. The proposed scheme also verified the medicine by validating its medicine status parameter. In this verification process, all valid nodes participate and check the medicine compliance with regularity standards of medicine, and if it is valid, add medicine to the BCN. The proposed scheme provides end-to-end visibility from supplier to end consumer. Thus, every end consumer can see and locate the medicine and verify the medicine status. We have also applied smart contract vulnerability analysis to make our system robust. Therefore, our proposed scheme provides transparency, security, decentralization, and robustness. In the future, we will make our system more reliable using a permissioned network.

References

1. Uddin M, Salah K, Jayaraman R, Pesic S, Ellahham S (2021) Blockchain for drug traceability: architectures and open challenges. *Health Inform J* 27(2). <https://doi.org/10.1177/14604582211011228>
2. Musamih A et al (2021) A blockchain-based approach for drug traceability in healthcare supply chain. *IEEE Access* 9:9728–9743. <https://doi.org/10.1109/ACCESS.2021.3049920>
3. Jangir S, Muzumdar A, Jaiswal A, Modi CN, Chandel S, Vyjayanthi C (2019) A novel framework for pharmaceutical supply chain management using distributed ledger and smart contracts. In: 2019 10th international conference on computing, communication and networking technologies (ICCCNT), pp 1–7. <https://doi.org/10.1109/ICCCNT45670.2019.8944829>
4. Abbas K, et al (2020) A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry. *Electronics* 9.5: 852. <https://doi.org/10.1109/HPDC.2001.945188>
5. Konapure RR, Nawale SD (2022) Smart contract system architecture for pharma supply chain. In: International conference on IoT and blockchain technology (ICIBT), pp 1–5. <https://doi.org/10.1109/ICIBT52874.2022.9807744>
6. Anjum N, Dutta P (2022) Identifying counterfeit products using blockchain technology in supply chain system. In: 2022 16th international conference on ubiquitous information management and communication (IMCOM), pp 1–5. <https://doi.org/10.1109/IMCOM53663.2022.9721789>
7. Lau WF, Liu DYW, Au MH (2021) Blockchain-based supply chain system for traceability, regulation and anti-counterfeiting. In: IEEE international conference on blockchain (blockchain), pp 82–89. <https://doi.org/10.1109/Blockchain53845.2021.00022>
8. Dave M, Patil K, Jaiswal R, Pawar R (2022) Monitoring supply chain of pharmaceutical drugs using blockchain. In: IEEE Delhi section conference (DELCON), pp 1–5. <https://doi.org/10.1109/DELCON54057.2022.9753598>
9. Alam N, Hasan Tanvir MR, Shanto SA, Israt F, Rahman A, Momotaj S (2021) Blockchain based counterfeit medicine authentication system. In: 2021 IEEE 11th IEEE symposium on computer applications & industrial electronics (ISCAIE), pp 214–217. <https://doi.org/10.1109/ISCAIE51753.2021.9431789>
10. Luu L, Chu D-H, Olickel H, Saxena P, Hobor A (2016) Making smart contracts smarter. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (CCS '16). association for computing machinery. New York, NY, USA, pp 254–269. <https://doi.org/10.1145/2976749.2978309>

Machine Learning Algorithms for Attack and Anomaly Detection in IoT



Rahul Kushwah and Ritu Garg

Abstract With the invention of IoT and its range of smart applications, people’s life has been reformed drastically. IoT infrastructure consists of actuators and sensors that generate a massive volume of data that requires extensive computing. Due to the constrained nature of IoT devices, they can be easily exploited. Moreover, some of the IoT nodes behaves abnormally which makes the IoT infrastructure vulnerable that could be exploited by the adversary to gain unauthorized access or to perform other malicious deed. Thus, anomaly detection is a primary concern in the IoT infrastructure, and the investigation of IoT network for detecting anomalies is a fast-emerging subject. There is a plethora of techniques which have been developed by the researchers for detecting anomalies. In this article, we have emphasized on the machine learning-based anomaly detection techniques due to their ability to bring out accurate results and predictions. Further, to provide a detailed overview, we have categorized the machine learning-based anomaly detection techniques into unsupervised and supervised learning. This study helps the researchers to get a better idea of machine learning techniques that have been employed for the detection of anomaly in the IoT infrastructure.

Keywords Internet of Things · Machine learning · Unsupervised learning · Supervised learning

R. Kushwah · R. Garg (✉)
Department of Computer Engineering, National Institute of Technology, Kurukshetra 136119,
Haryana, India
e-mail: ritu.59@nitkkr.ac.in

R. Kushwah
e-mail: rahul_32113211@nitkkr.ac.in

1 Introduction

The invention of the Internet transformed human communication. Likewise, IoT smart gadgets are changing how people perceive as well as interact with their surroundings. Its popularity has allowed it to dominate the digital market. IoT is the following trend that is steadily gaining popularity. One gadget may connect with other devices through the Internet without the need for a human intermediary, making it a more widely used and technologically advanced technology. This is the rationale behind IoT device numbers rising quickly, which suggests that IoT infrastructure is becoming more sophisticated. According to the survey, more than 76.54 billion IoT gadgets will be installed worldwide by 2026 which was 2 billion in 2006 [1]. Because IoT devices are more sophisticated and smarter than traditional technology, IoT has earned a position in every industry, including healthcare and agriculture. Apart from the extreme usage and popularity, IoT devices are more liable to security threats and attacks because mostly IoT devices are resource constrained and smaller in size. Security flaws in IoT devices are very frequent. Moreover, IoT devices can be easily exploited by the attacker as new vulnerabilities arise each day in IoT infrastructure. The wired network is considerably safer than the wireless network since, in the wireless network, communications conducted through air/radio waves which can be easily intercepted by the adversary. Thus, IoT devices are less secure since they use wireless media to transmit data, which makes it easier for attackers to target specific IoT devices. Further, the attacker can also exploit the abnormal IoT node in the network to get the illegitimate access or for other means. Thus, it is quite important to identify such abnormal nodes to prevent fore coming abnormalities in the network to avoid events like unauthorized access or data theft. Anomaly detection is the technique to identify abnormal nodes and other abnormalities.

IoT infrastructure anomaly detection techniques mainly rely on human intervention and optimization for an immediate cure. Theoretically, an anomaly is easy to understand, and an expert in domain may easily identify anomalous data given sufficient time but we need to develop a self-operating model to detect anomalies in the system. However, the creation of a smart anomaly detection model in an IoT infrastructure is difficult, and to develop such smart model, numerous researches have considered machine learning algorithms due to their better predicting quality.

IoT real-time applications produce exponentially large data streams, which present certain restrictions and challenges to machine learning algorithms [2]. Due to these challenges, the algorithm must be designed carefully to process data produced by IoT devices [3]. The majority of currently existing data stream methods are not so much effective and have fewer system requirements. Anomaly detection strategies have been the subject of several researches, including [4–7] that use statistical and machine learning techniques to handle static data and data streams. These studies, however, have not given much attention to evolving data streams.

For detecting anomalies in data streams, numerous approaches, such as non-parametric, Hierarchical clustering, and lightweight approach, identify connected

sensors and generate clusters [4]. A large number of data stream constraints/challenges should be taken into account for an effective anomaly detection method.

- Continuous data points are sent out, and the rate at which they arrive depends on the source of data. Therefore, it could be quick or slow.
- There may be no end to the incoming data since the dataset may be limitless.
- The characteristics and/or features of the incoming data sources might change.
- Data points are possibly one-pass, meaning they can only be utilized once before being deleted. Thus, retrieving important data attributes is essential.

Further, algorithms must be able to cope with the following challenges to have high-quality anomaly detection:

- Ability to handle rapid data: Since datasets may be large and should be processed in a single pass, the anomalies detection technique should be able to tackle as well as to analyze data in actual time if the data points from the data source come in a continuous manner.
- Ability to manage variable streams of data: Nature of data sources is diverse, constantly growing as well as migrating over time. As a result, the data generated would alter, and the results of the algorithm may change significantly. Stream activity is known to change over time, necessitating a unique approach to address them.
- The ability to deal with dimensionality: Choosing the appropriate feature vectors or dimensionality that might lead to better clusters in high-dimensional data is an extra challenge.

The majority of research literature has explored anomalies using batch processing-centric machine learning algorithms. Thus in this survey, we are analyzing sniping research on machine learning for finding anomalies in datasets, especially for dynamic data streams.

2 Background

Datasets are expanding quickly as the use of IoT devices increased in daily life. Datasets are never-ending, continuous records of data that are organized as well as followed by embedded/exact timestamps [3]. For many real applications, effective analysis of such datasets provides insightful information. Anomaly detection is the process of identifying unusual appearances or patterns in a dataset that is considerably different from the majority of items, these unusual patterns are also known as anomalies, noise, surprises, exceptions, deviations, or novelties [8–10]. Anomaly detection is one of the methods for efficiently assessing the data stream that has been gathered [4]. The approaches used for anomaly detection techniques heavily depend on types of anomaly, anomaly detection, and nature of data.

2.1 Nature of Data

The approach used to detect anomalies is selected on the basis of the type of data being analyzed. IoT is a significant source of data stream. This data series contains three distinguishing characteristics. Firstly, there is a continuous flow of data (i.e., data stream). Thus, the technique must analyze the data in a specific amount of time. Secondly, the limitless flow of data, i.e., the number of data points that may be entered is limitless. Finally, data streams develop, or alter over time [4]. Moreover, anomalies in data streams have significant and practical applications in a wide range of domains. The importance of detecting abnormalities in the data stream in practical implementations increases its growth by assuring accuracy as well as immediacy [11]. Various methods have been suggested for abnormality detection in data streams, including C-LOF [3], combination of BDLMs and RBPF [12], xStream [13], CEDAS [14], KPI-TSAD [15], HTM [16], TEDA Clustering [17], AutoCloud [18], Multiple kernel learning [19], MuDi-Stream [20], evolving spiking neural network [21], ensembles neural networks [22].

2.2 Anomaly Types

Depending on its nature anomalies can be categorized into three types.

1. **Point Anomaly:** A point anomaly is one in which one object may be shown to differ from other objects. Since this is the most basic kind of anomaly, many studies include it.
2. **Contextual Anomalies:** Contextual Anomalies occur only in case when an object is abnormal in a specific situation.
3. **Collective Anomalies:** Certain connected objects can be seen as an abnormality when compared to other objects. In this scenario, only a collection of objects may be abnormal.

Early anomaly detection is advantageous in virtually all use cases. When a heart patient heart rate is continually monitored by a gadget, then an anomaly or any abnormality might cause a heart attack. It is far desirable to detect such abnormalities minutes in advance rather than a few seconds or right after the incident [23]. A wide range of different fields can be benefited from and use the ability to detect abnormalities in the dataset [11].

Machine Learning (ML) algorithms is one the popular method used for the anomaly detection. ML allows computers to learn without being explicitly programmed [24]. It uses the knowledge from the past years or the present to forecast the future or make judgments [6]. Thus, “learning” is very important in ML. The primary goal of ML is to develop a model that really can accurately and automatically detect essential patterns in data [12]. Recently the use of deep learning-based techniques made it possible to enhance the anomaly detection efficiency as compared to the classical techniques.

3 Classification of Machine Learning-Based Anomaly Detection Techniques

There is a plethora of methods used for detecting anomalies in the IoT devices like machine learning, deep learning, etc. ML techniques appear to be a viable solution for the anomaly detection due to its ability to make accurate prediction via processing the past datasets. Thus, in this article, we have emphasized on the various cutting-edge methods for detecting anomalies in IoT data streams using machine learning which was presented by the researchers recently. We have divided the machine learning-based anomaly detection techniques into two categories, namely, supervised learning and unsupervised learning. The detailed taxonomy of the anomaly detection methods using machine learning is given in Fig. 1.

3.1 Unsupervised Learning Techniques

Unsupervised learning employs machine learning algorithms to examine and organize unlabeled information without the assistance of humans. Due to the ability of these algorithms to identify occult patterns or data clusters in the data streams, researchers have utilized unsupervised learning schemes for developing precise and accurate techniques for anomaly detection. We have presented a detailed overview of unsupervised techniques developed by the researcher for detecting anomalies in the IoT infrastructure.

In the article [3], the authors developed a novel data stream method for anomaly detection called Cumulative Local Outlier Factor (C-LOF) based on a popular and effective anomaly detection approach called Local Outlier Factor (LOF) which only works on batch data. The proposed C-LOF approach is able to resist masquerade while identifying irregularities in the data stream. The temporal complexity of C-LOF is a significant disadvantage of this method.

In the article [16], the author utilized hierarchical temporary memory technique to propose a real-time data stream anomaly identification approach for space imagers. The proposed method detects geographical and temporal anomalies in continuous and real-time data stream anomaly detection without supervision. The results suggest that the method is capable of detecting anomalies in real time. However, the proposed solution was unable to lower the false positive rate. Further, in the article [25], the author continuously analyzes the resource usage of completed process activities and utilizes HTM to find abnormalities in real-time network measurements. The proposed model can process unsupervised data streams and successfully adapt to changes in the data depending on the underlying statistics. The suggested model is capable to detect output variations appropriately based on the resource utilization characteristics caused by various competing tasks. However, this reasoning has a serious flaw in that there is still room for development in terms of accuracy and latency. Moreover, in the paper [26], HTM has been utilized by the authors to develop

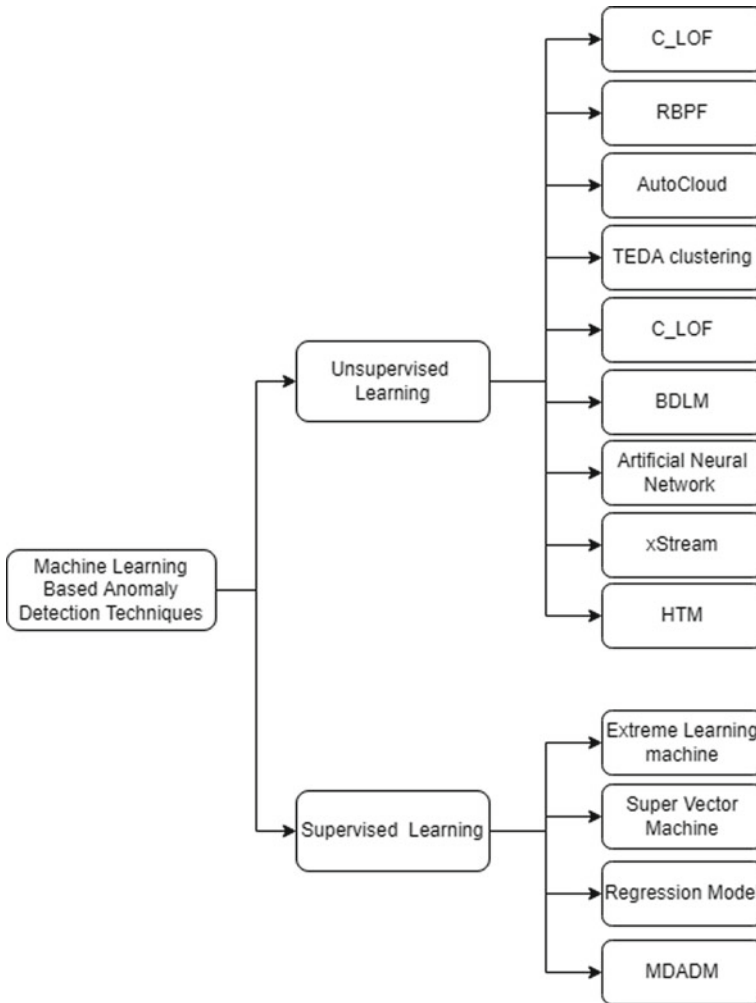


Fig. 1 Classification of techniques for anomaly detection

an online sequence memory-based technique for detecting anomalies. The proposed technique can identify temporal and spatial abnormalities within noisy domains and predictable. The approach accomplishes real-time online detection criteria without supervision. The proposed approach was much effective than any other HTM-based techniques. The research, however, did not explore assessing the method using an actual dataset with anomalies having a high dimensionality.

Cauteruccio et al. [27] utilized an artificial neural network technique inspired by the data edge exploration mixed with cloud data for automatic anomalies within heterogeneous sensor networks. The suggested method was evaluated experimentally using real data collected from the IoT sensor network and then deformed by numerous

virtual disabilities. The obtained result demonstrates that the suggested technique self-adapts to changes in the environment and appropriately detects abnormalities. Further, it also analyzes the performance of the short- and long-term algorithms that detect anomalies in Wireless sensor nodes. The research, however, does not take into account how the data's characteristics change over time or how drift is defined.

Manzoor et al. [13] proposed a technique called xStream to resolve the issue of abnormalities detection in evolving data streams. The proposed technique is a density-based anomaly detection method with three key features: firstly, it is a constant-time and constant-space techniques, then it measures abnormalities on granularities or multiple scales, and finally, by using distance-preserving projections, it can handle high-dimensionality. Experiments demonstrate that xStream is efficient and exact for evolving streams with low space overhead.

Bezerra et al. [18] introduced an improved technique based on data stream clustering called AutoCloud, which is an Eccentricity and Typicality Data Analytics approach for the detection of anomaly. AutoCloud is a recursive and online method that doesn't require any training or prior understanding of the dataset. Since the technique is online and recursive, it is appropriate for real-time applications because the technique can also handle concept evolution and concept drift, which are fundamental issues in the data streams. The suggested method, meanwhile, ignores various issues with distance that dynamically impact how the clouds are separated.

Maia et al. [17], proposed a dynamic clustering method formed on a combination of topicalities. It is inspired by the framework of TEDA and divides the clustering issue into two parts: macro- and microclusters. According to experiment results, the suggested technique can produce satisfactory outcomes for clustering data and estimating density even when there are variables that impact data distribution characteristics, like concept of drifting.

Hyde et al. [14] proposed an online approach for Clustering Evolving Data Streams into Arbitrary-Shaped (CEDAS) clusters. The proposed approach is a two-step approach, i.e., beneficial and noise-resistant with respect to computing and memory applications in the context of short latency as the count of data dimensions increases. The technique's first stage generates microclusters, while the second stage merges these microclusters into larger clusters. The author demonstrates the proposed technique's abilities to join and split macroclusters as they expand in a completely online mode, and compares it to other online/offline hybrid solutions in terms of accuracy, purity, and speed.

3.2 Supervised Learning Techniques

The process of educating or training a computer system using labeled data is referred to as supervised learning. A supervisor serves as an educator during supervised learning. Researchers have suggested that supervised machine learning techniques are the viable solution for developing flawless and precise methods for the detection

of anomaly. In this section, we have discussed the supervised learning-based anomaly detection schemes.

Peng et al. [28] developed, a multi-source Multi-Dimensional Data Anomaly Detection Method (MDADM) inspired by the hierarchical edge computing which aims to provide early warning of an accident in an underground mining environment using IIoT (Industrial Internet of Things). The obtained result demonstrates that the technique outperforms in terms of detection performance and transmission latency.

Bose et al. [29] designed an SVM (Support Vector Machine) algorithm to guarantee a safe driving experience by providing drivers with actual-time alerts and guidance on road. The system classifies driving events, such as braking and acceleration and road abnormalities, like potholes and bumps, using the popular machine learning approach SVM. It then employs the local Fast Dynamic Time Warping (FastDTW) approach to provide drivers instructions and real-time alerts. According to the experimental results locally running FastDTW provided an accuracy of 86.36%, while SVM provided an accuracy of 95.45% for the detection of driving events. However, the study disregards the types and characteristics of moving items on the highways.

Farshichi et al. [30] created a regression-based method for identifying contextual abnormalities in the air traffic control system. By utilizing the method for modification detection and time spans on contextual abnormalities, they also have details for an upgrade that can be put into practice. The results of the suggested model reveal minimal latency anomaly detection with excellent accuracy.

In the article [19], the author proposed a novel method based on a multi-kernel approach for categorizing non-stationary data streams. This method addresses problems in abnormalities detection such as concept evolution, recurring concepts, idea drift, and infinite period. After learning the exact labels of the instances, these kernels were regularly changed in the stream. Newly received instances will be rated in the function areas according to how far they are away from the limits of the already defined categories (Table 1).

4 Result

Going through the survey, we found that there are still several unresolved research problems despite the advancements in the anomaly detection. Nowadays, a significant amount of data is produced in the form of problematic data streams, it is imperative to effectively tackle the challenge of recognizing abnormalities within changing data streams. Due to the prevalence of high-dimensional data, the majority of existing data stream techniques for anomaly identification are becoming less effective. As a result, it's essential to rebuild the present models in order to properly and effectively detect abnormalities.

The overview of the studied anomaly detection strategies and their ability to satisfy the standards for a high-quality anomaly detection methodology is shown in Table 2. The studied approaches have been examined for their capacity to handle noisy data,

Table 1 An overview of machine learning methods for finding anomalies in data streams

Methods	Type of anomaly point contextual	Nature of data	Dataset	Type of anomaly detection	Pros	Cons
AutoCloud [18]	✓	Data stream	Artificial and real data	UL using clustering	Can handle drift	Do not consider clouds separation
TEDA clustering [17]	✓	Data stream	Own synthetic dataset	UL using clustering	Suitable for clustering data stream	Doesn't account independent process
HTM [16]	✓	Data stream	Space imagers	UL using HTM	Real-time anomaly detection	High false positive rate
CEDAS [14]	✓	Data stream	KKDCup99	UL using clustering	High accuracy and speed	Discard the feature development
xStream [13]	✓	Data stream	Spam-SMS Spam-URL	UL based on density	Low space overhead	High complexity
Multi-kernel [19]	✓	Data stream	KDD99 Cover type	UL based on multi-kernel learning	Handling non-stationary data	High complexity
ANN [27]	✓	Continuous and image data	NA	UL based on pattern of WSN nodes	Self-adapted technique	Doesn't take into account how drift is defining
C_LOF [3]	✓	Data stream	Synthetic and real-life dataset	UL using density	Resist masquerade	High temporal complexity
SVM [29]	✓	Continues data	Own dataset	SL on historical data	High accuracy	Discard the characteristics of moving item
MDADM [28]	✓	Continuous and image data	NA	SL	High scalability	Lower transmission latency
Regression model [30]	✓	Continuous data	NA	SL on historical data	Minimal latency	Unable to find faulty values
Combination of BDLM and RBPF [12]	✓	Data stream	Artificial dataset	UL using density	Performing real-time analysis	Complex

Table 2 Summary of anomaly detection techniques

Methods	Scalability	Evolving feature	Limited time	Limited memory	Projection	Accuracy (%)
AutoCloud [18]	✓			✓		95
TEDA clustering [17]	✓			✓		NA
HTM [16]				✓	✓	94
CEDAS [14]	✓			✓		92
xStream [13]			✓		✓	84
Multi-kernel [19]	✓	✓			✓	NA
ANN [27]	✓				✓	86
C_LOF [3]	✓				✓	NA
SVM [29]	✓				✓	95.45
MDADM [28]	✓				✓	92
Combination of BDLM and RBPF [12]	✓					86.4

conduct data projection, and operate within memory and temporal constraints. Additionally, they address issues with growing features, high-dimensional data, evolving data, and eventually scalability.

5 Research Gap and Future Direction

The study included many anomaly detection methods; however, there are still a lot of issues that need to be addressed. There is presently no one optimal technique for resolving the issue; rather, a variety of methods may be effective for specific data kinds and application areas. We discovered the following serious issues:

1. Accuracy

Despite learning algorithms' ability to detect and classify aberrant behavior in real time, such techniques are still vulnerable to modification to increase accuracy, including a reduction in spurious detection rate, particularly in large-scale sensor nodes [14, 16].

2. Scalability

Scalability is also another essential parameter for anomaly detection techniques since many algorithms lose efficiency when dealing with big amounts of data [19].

3. Time Complexity

The major feature of the data stream is the huge volume of data that arrives continually, requiring the algorithm to perform in real time. However, the time complexity of identifying the anomalies would be a significant difficulty [3], due to the constant exchange between speed and accuracy complexity.

4. Parameter Selection

IoT data streams are frequently produced from non-stationary settings with no prior knowledge of the data distribution, which has an impact on the selection of the right set of model parameters for anomaly detection [17].

5. Data Complexity and Noise

One of the biggest obstacles to developing a model for anomaly detection is data complexity, which includes unbalanced datasets, unexpected sounds, and redundancy within the data [28]. To gather pertinent data and expertise, well-developed dataset curation methods are needed.

6 Conclusion

Anomaly detection has gotten a lot of interest from researchers in recent years, due to the development of low-cost, high-impact sensor technologies in a variety of application domains. Anomaly detection dramatically reduces functional threats, eliminates unknown complexities, and minimizes process downtime. Machine learning algorithms play critical roles in detecting data stream abnormalities in a wide range of IoT application areas. However, various challenges remain to be solved in order to address the issue of anomaly detection. So, in this paper, we have presented a systematic review of the machine learning techniques used for anomaly detection. We have categorized the machine learning techniques into unsupervised and supervised learning for better clarification. This article helps the researchers to obtain extensive information about the recently established machine learning methods for detecting anomalies in IoT data.

References

1. Ratasich D, Khalid F, Geissler F, Grosu R, Shafique M, Bartocci E (2019) A roadmap toward the resilient internet of things for cyber-physical systems. *IEEE Access* 7:13260–13283
2. Kozitsin V, Katser I, Lakontsev D (2021) Online forecasting and anomaly detection based on the ARIMA model. *Appl Sci* 11(7):3194
3. Yu K, Shi W, Santoro N (2020) Designing a streaming algorithm for outlier detection in data mining—an incremental approach. *Sensors* 20(5):1261
4. Munir M, Siddiqui SA, Dengel A, Ahmed S (2018) DeepAnT: a deep learning approach for unsupervised anomaly detection in time series. *IEEE Access* 7:1991–2005
5. Donevski M, Zia T (2018) A survey of anomaly and automation from a cybersecurity perspective. In: 2018 IEEE Globecom workshops (GC wkshps), December. IEEE, pp 1–6

6. Habeeb RAA, Nasaruddin F, Gani A, Hashem IAT, Ahmed E, Imran M (2019) Real-time big data processing for anomaly detection: a survey. *Int J Inf Manage* 45:289–307
7. Fahim M, Sillitti A (2019) Anomaly detection, analysis and prediction techniques in iot environment: a systematic literature review. *IEEE Access* 7:81664–81681
8. Mahdavinejad MS, Rezvan M, Barekatin M, Adibi P, Barnaghi P, Sheth AP (2018) Machine learning for Internet of Things data analysis: a survey. *Digit Commun Netw* 4(3):161–175
9. Vilenski E, Bak P, Rosenblatt JD (2019) Multivariate anomaly detection for ensuring data quality of dendrometer sensor networks. *Comput Electron Agric* 162:412–421
10. Baydargil HB, Park JS, Kang DY (2021) Anomaly analysis of Alzheimer’s disease in PET images using an unsupervised adversarial deep learning model. *Appl Sci* 11(5):2187
11. Ding N, Ma H, Gao H, Ma Y, Tan G (2019) Real-time anomaly detection based on long short-term memory and Gaussian mixture model. *Comput Electr Eng* 79:106458
12. Nguyen LH, Goulet JA (2019) Real-time anomaly detection with Bayesian dynamic linear models. *Struct Control Health Monit* 26(9):e2404
13. Manzoor E, Lamba H, Akoglu L (2018) xStream: outlier detection in feature-evolving data streams. In: *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, July, pp 1963–1972
14. Hyde R, Angelov P, MacKenzie AR (2017) Fully online clustering of evolving data streams into arbitrarily shaped clusters. *Inf Sci* 382:96–114
15. Qiu J, Du Q, Qian C (2019) Kpi-tsad: a time-series anomaly detector for kpi monitoring in cloud applications. *Symmetry* 11(11):1350
16. Song L, Liang H, Zheng T (2019) Real-time anomaly detection method for space imager streaming data based on HTM algorithm. In: *2019 IEEE 19th international symposium on high assurance systems engineering (HASE)*, January. IEEE, pp 33–38
17. Maia J, Junior CAS, Guimarães FG, de Castro CL, Lemos AP, Galindo JCF, Cohen MW (2020) Evolving clustering algorithm based on mixture of typicalities for stream data mining. *Futur Gener Comput Syst* 106:672–684
18. Bezerra CG, Costa BSJ, Guedes LA, Angelov PP (2020) An evolving approach to data streams clustering based on typicality and eccentricity data analytics. *Inf Sci* 518:13–28
19. Siahroudi SK, Moodi PZ, Beigy H (2018) Detection of evolving concepts in non-stationary data streams: a multiple kernel learning approach. *Expert Syst Appl* 91:187–197
20. Amini A, Saboohi H, Herawan T, Wah TY (2016) MuDi-Stream: a multi density clustering algorithm for evolving data stream. *J Netw Comput Appl* 59:370–385
21. Xing L, Demertzis K, Yang J (2020) Identifying data streams anomalies by evolving spiking restricted Boltzmann machines. *Neural Comput Appl* 32(11):6699–6713
22. Dong Y, Japkowicz N (2018) Threaded ensembles of autoencoders for stream learning. *Comput Intell* 34(1):261–281
23. Chauhan S, Vig L, Ahmad S (2019) ECG anomaly class identification using LSTM and error profile modeling. *Comput Biol Med* 109:14–21
24. Shanthamallu US, Spanias A, Tepedelenioglu C, Stanley M (2017) A brief survey of machine learning methods and their sensor and IoT applications. In: *2017 8th international conference on information, intelligence, systems & applications (IISA)*, August. IEEE, pp 1–8
25. Rodriguez MA, Kotagiri R, Buyya R (2018) Detecting performance anomalies in scientific workflows using hierarchical temporal memory. *Futur Gener Comput Syst* 88:624–635
26. Ahmad S, Lavin A, Purdy S, Agha Z (2017) Unsupervised real-time anomaly detection for streaming data. *Neurocomputing* 262:134–147
27. Cauteruccio F, Fortino G, Guerrieri A, Liotta A, Mocanu DC, Perra C, Terracina G, Vega MT (2019) Short-long term anomaly detection in wireless sensor networks based on machine learning and multi-parameterized edit distance. *Inf Fusion* 52:13–30
28. Peng Y, Tan A, Wu J, Bi Y (2019) Hierarchical edge computing: a novel multi-source multi-dimensional data anomaly detection scheme for industrial Internet of Things. *IEEE Access* 7:111257–111270

29. Bose B, Dutta J, Ghosh S, Pramanick P, Roy S (2018) D&RSense: detection of driving patterns and road anomalies. In: 2018 3rd international conference on Internet of Things: smart innovation and usages (IoT-SIU), February. IEEE, pp 1–7
30. Farshchi M, Weber I, Della Corte R, Pecchia A, Cinque M, Schneider JG, Grundy J (2018) Contextual anomaly detection for a critical industrial system based on logs and metrics. In: 2018 14th European dependable computing conference (EDCC), September. IEEE, pp 140–143

A Mini Review on—Physically Unclonable Functions: The Hardware Security Primitives



Harsh Panchal, Naveen Kumar Chaudhary, and Sandeep Munjal

Abstract The Internet of Things (IoTs) is made up of several interconnected, resource-constrained devices, including sensors, actuators, and nodes that are connected to the Internet. These devices often have limited feature size/area and energy resources, making the cost of using traditional cryptography very expensive and infeasible. In recent years, physically unclonable functions (PUFs), a promising hardware security primitive, have emerged. In this paper, we discuss PUF as an alternative to these already available security protocols based on traditional mathematical cryptography. The PUFs are security primitives that address security issues such as IC authentication, Intellectual Property (IP) protection, etc. These PUFs are compatible with well-developed CMOS technology and are built on delay (such as; Arbiter, Loop, etc.) or memory structures (such as; RRAM, SRAM, etc.). IC Authentication Mechanism of PUFs and Types of PUFs have been discussed in brief in this paper. The potential of PUFs in different novel applications has also been discussed.

Keywords Security primitives · Hardware security · Physically unclonable functions · PUFs

1 Introduction

Many of our daily life's processes have been digitized and run over the vast domain of the Internet. The development of Internet of Things (IoTs) is one of the most important developments of the present era [1]. In this era of 5G and cryptocurrency, networks and security have become repositories of digital information [2]. IoT—The new edge of computing where vast amounts of confidential digital information and security tasks are being executed. The IoT devices such as routers and webcams have

H. Panchal · S. Munjal (✉)

National Forensic Sciences University, Goa Campus 403401, India

N. K. Chaudhary

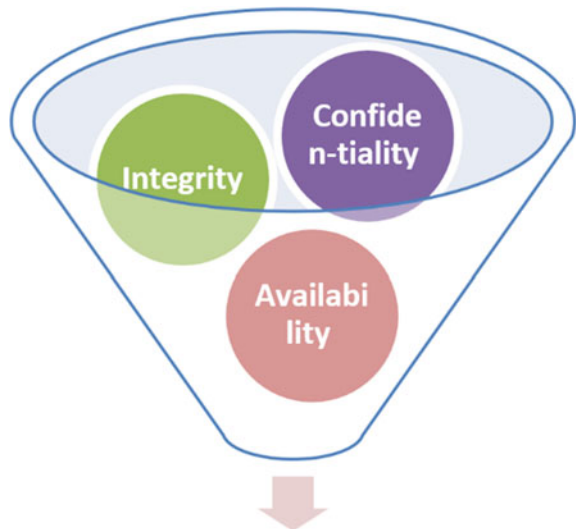
National Forensic Sciences University, Gandhinagar, Gujarat 382007, India

become targets for hackers because of their inherent lack of security, vulnerabilities, etc. [3]. As these IoT devices are mostly connected to the Internet and hence they pose a threat to the user in many ways. They also collect and share different types of data, which may raise critical concerns. By gaining unauthorized access to IoT equipment, an attacker may monitor or control the equipment and try to sabotage it. This may include device hijacking, data siphoning, device theft, access risks, and many more.

A reliable connection between entities, a proper authentication mechanism, and data secrecy are all security objectives for the IoTs [1–3]. As depicted in Fig. 1, this objective leads to the information security trinity, which provides CIA, or Data Confidentiality, Data Integrity, and Data Availability [4, 5]. Any threat in each of these areas has the potential to seriously harm the system and have a direct effect on the functions of the system, which may create critical IT/OT infrastructure damage and bigger financial losses.

The secrecy of the information, message integrity, and device authentication security characteristics are taken into consideration during secure communication. Due to all such threats, the IoT requires a complete set of precautions. One of the greatest methods to equip an IoT device with appropriate protection is to provide protection at the chip level, and this field of information security falls under hardware security [6]. Due to their widespread use and easy access by adversaries, IoT hardware devices are susceptible to physical attacks. IoT hardware has less focus to lead a secure “digital life”. Hardware, which is considered to be the key to information security, can play a crucial role in developing secure communication. Being at the base of our computer systems, it becomes essential to secure it to protect anything that comes above. A small flaw in the hardware can be used to compromise or steal

Fig. 1 CIA security model



everything about that hardware, like the operating system, BIOS, and various applications. Hardware security [7] is a part of a physical computing device that secures and manages the cryptographic key for critical functions such as encryption, authentication, and crypto processing, which deals with security in processors, smartcards, chipsets, flash drives, cryptosystems, etc.

The devices that serve to transmit data between the cloud and local network are sophisticated IoT devices like pieces of hardware such as sensors, actuators, different portable gadgets, or their integrated combinations. These devices are called edge computing devices [8] and they perform some degree of data processing within the device itself with the help of AI and machine learning to bring intelligence systems to do smart things such as autonomous driving, IT/OT infrastructure, medical equipment, and home automation. These edge devices are low-powered devices and have a very small memory footprint [8, 9].

These edge devices are often connected to different (i) sensors—to monitor, collect, or sense the physical properties. The sensors detect changes in measurable conditions and send a signal when one occurs; and (ii) actuators receive signals and act on them. In a mechanical machine, the actions performed by these actuators are often in the form of movement. By manipulating the signal processing within these edge devices, an attacker can control the physical motion of different parts of the connected devices. Hence, the authentication of these devices cannot be ignored and that plays a major role if we talk about critical infrastructure.

1.1 Hardware Security Primitives

IoT edge devices require secret keys in order to protect their data, IP, and operations from possible threats. These devices depend upon software or algorithms for the generation of random keys [10, 11], which works on a deterministic principle and hence cannot generate a genuine random key. Further, these software-based random key generators cannot defend against physical attacks, and because of their complexity, there are always vulnerabilities that hackers can take advantage of. For instance, malicious codes, that can read the keys kept in the software and send them to the hackers, can be introduced. Many IoTs, which are used to monitor patients and collect their daily vital data and send this data to the concerned health professionals to make medical or treatment decisions, are generally lightweight IoTs. Due to their insufficient entropy to generate adequately random keys, these lightweight IoTs are especially at risk.

Most of the cryptographic algorithms that can authenticate these IoT devices are quite complex, and they consume a considerable amount of processing power and memory [10]. Furthermore, the injection of secret keys into chips in this way adds cost and complexity, and limits flexibility. Integral circuits (ICs) and electronic systems' trustworthiness, integrity, and authenticity are significantly enhanced by hardware security primitives [12]. For example, primitives, like true random number generators (TRNGs), generate secret cryptographic keys and IDs that are frequently

used for device authentication, generating session keys, cloning prevention, and many other purposes. These primitives produce device-specific electronic fingerprints and random digital signatures.

By creating the secret keys inside the chips, one can also reduce this complexity and speed up the device authentication process. Recently, a novel method known as Physically Unclonable Functions (PUFs) has been proposed as an interesting cryptographic primitive, which has the potential to take the place of using private keys as a method of device authentication. The brief details of these PUFs are given in the next section.

1.2 Physically Unclonable Functions

The PUFs are devices, which are practically infeasible to clone, duplicate or predict, even if the exact manufacturing process is followed very precisely [13]. Due to this extraordinary property of PUFs, they are referred to as “fingerprints of devices”. Since the edge devices have low power and a smaller memory footprint, a major application for PUFs is for authentication, especially for edge devices since the PUFs are known to work on extremely small power. A PUF is a physical challenge–response system (CRP) that generates a robust, unclonable, and unpredictable response on giving a query or challenge [14]. The PUFs use these challenge–response pairs to authenticate the devices instead of using a private key that is linked to the device identity. In these challenge–response systems, an electrical stimulus of very low power is applied to the device, and this physical stimulus is called the challenge. The reaction to this challenge or stimulus is called a response, which is then recorded.

This response of the device strongly depends upon the physical microstructures of the device, which depend upon several factors or steps followed during the fabrication/manufacturing process. For similar challenge given to different devices of same structure, even manufactured by following the same steps by the same manufacturer gives different responses are obtained, and hence these responses are unpredictable and unclonable. PUFs have already been used for IoT device and IC authentication/identification, keyless secure communication, secure key storage, binding hardware to software platforms, etc.

As discussed above, for the IoTs, authentication is one of the essential security features and PUFs have emerged as promising candidates for the development of low-power authentication protocols for edge devices and IoTs. However, designing the PUF-based IoT authentication protocols has several challenges to face. Many possible architectures to construct secure and reliable device authentication protocols by using noisy-PUFs, ideal-PUFs, and machine learning attacks on PUF are being extensively studied by the researchers [15–17]. Many of the PUFs are vulnerable to machine learning-based modeling attacks; however, strong PUFs like RRAM-based PUFs have shown high reliability against machine learning attack [18–26].

2 Types of Physically Unclonable Functions

Silicon PUFs can be broadly classified into two categories:

- (i) Delay-based PUFs
- (ii) Memory-based PUFs.

Delay-based Physical unclonable function uses small delay variations as a characteristic, whereas the memory-based PUF uses the intrinsic characteristics of individual memory cells. Many of these delay-based and memory-based PUFs are depicted in Fig. 2.

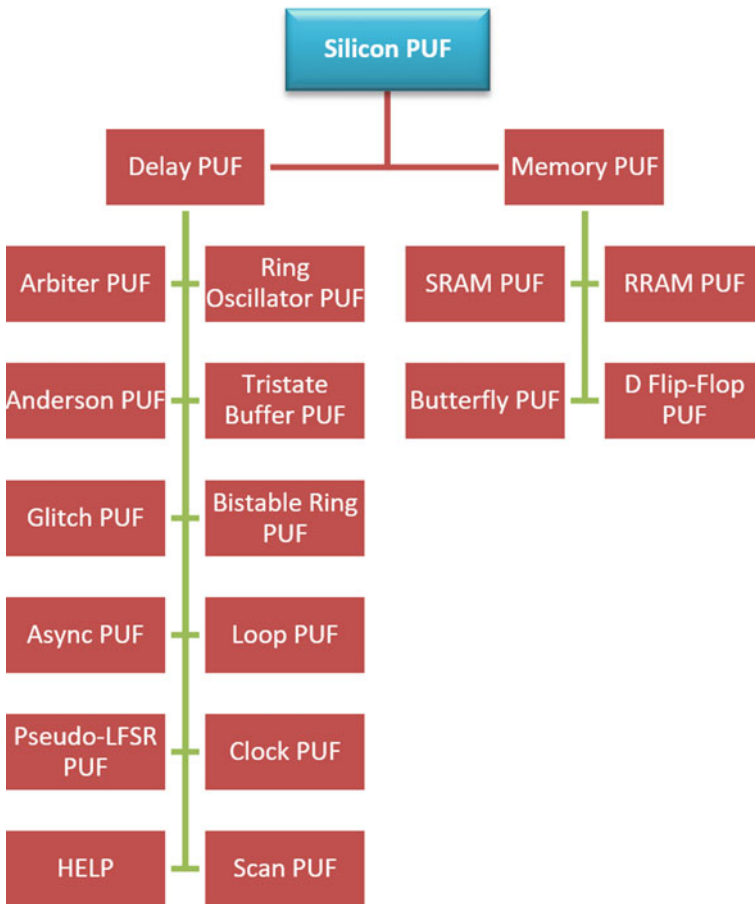


Fig. 2 Silicon-based PUFs

2.1 Delay-Based PUFs

Delay-based PUFs are based on minor changes in electronic delay paths and can be integrated into both Application-Specific Integrated Circuit (ASIC) and Field Programmable Gate Array (FPGA) technology [27]. Among a number of delay-based PUFs (see Fig. 2), the following PUFs are being studied extensively.

- Ring Oscillator PUF
- Arbiter PUF
- Bistable Ring PUF
- Loop PUF.

2.2 Memory-Based PUF

For any non-volatile memory technology, the memory characteristics of the start-up state become device-specific, then such memory devices can be used in PUFs [28]. These device-specific memory characteristics arise from the positive feedback loop, which is used to store bits. For an undetermined memory state at power-up, a perfectly matched loop has equal chances of going to ON state or OFF state and hence of becoming logic 1 or logic 0. However, during the fabrication of the device, some slight variations are introduced unwillingly, which unbalances the feedback loop (which consists of cross-coupled gates) uniquely for each memory cell. This uniqueness of the feedback loops makes the memory characteristics device-specific and hence useful for PUF applications. Some of the memory-based PUFs are

- SRAM PUF
- Flip-Flop PUF
- Butterfly PUF
- RRAM PUF.

A comparison of above-mentioned PUFs is given in the following table (Table 1).

3 IC Authentication Mechanism of PUFs

There are generally two classes of security applications for PUFs—(i) PUF-Based Encryption/Decryption and (ii) PUF-Based Authentication. In this paper, we are particularly interested in PUF-Based device authentication.

The identity of an IoT device must be verified in order to establish trust in the device. A PUF-based secure IoT device authentication method is depicted in Fig. 3. As discussed earlier, a PUF is a physical challenge–response system that, when queried with a challenge x , generates a response y that is robust, unclonable, and

Table 1 Comparison of the existing PUFs

Type of PUF	Name	Weak/strong	Refs.	Comments
Delay-based	Ring oscillator PUF	Weak	[29–31]	Needs large power and space
	Arbiter PUF	Strong	[29, 31]	Vulnerable to attacks
	Bistable ring PUF	Strong	[29]	–
	Loop PUF	Weak	[29, 30]	–
Memory-based	SRAM PUF	Weak	[29–31]	Vulnerable to side-channel attacks
	Flip-Flop PUF	Weak	[29, 30]	–
	Butterfly PUF	Weak	[29–31]	Unstable adjoining affects PUFs response
	RRAM PUF	Strong	[29, 31]	Very sensitive to environmental and voltage fluctuations

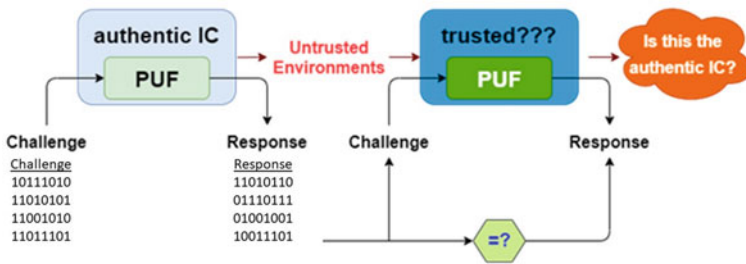


Fig. 3 PUF-based IoT device authentication method

unpredictable. In the typical enrollment process, which is the first step of the authentication process, In the enrollment process, a large number of randomly chosen challenges are presented to the PUFs, and corresponding responses are stored in a secure database for the IC authentication process. After the completion of the enrollment process, the corresponding IoT devices can be deployed.

In verification, which is the second stage of authentication, the verification authority selects a challenge from the pre-recorded sets of CRPs that have never been used. This challenge is presented to the PUF of the IoT device and the returned response is checked against the previously recorded responses in the database. The IoT is considered authentic if the response returned by the PUF of the IoT corresponds to the one that is recorded in the database.

PUF-based identification and authentication are essentially similar in that a server again builds a database of (part of) the CRPs of a PUF. This time, however, the PUF and the associated device go through untrusted environments where counterfeits could be used in their place. So, if a client wishes to make sure that a device is authentic, it can be done by submitting a request to the server for authentication of

the required PUF. The server then sends a request for the PUF to be identified “as a challenge”. The device is authenticated if the generated answer matches the relevant response stored in the server’s database; otherwise, it is not, as shown in Fig. 3.

The authentication in this case is completed without explicitly disclosing any of the stored CRPs. However, a CRP of the verified PUF may once again be entirely revealed in the event of successful authentication, and therefore this CRP should not be used in any future authentication. Such reuse of a revealed CRP can be avoided by using a Hashing scheme. Finally, if the PUF responses are noisy, error correction may once more be done.

4 Key Metrics for PUFs and Applications of PUFs

As a PUF is a physical challenge–response system, a response should, in theory, have a uniform distribution of ON and OFF states, meaning that there should be an equal number of 0 and 1 s in the response bits. This is one of the most important criteria for a challenge–response system to behave like a PUF. Further, a strong PUF has many other varieties of criteria and qualities such as—Uniqueness, Reliability, Randomness, Steadiness, Security, Diffuseness, Bit-aliasing Correctness Uniformity, etc., as shown in Fig. 4. All these criteria are considered to be key metrics for determining the PUF performance [32]. These metrics have different importance in different applications depending upon the set of requirements of the application.

In particular, a PUF must always have Uniqueness, Reliability, and Randomness [32]. These qualities of a PUF system can be stated as

- The capacity of a PUF to produce responses in a distinctive manner is referred to as its uniqueness. To put it another way, PUF uniqueness refers to the fact that many PUFs provide various replies to the same task.
- A PUF’s reliability shows that it can reliably regenerate the same response to a certain challenge.
- Finally, the randomness of a PUF’s responses to different challenges reveals the randomness of the answer bits.

The PUFs have a variety of applications. PUFs can be used in applications that require small amounts of randomness to operate. In applications like random number generators, RFID tags, secret key generation, and device authentication where the necessary randomness property is obtained from process variation, PUFs appear to be a simple yet elegant solution. PUFs have also been used in consumer electronics for low-cost authentication of IoT devices.

The PUFs can be used as Hardware Random Number Generators (HRNGs), which can generate true random numbers, not pseudo-random numbers, which are generally generated using mathematical algorithm-based Pseudo-Random Number Generators (PRNGs) in many cryptographic applications. Many other applications of PUFs are depicted in Fig. 5.



Fig. 4 Key metrics for the determination of PUF performance

5 Conclusion

In summary, we have discussed the importance of information security and the security of edge devices in the current era of the Internet of Things. As these edge devices are low-power architectures with very small feature sizes or areas and small memory footprints, the application of conventional cryptography systems to protect them becomes difficult. To protect their data and operations from potential threats, IoT edge devices typically use algorithm-based software to generate pseudo-random keys that are stored in the system or IC. PUFs have been recently proposed by a number of studies to generate an unclonable and unpredictable cryptographic root key, which is never stored, but rather reconstructed from the device-specific silicon fingerprint every time it is required for authentication. This way, the PUFs alleviate the drawbacks of authentication of the IoT edge devices with the stored keys, which makes the PUF-based hardware security primitives highly secure and inexpensive. As PUFs provide volatile secrets that do not need to be stored anywhere and do not require any special fabrication technology or manufacturing process. Furthermore, different types and applications of PUFs have also been discussed.

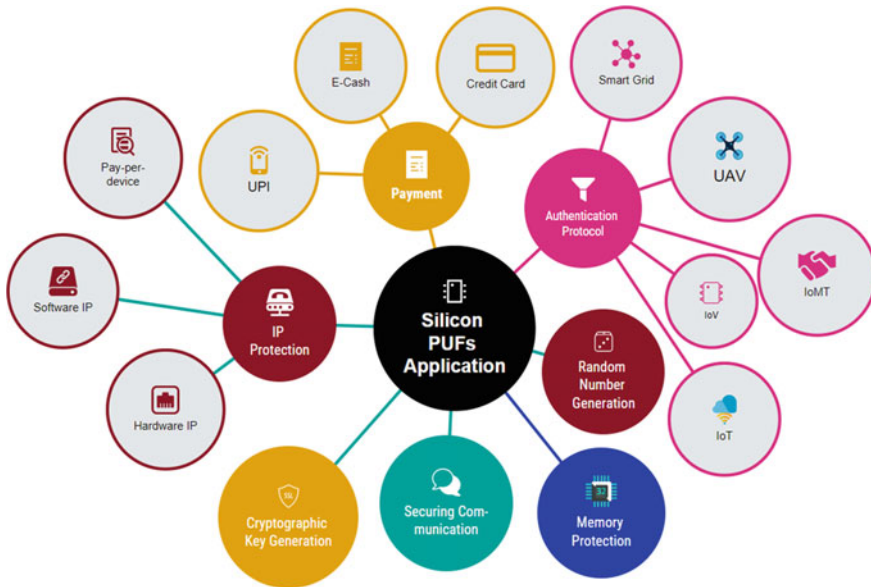


Fig. 5 Applications of physically unclonable functions

References

- Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M (2014) Internet of things for smart cities. *IEEE Internet Things J* 1:22–32. <https://doi.org/10.1109/JIOT.2014.2306328>
- Panharwar MA, Sullemannemmon M, Saddar S, Rajput U, Fellow R (2017) 5G future technology: research challenges for an emerging wireless networks. *IJCSNS Int J Comput Sci Netw Secur* 17
- Airehrour D, Gutierrez J, Ray SK (2016) Secure routing for internet of things: a survey. *J Netw Comput Appl* 66:198–213. <https://doi.org/10.1016/J.JNCA.2016.03.006>
- Kumar M, Meena J, Singh R, Vardhan M (2016) Data outsourcing: a threat to confidentiality, integrity, and availability. In: *Proceedings of the 2015 international conference on green computing and internet of things, ICGCIoT*, pp 1496–1501. <https://doi.org/10.1109/ICGCIOT.2015.7380703>
- Yee CK, Zolkipli MF (2021) Review on confidentiality, integrity and availability in information security. *J ICT Educ* 8:34–42. <https://doi.org/10.37134/JICTIE.VOL8.2.4.2021>
- Lv S, Liu J, Geng Z (2021) Application of memristors in hardware security: a current state-of-the-art technology. *Adv Intell Syst* 3:2000127. <https://doi.org/10.1002/AISY.202000127>
- Hardware is the Key to Information Security. <https://www.secuosys.com/about/stories/hardware-is-the-key-to-information-security>. Accessed 25 Oct 2022
- Mateos Matilla D, Lozano Murciego Á, Jiménez-Bravo DM, Sales Mendes A, Leithardt VRQ (2021) Low-cost edge computing devices and novel user interfaces for monitoring pivot irrigation systems based on Internet of Things and LoRaWAN technologies. *Biosyst Eng* <https://doi.org/10.1016/J.BIOSYSTEMSENG.2021.07.010>
- Mansouri Y, Babar MA (2021) A review of edge computing: features and resource virtualization. *J Parallel Distrib Comput* 150:155–183. <https://doi.org/10.1016/J.JPDC.2020.12.015>
- Sreekumar L, Ramesh P (2016) Selection of an optimum entropy source design for a true random number generator. *Procedia Technol* 25:598–605. <https://doi.org/10.1016/J.PROTCY.2016.08.150>

11. (2017) Random number generators for cryptography. *Circuits Syst Secur Priv* 269–310 <https://doi.org/10.1201/B19499-14>
12. Rose GS (2016) Security meets nanoelectronics for internet of things applications. In: *Proceedings of the ACM international conference of the great lakes symposium on VLSI, GLSVLSI. 18-20-NaN-2016*. pp 181–183. <https://doi.org/10.1145/2902961.2903045>
13. How Physical Unclonable Functions (PUFs) are Creating Trust. <https://www.wevolver.com/article/how-physical-unclonable-functions-pufs-are-creating-trust>. Accessed 25 Oct 2022
14. Gao Y, Li G, Ma H, Al-Sarawi SF, Kavehei O, Abbott D, Ranasinghe DC (2016) Obfuscated challenge-response: a secure lightweight authentication mechanism for PUF-based pervasive devices. In: *2016 IEEE international conference on pervasive computing and communications workshops and other affiliated events. PerCom Work*. <https://doi.org/10.1109/PERCOMW.2016.7457162>
15. Roy S, Das D, Mondal A, Mahalat MH, Roy S, Sen B (2021) PUF based lightweight authentication and key exchange protocol for IoT. In: *Proceedings of 18th international conference on security and cryptography. SECRYPT*. pp 698–703. <https://doi.org/10.5220/0010550906980703>
16. Li Z, Chu Y, Liu X, Zhang Y, Feng J, Xiang X (2021) Physical unclonable function based identity management for IoT with blockchain. *Procedia Comput Sci* 198:454–459. <https://doi.org/10.1016/j.procs.2021.12.269>
17. Kim JH, Jeon S, In JH, Nam S, Jin HM, Han KH, Yang GG, Choi HJ, Kim KM, Shin J, Son SW, Kwon SJ, Kim BH, Kim SO (2022) Nanoscale physical unclonable function labels based on block co-polymer self-assembly. *Nat Electron* 5:433–442. <https://doi.org/10.1038/s41928-022-00788-w>
18. Dai L, Yan Q, Yi S, Liu W, Qian H (2019) A Novel RRAM based PUF for anti-machine learning attack and high reliability. *J Shanghai Jiaotong Univ* 24:101–106. <https://doi.org/10.1007/s12204-019-2043-0>
19. Oh J, Kim S, Choi J, Cha J, Im SG, Jang BC, Choi S (2022) Memristor-based security primitives robust to malicious attacks for highly secure neuromorphic systems. *Adv Intell Syst* 2200177. <https://doi.org/10.1002/aisy.202200177>
20. Kim D, Kim T-H, Choi Y, Lee GH, Lee J, Sun W, Park B-G, Kim H, Shin H (2021) Selected bit-line current PUF: implementation of hardware security primitive based on a memristor crossbar array. *IEEE Access* 9:120901–120910. <https://doi.org/10.1109/ACCESS.2021.3108534>
21. Uddin M, Majumder MB, Rose GS (2017) Robustness analysis of a memristive crossbar PUF against modeling attacks. *IEEE Trans Nanotechnol* 16:396–405. <https://doi.org/10.1109/TNANO.2017.2677882>
22. Santikellur P, Subhra R (2021) Deep Learning for computational problems in hardware security modeling attacks on strong physically
23. Khan MI, Ali S, Al-Tamimi A, Hassan A, Ikram AA, Bermak A (2021) A robust architecture of physical unclonable function based on Memristor crossbar array. *Microelectronics J* 116:105238. <https://doi.org/10.1016/j.mejo.2021.105238>
24. Munjal S, Khare N (2021) Compliance current controlled volatile and nonvolatile memory in Ag/CoFe2O4/Pt resistive switching device. *Nanotechnology* 32:185204. <https://doi.org/10.1088/1361-6528/abdd5f>
25. Munjal S, Khare N (2020) Forming free resistive switching characteristics in Al/NiFe2O4/FTO device. *AIP Conf Proc* 2220:20171. <https://doi.org/10.1063/5.0001806>
26. Pandey V, Adiba A, Ahmad T, Nehla P, Munjal S (2022) Forming-free bipolar resistive switching characteristics in Al/Mn3O4/FTO RRAM device. *J Phys Chem Solid* 165:110689. <https://doi.org/10.1016/j.jpcs.2022.110689>
27. Aghaie A, Moradi A, Tobisch J, Wisiol N (2022) Security analysis of delay-based strong PUFs with multiple delay lines. In: *Proceedings of 2022 IEEE international symposium on hardware oriented security and trust. HOST 2022*. pp 125–128. <https://doi.org/10.1109/HOST54066.2022.9840099>
28. Williams P, Idriss H, Bayoumi M Mc-PUF (2021) Memory-based and machine learning resilient strong PUF for device authentication in internet of things. In: *Proceedings of the 2021 IEEE*

- international conference on cyber security and resilience, CSR. pp 61–65. <https://doi.org/10.1109/CSR51186.2021.9527930>
29. Zhang JL, Qu G, Lv YQ, Zhou Q (2014) A survey on silicon PUFs and recent advances in ring oscillator PUFs. *J Comput Sci Technol* 29:664–678. <https://doi.org/10.1007/s11390-014-1458-1>
 30. Yehoshuva C, Raja Adhithan R, Nalla Anandakumar N (2021) A survey of security attacks on silicon based weak PUF architectures. In: Thampi SM, Wang G, Rawat DB, Ko R, Fan C-I (eds) *Communications in computer and information science*. Springer Singapore, Singapore, pp 107–122. https://doi.org/10.1007/978-981-16-0422-5_8
 31. Shamsoshoara A, Korenda A, Afghah F, Zeadally S (2020) A survey on physical unclonable function (PUF)-based security solutions for internet of things. *Comput Netw* 183:107593. <https://doi.org/10.1016/j.comnet.2020.107593>
 32. Likhithashree R, Kiran D (2020) Area-efficient physically unclonable functions for FPGA using ring oscillator. In: 2nd international conference on innovative mechanisms for industry applications. ICIMIA. pp 403–408. <https://doi.org/10.1109/ICIMIA48430.2020.9074968>

An Intelligent Analysis of Mobile Evidence Using Sentimental Analysis



G. Maria Jones, P. Santhiya, S. Godfrey Winster, and R. Sundar

Abstract Smartphones are compatible and easily accessible compared to computers irrespective of place and time. Smartphones merge with our routine which acts as a medium of communication in several ways such as messaging, voice and video calling, sharing of audio and video contact, and many multimedia contents. We express our emotions through words and these communications can be analyzed by a technique called Natural Language Processing (NLP). Many researchers in the field of text mining and NLP have analyzed the polarity of online product reviews, blogs, social media comments, tweets, and many more. Mobile forensics investigation will be more effective by providing the evidences based on sentimental analysis of the text where the suspected messages contain a negative content of having threats, harassments, text against the community guidelines, and some keywords related to the cases which may be the turning point during the forensics investigation. In this paper, a novel framework called Integrated Mobile Device Forensic Investigation Process (IMDFIP) with inclusion of sentimental analysis techniques in the text which is used to emotionally analyze the texts during the communication process. This framework provides additional information to forensic analysts which will trap the suspect during investigation. The proposed framework of IMDFIP identifies the important terms, emotions, and polarity of the text during the crime. IMDFIP capitulates results compared with the previous models.

Keywords Natural language processing · Mobile forensics investigation · Sentiment analysis · IMDFIP

G. M. Jones (✉) · P. Santhiya
Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India
e-mail: joneofarc26@gmail.com

P. Santhiya
e-mail: santhiya.cse@sathyabama.ac.in

S. G. Winster · R. Sundar
Department of Computer Science and Engineering, SRM Institute of Science and Technology and Panimalar Engineering College, Chennai, India
e-mail: godfrey@live.in

1 Introduction

The ubiquity of smartphones has various purposes among people but it plays a major role in communications across the globe. Communication on a global scale leads to the development of valuable sources of information from SMS, and social media networks like Messenger, Facebook, Whatsapp, and so forth. The main usage of digital forensics is to reconstruct and retrieve the digital data from electronic devices which are utilized for legal proceedings by identifying, analyzing, and capturing the data. The existence of digital information in electronic and digital data represents each and everyone's activity while working, living, and playing in a virtual environment which creates electronic trace in our daily lives. The evolution of cybercrime in today's world is at an unprecedented phase. In recent years, social networking usage by individuals and companies is drastically raising. This is because, there is a rapid growth in the usage of smart devices, internet facilities, storage space, etc. Internet has speed and the ability to transfer the data which is mainly used for communication purpose and also opens the door for criminals to indulge themselves in crimes. In traditional times, usually criminals leave traces of a crime by either fingerprints or physical evidences which required a short period of time to investigate. Since the technology is proliferating rapidly, the cybercrimes are also raising exponentially. Most of the cybercrimes target information about individuals, governments, or corporations. The investigator should answer the six important questions during the investigation, which include Who, How, What, Why, When, and Where [16]. Cell phones and smartphones come under the mobile phone category which are portable devices and are vital for day-to-day activities, so they are vulnerable to criminal activity or also being part of crime scene. A variety of smart devices contain sensitive information of users including their phone call logs, SMS, MMS, electronic mails, photographs, videos, memos, passwords, web history, and credit/debit card numbers. These device holders use the smartphones for communication, to exchange photos, connect to social networks, to write blogs, to record audio and video, etc. Due to technology and transmission, data rate is at its peak, it allows most of the individuals to transfer digital data (e.g., digital video, digital images, etc.). The behavioral pattern hidden in the communication can be extracted and identified by sentimental analysis which could be useful during forensics investigation. Sentimental analysis is the process of analyzing and classifying the text based on emotions. The text information or conversations can be retrieved by forensics tools and ML techniques are used to classify the further processes. The sentimental analysis is widely used in various fields like movie reviews, Twitter analysis, customer reviews, and many more. In this paper, we investigate the mobile forensic messages using sentimental analysis to identify the pattern of text and emotion of the people. The digital information stored in smartphone devices help to address the critical questions during a forensics investigation, revealing with whom a particular person has been in contact with, what they have been communicating about, and where they have been.

1.1 Contribution

The contributions of our work are as follows:

- We study the present state investigation of mobile forensics and its flow process. The detailed analysis of its requirements and challenges in mobile forensics compared with traditional process.
- We present the framework of the proposed methodology that could provide additional evidence of suspect.
- We conduct a mobile forensics examination on volatile memory of two mobile devices such as Samsung A50 and J7. With this effect, we additionally used sentiment analysis to understand the polarity of each message and pattern of communications.
- We list the digital evidence that can be recovered from mobile devices and also keywords evidence tagged from the text messages, Images, documents, and so on.
- We present a performance metrics of the proposed methodology in corporate of digital evidence that could help the investigators and also prove that the proposed work is effective in identifying the pattern and appropriate to be followed.

The rest of the paper is organized in the following manner. Section 2 describes the standards and structured investigation flow process of mobile forensics, Sect. 3 is a literature review given by several researchers, while Sect. 4 provides a proposed generalized flow process of mobile forensics, Sect. 4 novel framework of mobile forensics and detailed methodology used in this paper. In Sect. 5, we discuss the findings of the proposed work, and finally Sect. 6 concludes the paper.

2 Related Work

The various existing works on analyzing digital forensics in mobile environment are presented in this segment. Some of the techniques used in the existing system are discussed in this session: [25] explored Wechat application with some queries that came to forefront during mobile forensics examination and technical methods are proposed for answering the questions. Anglano [3] and Anglano et al. [4] has analyzed the Whatsapp and chatSecure for reconstructing the chats and contacts from smartphone and also decrypted the database of all messages and they concluded that none can reconstruct the data from database by using the SQL cipher deletion technique. Rocha et al. [17] proposed a method of authorship attribution applied to specific problems of social media forensics and also examined supervised learning methods for small samples with a step-by-step explanation. Al Mutawa et al. [2] and Jones et al. [12] has conducted and proposed Behavioral Evidence Analysis (BEA) to analyze the motive of offenders and also analyzed the forensics technique on Facebook, Twitter, and MySpace from three BlackBerrys, iPhones, and Android phones to acquire logical image of smartphones with final conclusion that no traces

could be recovered from BlackBerry devices, while the valuable data from iPhones and Android phones can be recovered and used by forensic examiners. Davies et al. [7] provided a technical method for retrieving data from PlayStation 4 and also identified specific issues when the device is online and offline during investigation process.

Akbal et al. [1] analyzed the BiP Messenger (BM) using forensics methodology for both rooted and unrooted devices which showed the differences for data extraction. Jones and Winstler [11] focused on the problem of evidence gathering by detecting malicious activities. Stüttgen [23] illustrated firmware manipulation techniques and proposed a method for identifying firmware-level threats in memory forensic investigations and also evaluated memory forensic tools within both physical and virtual environments. Jones and Winstler [9] and Saleem et al. [18] evaluated two tools and showed that XRY dominates UFED in many cases in analyzing both performance and relevance. Harichandran et al. [8] conducted survey in cyber forensics to optimize resource allocation and to prioritize problems and possible solutions with more efficiency and finally obtained the results from 99 respondents who gave testimony which will be vital in the future. Koch et al. [15] proposed degree and characteristics of logging based on geo-location. They gathered additional information and stored in geo-reputation for conspicuous locations, since IP addresses are not static, additional information is also stored and this information was used to reconstruct the attack routes, to identify and analyze distributed attacks.

Sharma and Jain [21] provided a various sentimental analysis for analyzing the social media security. They also provided a deeper study of security issues related to social media and with the help of machine learning techniques; the performance measure for sentimental analysis was also measured. Jones et al. [13], Kirange and Deshmukh [14], Santhiya and Chitrakala [19, 20] proposed a framework for predicting the emotions like Joy, Fear, Surprise, Sadness, Disgust, Normal, and mixed from news articles text. Various decision tree algorithms were used to perform the sentimental analysis with the effect of feature selection and ensembles. Finally, the highest accuracy of 87.83% was achieved with C4.5 grafted algorithm. Humaira et al. [5] presented a new framework for investigation, i.e., Online Social Networks (OSN). The existing work failed to have automated or semi-automated forensics investigation process. So, the proposed work incorporates the standard model and new model to address the investigation model for OSN. Deepika et al. [6] designed a forensics tool for analyzing cyberbullying and hate text from Twitter using Apache Spark. Using sentimental analysis, cyberbullying and hate words are detected. Spark API Streaming and MLAPI are used for streaming the social media data and classification of data, respectively. Hudan et al. [22] proposed a sentimental analysis to extract the log messages automatically from forensics timeline. The proposed methodology achieved 99.64% of accuracy and 98.43% of F1-Score. The investigators also recorded other activities within the time period. Jones and Winstler [10] and Weng and Tu [24] analyzed jumplist for digital forensics in the field of financial fraud and also proposed a tool to identify the user behavioral pattern on computer.

3 Proposed Flow Process

We proposed a novel framework for Mobile forensics using NLP and machine learning techniques to analyze the text messages which could be helpful for law enforcement. In this section, we present a generalized workflow process which provides the outline of our work. The proposed framework is called Integrated Mobile Device Forensic Investigation Process (IMDFIP) is an integrated process of both forensics process and machine learning techniques to acquire good results as presented in Fig. 1.

3.1 Seizure

The seizure process is the first step of mobile forensics investigation process. Prior to the examination, the mobile devices are seized from the crime scene. During the seizure, all wireless connections connected to the mobile devices are cut off.

3.2 Preparation

The second stage is the preparation phase where the investigation team is used to prepare for the forensics environment with authentic forensics tools. The seized device should be isolated from the network to prevent from tampering.

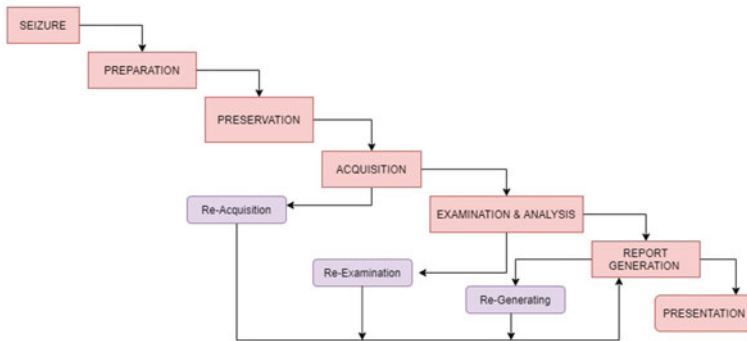


Fig. 1 Generalized flow process of proposed system

3.3 *Preservation*

Based on the previous process, the preservation process takes place. The preservation process is carried out to protect the integrity of digital evidence.

3.4 *Acquisition*

Data acquisition is the process of acquiring data from both volatile and non-volatile memories of suspected mobile phones. There are three types of data acquisition present in mobile forensics.

3.5 *Examination and Analysis*

After the data acquisition process, investigators use various tools for recovering the evidence. Forensics examiners use specialist forensics tools like Encase, Oxygen Forensics, Cellebrite, and so on to recover the data. In this phase, we have additionally included another technique called sentimental analysis for text messages to identify the text pattern which may be helpful for crime investigators.

3.6 *Report Generation and Presentation*

Report generation is the final stage of the proposed workflow process. After gathering all evidence, it should be generated with all necessary documents, expert testimony, clarification, and so on are need to be presented in the court of law.

Three re-consideration processes are further considered for the proposed framework. The re-considered process is examined for data acquisition, examination, and report generation. Sometimes, the initial evaluated report requires revision to collecting more evidence, or the update of the acquired results. However, each forensics toolkit has a different ability of acquiring data. Multiple usages of tools provide the enhanced results. After the evidence preservation phase, the re-acquisition is used to acquire more data. Given that re-acquisition process feeds to re-examination stage which could process with more knowledgeable data. In this scenario, a new evidence pattern is observed with additional key evidence. Despite this, the standard chain of custody and evidence process is required to proceed with the process. In this instance, the new knowledge is observed and whole investigation is required to obtain the knowledge. Finally, the presentation phase is used to present all necessary documents and evidence to court.

4 Proposed Flow Process

Mobile phone developers offer information features and capabilities like personal information storage or management, messaging, audio, video, web browsing, and many more features. These features are varying based on the device and developers, and the modification is updating in each version and also the application installed by users. The following Fig. 2 represents the potential evidence that resides in the mobile phones.

The author has collected almost 7k text messages from several sources like sms, emails, Instagram, and Whatsapp from two mobile devices. There are about 38%, 30%, 20%, and 12% of text content from Whatsapp, emails, messages, and Instagram, respectively. Sentimental analysis is used to detect the polarity (Positive, Negative, and Neutral) as shown in Fig. 3 within the text messages which helps to discover the pattern of the users communication. Based on each sentence, the polarity has determined using libraries and there were about 49% of Negative content, 27% of positive text, and 24% of neutral text are classified as represented in Fig. 4.

4.1 Pre-processing

The cycle engaged with pre-preparing is named as the way toward cleaning crude information prior to utilizing machine learning calculations. Pre-preparing text guarantees the expulsion of loud information from the informational collection. Prior to

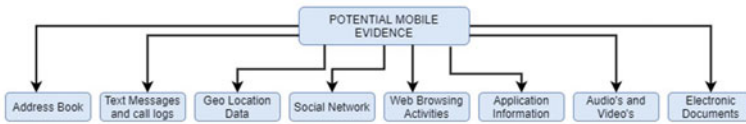


Fig. 2 Potential evidence from mobile devices

Fig. 3 Categories of text sources from mobile devices

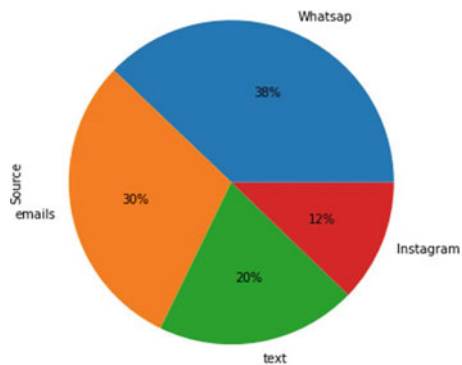
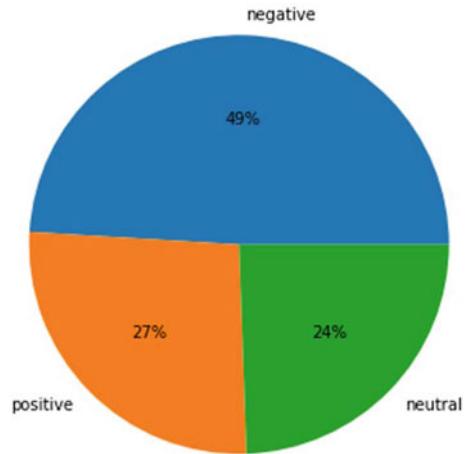


Fig. 4 Categories of polarity from text messages



encoding the content into numeric vector, the pre-handling should be finished by utilizing the following methods: Eliminating URL, converting all capitalized characters to bring down case, eliminating undesirable characters like accentuation, eliminating stop words, stemming and lemmatization, normalization, and a lot more strategies. In this paper, a couple of methods are depicted underneath.

Removing Stop Words The words in the text messages with high frequency or highly used words like if, can, she, them, and may can be removed without changing the sentiment of the sentence. The stop word removal is mainly used to improve the performance of the model. The NLTK library is used to remove the stop words which consists of a bag of words to remove the stop words. **Normalization** The next process is to normalize the text and also stemming and lemmatization are used for further procedure. In order to perform the process efficiently, the words are converted into their root form which helps for better work. For example, the word dance can be dancing, danced, and dancer for stemming process, whereas lemmatization works as removing the inflectional end letters and replaced with the base words. The example of lemmatization is “begging” “beggar” “beginning” will be replaced as “beg” while in lemmatization these words will be replaced with “beg” , “beg” and “begin”.

4.2 Validation and Performance Measure

The author divided the dataset into 80% of training data and 20% of testing data. The aforementioned technique produces performance metrics and confusion matrix in terms of Accuracy, Precision, F1-Score, and Recall. In any classification problem in machine learning, the four combinations of predicted and actual values classification possibilities with respect to suspiciousness act are employed as shown in Table 1, where TP is True Positive and is also known as sensitivity where observation and

Table 1 Confusion matrix

		Actual data	
		Positive (1)	Negative (0)
Predicted data	Positive (1)	True positive (TP)	False positive (FP)
	Negative (0)	False negative (FN)	True negative (TN)

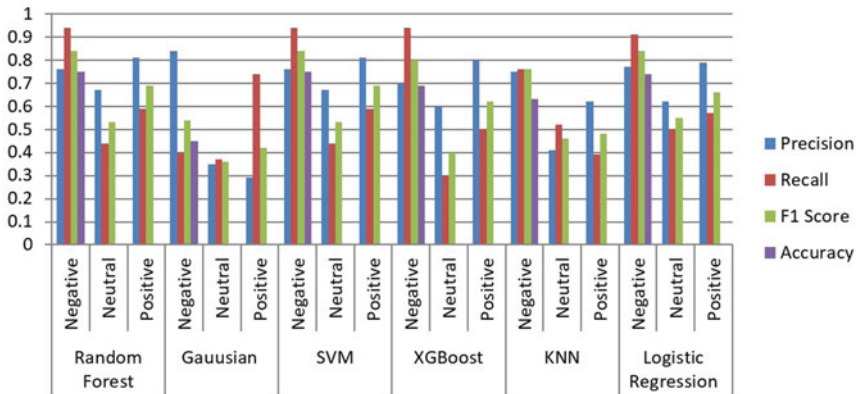


Fig. 5 Graphical representation of evaluation

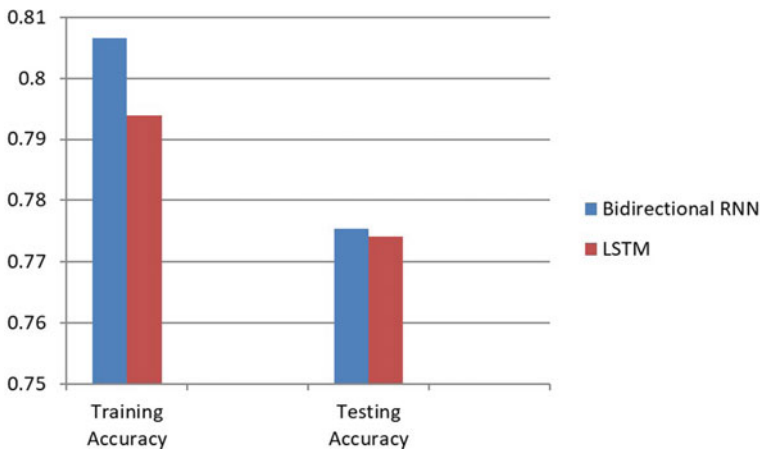


Fig. 6 Graphical illustrations of RNN and LSTM

predicted values are positive, FN is False Negative where observations are positive but predicted values are negative, FP is False Positive where the number of observation is negative but predicted is positive, and TN is True Negative is also known as specificity where observation and predicted values are negative (Figs. 5 and 6).

- Accuracy (A): Accuracy refers to the ratio of correctly classified samples to the total number of samples from the dataset given below:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Precision (P): The ratio of correctly classified samples to the total number of positively predicted/detected samples as given below:

$$P = \frac{TP}{TP + FP} \quad (2)$$

- Recall (R): Recall is also known as sensitivity or true positive rate. It's the ratio of correct positive prediction/detection to the total number of positives as given below:

$$R = \frac{TP}{TP + FN} \quad (3)$$

- F_1Score : Harmonic mean is used to calculate as given below:

$$F_1Score = \frac{2PR}{P + R} \quad (4)$$

4.3 Report Generation and Presentation

After extracting and analyzing the evidence which is collected from mobile devices, the report should be presented to the juries, technical experts, and many more. A detailed summary of all steps involved during the investigation process and also an expert testimony report are presented to the juries by forensics examiners. The expertise and knowledge of the forensic examiner, the methodology adopted, the tools and techniques used, etc., are all likely to be challenged before the jury. Along with the detailed case report, additional documents such as copies of mobile evidence, chain of custody, etc., should also be submitted for strong proof against suspect.

5 Results and Discussion

This work is comprised of three segments and the experimental results of both ML and DL algorithms are evaluated by using metrics which are explained in this session. The first segment is comprised of the use of mobile forensics tool called oxygen forensics software for data acquisition, which acquires the data from two different mobile devices. There are no advanced features or techniques like NPL include grammatical correction, stemming, word cloud, and so on built in this software.

Fig. 7 Keyword evidence of mobile devices

<input checked="" type="checkbox"/> Sources	1
<input checked="" type="checkbox"/> Files	439
<input checked="" type="checkbox"/> Tags	13
<input checked="" type="checkbox"/> Alcohol	3
<input checked="" type="checkbox"/> Chat	76
<input checked="" type="checkbox"/> Child abuse	13
<input checked="" type="checkbox"/> Currency	5
<input checked="" type="checkbox"/> Document	333
<input checked="" type="checkbox"/> Drugs	5
<input checked="" type="checkbox"/> Extremism	8
<input checked="" type="checkbox"/> Gambling	4
<input checked="" type="checkbox"/> Graphic violence	13
<input checked="" type="checkbox"/> ID / Credit card	14
<input checked="" type="checkbox"/> Pornography	2
<input checked="" type="checkbox"/> Vehicles	10
<input checked="" type="checkbox"/> Weapon	1

One main advanced feature is keyword evidence which plays an important role in identifying the criminals which is shown in Fig. 7. The second segment is associated with the use of ML and DL algorithms for NLP. Once the text messages are extracted, the messages assigned with the polarity are either positive, negative, or neutral. Text processing, stemming, and removing stop words are used for better performance and to carry out the process at a higher speed of the sentimental process. The supervised classification algorithms are used for evaluating the performance measure based on training and testing datasets in terms of 80:20 ratios. The following algorithms are used for sentimental analysis: Random Forest, GaussianNB, SVM, XGBoost, KNN, and Logistic Regression. The deep learning algorithms called RNN and LSTM are also used in this work where the process is different from ML algorithms.

The third segment is the presentation of report with all necessary and detailed forensics reports to the juries. The text messages classification was acquired from two mobile devices and categorized the messages based on the polarity which represents the highest acquired messages from Whatsapp and also the polarity of negative text is the highest. This work aims to analyze and examine if machine learning techniques are successfully applied or not to obtain the desired results. This segment compares the performance measures of eight algorithms, where six ML and two DL algorithms are used to identify which algorithm performed well. The performance metrics for six classification algorithms is considered from ML classification algorithms as shown in Fig. 5 and Table 2. The Random Forest and SVM algorithm manifest the highest accuracy of 75% among others. On the other hand, Gaussian produced the lowest accuracy rate of 45% followed by KNN, XGBoost, and Logistic Regression with rate of 63%, 69%, and 74%, respectively. The Logistic Regression algorithm produces 74% of accuracy which is very close to each RF and SVM.

Table 2 Performance measures of various algorithm

Algorithms	Polarity	Precision	Recall	F_1 score	Accuracy
Randon forest	Negative	0.76	0.94	0.84	0.75
	Neutral	0.67	0.44	0.53	
	Positive	0.81	0.59	0.69	
Gaussian NB	Negative	0.84	0.4	0.54	0.45
	Neutral	0.35	0.37	0.36	
	Positive	0.29	0.74	0.42	
SVM	Negative	0.76	0.94	0.84	0.75
	Neutral	0.67	0.44	0.53	
	Positive	0.81	0.59	0.69	
XGBoost	Negative	0.7	0.94	0.8	0.69
	Neutral	0.6	0.3	0.4	
	Positive	0.8	0.5	0.62	
KNN	Negative	0.75	0.76	0.76	0.63
	Neutral	0.41	0.52	0.46	
	Positive	0.62	0.39	0.48	
Logistic regression	Negative	0.77	0.91	0.84	0.74
	Neutral	0.62	0.5	0.55	
	Positive	0.79	0.57	0.66	

Table 3 Performance of RNN and LSTM

Class	RNN	LSTM
Training accuracy	0.8066	0.7939
Testing accuracy	0.7753	0.7642

Finally, the neural network algorithms RNN and LSTM are also used for sentimental analysis. Figure 9 represents the training and validation accuracy of RNN and it shows the validation loss for training. Table 3 represents the performance measures of training and testing accuracy for RNN and LSTM in which RNN performed well for sentimental analysis compared to LSTM. The training accuracy and testing accuracy reached about 80% and 77%, respectively, for RNN, whereas the accuracy for LSTM reached about 79% and 76% for training and testing as shown in Fig. 8. The representation of accuracy for RNN and LSTM is illustrated in Fig. 6.

6 Conclusions

In this work, we presented a novel mobile forensic framework that is proposed to detect additional evidences, which are acquired in text messages from various sources like chats, emails, Whatsapp, and Instagram. Specifically, the ML and DL algorithms

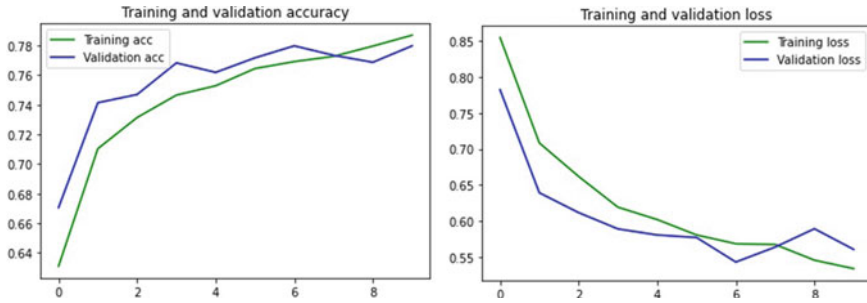


Fig. 8 Training and validation accuracy of RNN

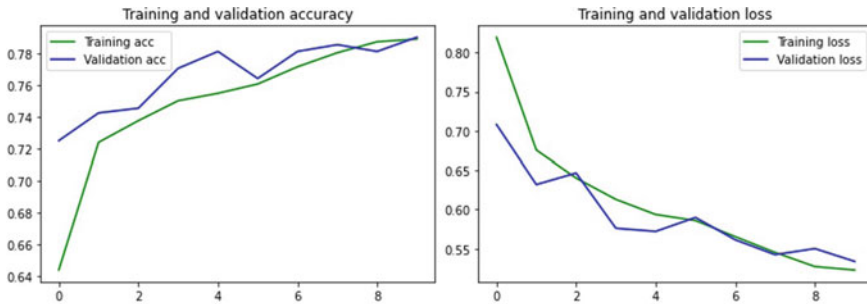


Fig. 9 Training and validation accuracy of LSTM

are used for the detection of patterns in social networks and instant messaging applications that are most commonly used today. For the proposed work, a dataset of 7418 text messages was prepared initially by forensics software and classified all the text messages based on polarity (Positive, Negative, and Neutral). Secondly, supervised machine learning models were built and extensively evaluated with four metrics (Accuracy, Precision, Recall, and F1-score). All the models achieved excellent performance because each class is significantly different from the others. However, the random forest and SVM model were the best performers with an accuracy of 75%. Finally, the two neural networking called RNN and LSTM were used for sentimental analysis. In this case, the RNN algorithm works better than LSTM in terms of training and testing accuracy. Through the IMDFIP framework, we are able to analyze the text messages to identify the keyword evidences. In particular, this framework can be used to search keywords and analyze the sentimental percentage to enrich the forensics investigation by revealing the user pattern of text messages. The paper started with analyzing and reviewing the study about the flow process of existing work, then a brief survey on mobile forensics, machine learning, and a sentimental analysis in the field of cybercrime. In the following, the integrated work for analyzing text message evidences are validated through ML and DL algorithms.

References

1. Akbal E, Baloglu I, Tuncer T, Dogan S (2020) Forensic analysis of BiP Messenger on android smartphones. *Aust J Forensic Sci* 52(5):590–609
2. Al Mutawa N, Baggili I, Marrington A (2012) Forensic analysis of social networking applications on mobile devices. *Digit Investig* 9:S24–S33
3. Anglano C (2014) Forensic analysis of whatsapp messenger on android smartphones. *Digit Investig* 11(3):201–213
4. Anglano C, Canonico M, Guazzone M (2016) Forensic analysis of the chatsecure instant messaging application on android smartphones. *Digit Investig* 19:44–59
5. Arshad H, Omlara E, Abiodun IO, Aminu A (2020) A semi-automated forensic investigation model for online social networks. *Comput Secur* 97:101,946
6. Chaturvedi D, Jangade A, Aietawr H, Pharate A, Shaikh F (2019) Detection of social network based cyber crime forensics using apache spark. *Int J Eng Sci* 21434
7. Davies M, Read H, Xynos K, Sutherland I (2015) Forensic analysis of a sony playstation 4: a first look. *Digit Investig* 12:S81–S89
8. Harichandran VS, Breitinger F, Baggili I, Marrington A (2016) A cyber forensics needs analysis survey: revisiting the domain's needs a decade later. *Comput Secur* 57:1–13
9. Jones GM, Winstler SG (2017) Forensics analysis on smart phones using mobile forensics tools. *Int J Comput Intell Res* 13(8):1859–1869
10. Jones GM, Winstler SG (2021) Analysis of crime report by data analytics using python. In: *Challenges and applications of data analytics in social perspectives*. IGI Global, pp 54–79
11. Jones GM, Winstler SG (2022) An insight into digital forensics: history, frameworks, types and tools. *Cyber Secur Digital Forensic*:105–125
12. Jones GM, Winstler SG, Valarmathie P (2022) An advanced integrated approach in mobile forensic investigation. *Intell Autom Soft Comput* 33(1):87–102
13. Jones GM, Winstler SG, Valarmathie P (2022) Integrated approach to detect cyberbullying text: mobile device forensics data. *Comput Syst Sci Eng* 40(3):963–978
14. Kirange D, Deshmukh R (2012) Emotion classification of news headlines using SVM. *Asian J Comput Sci Inf Technol* 5(2):104–106
15. Koch R, Golling M, Stiemert L, Rodosek GD (2015) Using geolocation for the strategic pre-incident preparation of an it forensics analysis. *IEEE Syst J* 10(4):1338–1349
16. Quick D, Choo KKR (2017) Pervasive social networking forensics: intelligence and evidence from mobile device extracts. *J Netw Comput Appl* 86:24–33
17. Rocha A, Scheirer WJ, Forstall CW, Cavalcante T, Theophilo A, Shen B, Carvalho AR, Stamatatos E (2016) Authorship attribution for social media forensics. *IEEE Trans Inf Forensics Secur* 12(1):5–33
18. Saleem S, Popov O, Baggili I (2016) A method and a case study for the selection of the best available tool for mobile device forensics using decision analysis. *Digit Investig* 16:S55–S64
19. Santhiya P, Chitrakala S (2019) A survey on emotion recognition from EEG signals: approaches, techniques & challenges. In: *2019 International conference on vision towards emerging trends in communication and networking (ViTECoN)*. IEEE, pp 1–6
20. Santhiya P, Chitrakala S (2022) Ptcere: personality-trait mapping using cognitive-based emotion recognition from electroencephalogram signals. *Visual Comput*:1–15
21. Sharma S, Jain A (2020) Role of sentiment analysis in social media security and analytics. *Wiley Interdiscip Rev Data Mining Knowl Discov* 10(5):e1366
22. Studiawan H, Sohel F, Payne C (2020) Sentiment analysis in a forensic timeline with deep learning. *IEEE Access* 8:60664–60675
23. Stüttgen J, Vömel S, Denzel M (2015) Acquisition and analysis of compromised firmware using memory forensics. *Digit Investig* 12:S50–S60
24. Weng SK, Tu JY (2020) A visualization jump lists tool for digital forensics of windows. *KSII Trans Internet Inf Syst (TIIS)* 14(1):221–239
25. Wu S, Zhang Y, Wang X, Xiong X, Du L (2017) Forensic analysis of wechat on android smartphones. *Digit Investig* 21:3–10

Forensics Analysis of TOR Browser



Adarsh Kumar, Kumar Sondarva, Bhavesh N. Gohil, Sankita J. Patel,
Ramya Shah, Sarang Rajvansh, and H. P. Sanghvi

Abstract The Onion Router is a web browser that uses the Tor network to anonymize web traffic by making it simple to conceal one's identity on social media. It uses the onion routing technology to access the multiple-level encrypted, Internet-impossible private mode. These features are being abused to engage in a variety of criminal activities, including cyber terrorism and the black market. The TOR erases all browsing history and other network traces, making it impossible for investigators to gather evidence. This study extracts and examines any potential artifacts that the TOR browser may have produced in local system files and memory dumps.

Keywords Deep web · Anonymous browser · Tor browser · Dark web forensics · The onion routing protocol

A. Kumar (✉) · K. Sondarva · B. N. Gohil · S. J. Patel
Computer Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat
395-007, Gujarat, India
e-mail: adarshraj.ar2000@gmail.com

K. Sondarva
e-mail: kumar.sondarva@gmail.com

B. N. Gohil
e-mail: bng@coed.svnit.ac.in

S. J. Patel
e-mail: sjp@coed.svnit.ac.in

R. Shah · S. Rajvansh
School of Cyber Security and Digital Forensics, National Forensic Sciences University,
Gandhinagar 382007, Gujarat, India
e-mail: ramya.shah@nfsu.ac.in

S. Rajvansh
e-mail: sarang.rajvansh@nfsu.ac.in

H. P. Sanghvi
Directorate of Forensic Science, Gandhinagar, Gujarat, India

1 Introduction

The term “Dark web” refers to the bottom-most and deepest layer of the internet, which is not indexed by search engines and whose contents are not visible to regular browsers and users. People must use specialised browsers like Tor, which offer users anonymity and protect their data, in order to access the Dark web. Criminals undertake a variety of illicit operations on the dark web due to the anonymity provided to users. By using encryption, Tor safeguards user identities and secure networks. Digital marketplace, payment, and community forum users are calling for increased anonymity in the sharing of their online interactions and transactions. Dark wallets and underground networks are the data anonymization platforms addressing these needs. One of these covert networks that was put into place to safeguard users’ identities is Tor. However, this benefit of Tor is frequently exploited to engage in illegal activities including drug trafficking, gambling, the selling of weapons, and violent crimes, among others [1].

2 Tor Relay

The Tor network has three relays: an entry/guard relay, a middle relay, and an exit relay [2]. These relays are also called nodes or routers and allow network traffic to pass through them.

1. **Entry/Guard Relay:** The Tor network is accessible through this relay. The client’s IP address can be read when attempting to connect via the entry relay. The middle node receives the client’s data through the entry relay/guard node.
2. **Middle Relay:** Data in an encrypted format is transmitted using the intermediary relay. It transfers the client’s data to the exit relay after receiving it from the entry relay.
3. **Exit Relay:** The exit relay, which completes the Tor circuit, receives client data from the middle relay and transmits it to the server of the target website. The destination has direct access to the IP address of the exit relay. As a result, in the case that malicious traffic is transmitted, the exit relay is presumed to be at fault because it is thought to be the source of such bad traffic. Therefore, even when it is not the source of bad traffic, the exit relay is most exposed to legal problems, take-down notices, complaints, etc.

3 Working of Tor

The Mozilla Firefox web browser is the foundation of the Tor browser. This browser uses a method known as “onion routing,” in which user data is encrypted using numerous levels analogous to the layers of an onion before being relayed through

various Tor network relays [3]. One layer of the data's multi-layered encryption is decoded at each succeeding relay as user data travels across the network's several relays. The final layer of encryption is removed when the data reaches the exit relay, the last relay in the Tor network, and the data then arrives at the destination server. The final relay in the Tor network, or the exit relay, is what the destination server considers to be the data's source. As a result, it is very challenging to determine the source of data on the Tor network using any kind of surveillance system. As a result, the Tor browser protects and anonymizes user information as well as information about websites and servers. Access to .onion websites on the dark web is made possible through the Tor browser. By using .BIT domains and the hidden service protocol of Tor, users can host websites that are accessible only to other users of the Tor network.

4 Tor Bridge Node

The directory list of Tor relay nodes is accessible to the general public, however, bridge nodes are distinct from relay nodes. Nodes that are not published or included in the public directory of Tor nodes are referred to as bridge nodes. In order to prevent the use of Tor, businesses and governments can block a number of the network's entry and exit nodes because they are listed and easily accessible online. Governments, Internet Service Providers (ISPs), and commercial entities prohibit the usage of the Tor network in many totalitarian nations. Bridge nodes assist in getting around restrictions and enabling users to access the Tor network in situations when the use of the Tor network is banned. Governments, corporations, and ISPs find it challenging to censor the use of the Tor network because of the use of bridge nodes.

4.1 How Bridge Nodes Help Circumvent Restrictions on the Tor Network

The Tor network has bridge nodes that act as proxies, but not all of them are visible in the Tor directory of nodes; some bridge nodes are hidden. ISPs, businesses, and governments are thus unable to identify or ban their IP addresses. Users can simply switch to other bridge nodes even if ISPs and organisations find some of the bridge nodes and censor them. When a Tor user sends traffic to the bridge node, it is sent to the guard node of their choice. Normal communication with a distant server takes place, but a bridge node—an additional transmission node—is also involved. Users can get around the Tor network's restrictions by using hidden bridge nodes as proxies.

5 Related Works

Dingledine et al. [4] detailed the Tor browser's operation, benefits, design aims, and resistance to various threats. Chivers [5] examined whether artifacts can be recovered from browsers and detailed how browsing artifacts are stored in the Windows file system. Goldschlag et al. [6] onion routing, its network architecture, and its fundamental operations were discussed. Huang et al. [7] analysed the traces of the Tor browser and extracts the surfing histories from the memory dump.

6 Determine How to Discover the Tor Browser's Traces

Although the Tor browser offers users anonymity, as long as the system is not shut off, artifacts related to the actions carried out on it remain in the system RAM. The artifacts related to the malicious use of the Tor browser can be found and examined by investigators by obtaining a RAM dump of the active suspect system. Investigators can also find and examine evidence related to the Tor browser that are stored in the Windows Registry and the prefetch folder if the Tor browser was used on a Windows system. How to spot the signs of the Tor browser during a forensic investigation is covered in this section.

6.1 Dark Web Forensics

Investigating illegal and disruptive behaviours that are carried out on the dark web by harmful users is known as "dark web forensics". Drug trafficking, fraud involving people's credit cards, banking information, other financial details, and terrorism are a few examples of illegal and antisocial actions that take place on the dark web. The Tor browser is used to access the dark web, which makes it incredibly difficult for forensic investigators to investigate dark web crimes because it protects user data and keeps users anonymous. Forensic investigators should gather RAM dumps from the suspect machine and analyse them to identify the criminal activities carried out using the Tor browser, such as websites viewed, emails accessed, and programmes downloaded, in order to look into cybercrimes committed using this browser.

6.2 Identifying Tor Browser Artifacts: Command Prompt

- Port 9150/9151 is used by the Tor browser when it is installed on a Windows computer to connect to Tor nodes.
- Investigators can determine whether Tor was used on the machine by checking the machine's active network connections with the command `netstat -ano`.

6.3 Identifying Tor Browser Artifacts: Windows Registry

- When the Tor browser is installed on a Windows PC, the Windows Registry keeps track of user behaviour.
- Forensic investigators can obtain the path from where the TOR browser is executed in the following Registry key: HKEY_USERS\SID\SOFTWARE\Mozilla\Firefox\Launcher

Extract last execution date and time of the Tor browser:

- The investigator examines the “State” file present in the route where the Tor browser was launched on a suspect computer.
- The directory of the State file in the Tor browser folder is \Tor Browser\Browser\TorBrowser\Data\Tor\

6.4 Identifying Tor Browser Artifacts: Prefetch Files

- It will be challenging for investigators to determine whether the Tor browser was used or where it was installed if it is removed from a computer or placed anywhere other than the desktop (in Windows).
- The prefetch files will be examined by investigators to help them get this data.
- The prefetch files are located in the directory, C:\WINDOWS\Prefetch, on a Windows machine.
- Investigators can get information about the browser’s metadata, which includes the following:
 - Browser created timestamps
 - Browser last run timestamps
 - Number of times the browser was executed
 - Tor browser execution directory
 - Filename.

7 Perform Tor Forensics

Depending on the state of the Tor browser on the suspect machine, the procedure and results of executing Tor browser forensics vary. The following scenarios are provided for simplicity in showing how to carry out Tor browser forensics dependent on the condition of the browser [8]:

1. Investigating email artifacts with forensic memory dump analysis while the Tor browser is active

2. Investigating a storage device forensically to obtain email attachments while the Tor browser is active
3. When the Tor browser is closed, forensic examination of memory dumps to look at email artifacts
4. Investigating a storage device forensically to find email attachments after the Tor browser has been closed
5. The ability to conduct forensic investigation after uninstalling the Tor browser

7.1 Tor Browser Forensics: Memory Acquisition

- RAM is where the system's different processes and programmes that are now operating store their volatile data [9].
- Examining RAMdumps can offer in-depth perceptions of the actions that took place on the system [10].
- These dumps can be examined by forensic investigators in an effort to recover various Tor browser artifacts that aid in reconstructing the incident.
- The following factors affect how these artifacts' examination yields different results:
 - Tor Browser Opened
 - Tor Browser Closed
 - Tor Browser Uninstalled
- Memory dump captured while the browser is running captures the highest number of artifacts, whereas a dump captured after uninstalling the browser captures the fewest.
- Memory dumps collected while the browser is closed contain the majority of the information found in memory dumps gathered while the browser is active.

7.2 Collecting Memory Dumps

- To start the forensic investigation, investigators first obtain a memory dump of the suspect computer [11].
- RAM can be captured using tools like FTK Imager and Belkasoft LIVE RAM Capturer.
- The RAM dump taken from the suspect system not only includes browser-related artifacts, but also all of the actions that took place there.

7.3 Memory Dump Analysis: Bulk Extractor

- In order to find artifacts that can be useful during the investigation, the memory dump that was obtained from the machine needs to be reviewed on the forensic workstation.
- Analysing these dumps with tools like Bulk Extractor helps provide useful data including the URLs browsed, email IDs used, and personally identifiable information entered into the websites.

7.4 Forensic Analysis of Memory Dumps to Examine Email Artifacts (Tor Browser Open)

In this case, we'll be performing forensic analysis on the memory dump obtained when the browser is opened in order to look through all the email artifacts and try to piece together the email activities that took place on the machine. Therefore, our attention will be on the email artifacts, such as domain.txt, email.txt, url.txt, url facebook-id.txt, and all of its histogram files, among others. Information that is essential to the browsing activity is contained in the json.txt file.

Step 1: Identify the Domain

- When looking at the artifacts, the first thing to do is look for any domains that are special to email service providers.
- All the domains listed in the dump are listed in the domain.txt file; search the domain entries for domains that are only used for email.

Step 2: Identify the Email IDs

- Since mail.google.com, a domain connected to Gmail, has been discovered, the following step is to look for Gmail IDs saved in the memory dump.
- To identify any recorded Gmail IDs, look through the email.txt file's contents, which contain a list of all the Email IDs found in the dump.

Step 3: Search for Composed Emails

- Start by using the email ID discovered in the preceding step. filtering all the entries in the json.txt file that include the email address
- json.txt is the primary repository for email-related data. content

Step 4: Search for Accessed/Opened Emails

- The memory dump includes both the emails that were composed and those that were open when Tor was last visited.
- These emails begin with the subject of the email, followed by the body, the sender and recipient email addresses, and, if applicable, an attachment.
- Look for emails with the aforementioned pattern in them.

7.5 Forensic Analysis of Storage to Acquire Email Attachments (Tor Browser Open)

So far, we have been successful in locating emails that were written and accessed. We now try to recover the attachments from these emails by forensically inspecting the computer's storage because they contain attachments. Note: Depending on the evidence, the storage medium differs. If the evidence is a physical device, we must obtain a copy of the computer's local storage in bit-by-bit format.

If a virtual machine serves as the proof, we must look at the virtual disc files, which include VMDK, VHDX, and OVF depending on the virtualization programme utilised. In this case, we look at the VMDK file of a virtual machine that was created while the browser was open. The file will be examined using an autopsy.

Step 5: Retrieve the Uploaded File

- Open Autopsy after copying the VMDK file to the forensic workstation.
- Through the emails, a.zip file was uploaded, and a.txt file was downloaded, according to memory dump analysis.
- Search for .zip in the Archives folder under Views→ File Types→ By Extension to find the zip file that was submitted.
- When a file is identified, we have two options: we can extract it to the forensic workstation or we can inspect its contents directly on Autopsy.

Step 6: Retrieve the Downloaded File

- Search for the Secret Credentials.txt file in the Plain Text folder, which can be accessed under Views→ File Types→ Documents, to obtain the.txt file that was downloaded.
- As an alternative, the investigator might search the Web Downloads folder under Results→ Extracted Content for the downloaded file.

7.6 Forensic Analysis of Memory Dumps to Examine Email Artifacts (Tor Browser Closed)

In this instance, all the email artifacts will be examined through forensic analysis of the memory dump that was taken when the browser was closed. Recreate as closely as you can the system's email activity. Therefore, we will concentrate on the email-related artifacts, such as domain.txt, email.txt, url.txt, the histogram files, url facebook-id.txt and all of them Information that is essential to the browsing activity is contained in the json.txt file.

Step 1: Identify the Domain

- When looking at the artifacts, the first thing to do is look for any domains that are special to email service providers.

- All the domains listed in the dump are listed in the domain.txt file; search the domain entries for domains that are only used for email.

Step 2: Identify the Email IDs

- Since mail.google.com, a domain connected to Gmail, has been discovered, the following step is to look for Gmail IDs saved in the memory dump.
- All of the Email IDs listed in the email.txt file are dump, look over its records to see if any Gmail IDs were noted.

Step 3: Search for Composed Emails

- Start by using the email ID discovered in the preceding step. filtering all the entries in the json.txt file that include the email address
- json.txt is the primary repository for email-related data. content

Step 4: Search for Accessed/Opened Emails

- The memory dump that was taken when the browser was closed may or may not contain the artifacts associated to accessed or opened emails. The outcome typically varies, however, it is possible to recover at least some artifacts.

7.7 Forensic Analysis of Storage to Acquire Email Attachments (Tor Browser Closed)

Step 5: Retrieve the Uploaded File

- Open Autopsy after copying the VMDK file to the forensic workstation.
- Through the emails, a.zip file was uploaded, and a.txt file was downloaded, according to memory dump analysis.
- Search for .zip in the Archives folder under Views→ File Types→ By Extension to find the zip file that was submitted.
- When a file is identified, we have two options: we can extract it to the forensic workstation or we can inspect its contents directly on Autopsy.

Step 6: Retrieve the Downloaded File

- The RAM dump does not contain information on emails that have been accessed or opened.
- However, the Online Downloads folder found under Results→ Extracted content can be used to obtain the downloaded data (including email attachments) from the web browser.

Note :Retrieve the file's metadata, such as its location and filesystem type, by looking at the File Metadata section.

7.8 Forensic Analysis: Tor Browser Uninstalled

- When a memory dump is taken from a computer where the Tor browser was installed and later uninstalled, no record of the actions taken on the computer while using the Tor browser is included in the memory dump.
- As a result, while utilising the Bulk Extractor tool to examine the memory dump, the investigator is unable to find any traces pertaining to Tor browser activity.
- In this situation, the investigator must examine prefetch files to find evidence of the machine's use of the Tor browser.

8 Conclusion

The Tor browser's anonymity feature enables criminals to conceal their identities while engaging in illegal activities including the sale of drugs and firearms. The use of memory forensics to detect the Tor browser's existence on a machine and track down the perpetrator's activity is also covered in this study. The correlation of various data from many sources, in addition to memory dump analysis, can aid in the identification of suspects. Cybercriminals are prone to errors like utilising the same email address or browser. For instance, the previous usage of an email that results in an IP address and identifies the suspect could provide useful information.

References

1. Sajjan PP, Balan C, Priya MJD, Sreedeeep AL (2021) Tor browser forensics. *Turkish J Comput Math Educ* 12(11):5599–5608
2. Mulr M, Lelmich P, Buchanan WJ (2019) A forensic audit of the tor browser bundle. *Digital Investig*, Research Gate
3. Darcie W, Boggs RJ, Sammons J, Fenger T (2014) Online anonymity: forensic analysis of the tor browser bundle. Technical Report, pp 1–34
4. Dingedine R, Mathewson N, Syverson P (2004) Tor: the second-generation onion router. Technical report, Naval Research Lab, Washington, DC
5. Chivers H (2014) Private browsing, a window of forensic opportunity. *Digital Invest* 11:20–29
6. Goldschlag DM, Reed MG, Syverson PF (1996) Hiding routing information. In: *International workshop on information hiding*, Springer, pp 137–150
7. Huang MJC, Wan YL et al (2018) Tor browser forensics in exploring invisible evidence. In: *2018 IEEE international conference on systems, man, and cybernetics*
8. Arshad MR, Hussain M, Tahir H, Qadir S, Memon FIA, Javed Y (2021) Forensic analysis of tor browser on windows 10 and android 10 operating systems
9. Ghafarian A, Seno S (2015) Analysis of privacy of private browsing mode through memory forensics. *Int J Comput Appl* 132

10. Dave R, Mistry NR, Dahiya MS (2014) Volatile memory based forensic artifacts and analysis. *Int J Res Appl Sci Eng Technol* 2(1):120–124
11. Dayalamurthy D (2013) Forensic memory dump analysis and recovery of the artifacts of using tor bundle browser the need

Phishing Classification Based on Text Content of an Email Body Using Transformers



M. Somesha and Alwyn R. Pais

Abstract Phishing attacks steal sensitive credentials using different techniques, tools, and some sophisticated methods. The techniques include content injection, information re-routing, social engineering, server hacking, social networking, SMS and WhatsApp mobile applications. To overcome such attacks and minimize risks of such attacks, many phishing detection and avoidance techniques were introduced. Among various techniques, deep learning algorithms achieved the efficient results. In the proposed work, a transformers-based technique is used to classify phishing emails. The proposed method outperformed the other similar mechanisms for the classification of phishing emails. The phishing classification accuracy achieved by the proposed work is 99.51% using open-source datasets. The proposed model is also used to learn and validate the correctness of the in-house created datasets. The obtained results with in-house datasets are equally competitive.

Keywords Email phishing · Transformers · BERT · Text classification

1 Introduction

Phishing is a fraudulent activity performed by an individual fraudster or a group to obtain the sensitive personal credentials of a user from unknown suspicious users through the use of social engineering techniques. Fraudulent emails are designed to accept as originating from a trustworthy organization or a known individual. Such emails frequently broadcast and create attention to persuade recipients to select a web

M. Somesha (✉) · A. R. Pais

Department of Computer Science and Engineering, Information Security Research Lab, National Institute of Technology, Surathkal 575025, Karnataka, India
e-mail: somesha.187co004@nitk.edu.in

A. R. Pais
e-mail: alwyn@nitk.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
S. J. Patel et al. (eds.), *Information Security, Privacy and Digital Forensics*,
Lecture Notes in Electrical Engineering 1075,
https://doi.org/10.1007/978-981-99-5091-1_25

343

link that will redirect them to a fake site that looks genuine¹. The user may be asked to provide sensitive credentials, such as account number, surname, phone number, and password, which could harm the financial and personal status. Furthermore, these spurious websites may incorporate malicious code.

To overcome phishing attacks, the user should have awareness about the anti-phishing tools. There are many anti-phishing tools developed by researchers to create awareness about phishing such as BitDefender, URLcheck, PhishDetector, AntiPhishing, and so on [1]. Due to the lack of education and usage of mobile applications, still users are trapped by phishers and lose their money.

According to APWG 2020-21 statistics, the most phishing emails were detected in the fourth quarter of 2020 as shown in Fig. 1. APWG-2022's first quarter report [2] recorded 1,025,968 phishing attacks. This is the highest phishing attack of more than one million, never observed before by APWG. The financial sectors were targeted majorly in this quarter which is 23.6% of all attacks, and also there were 53,638 unique phishing email subject attacks detected. The data showed a significant increase in phishing scams and their impact, especially during the COVID-19 pandemic [3].

The researchers have proposed email phishing detection algorithms based on supervised learning and unsupervised learning. These algorithms use different Machine Learning (ML) and Deep Learning (DL) algorithms for the classification. The existing algorithms make use of hybrid features (Body and Header) for the classification [4–16]. Some works have used email header features only for the detection of phishing email [17]. Some researchers [18–22] have used only the body part of the email for the phishing email detection. In the current work, we are proposing a novel technique based on transformers algorithm that uses only the email body text content.

The following are the contributions of this work:

- This paper describes a novel deep learning technique that uses transformers to identify phishing emails using only the email body text as input.
- The model performance is evaluated on open-source and in-house generated datasets.
- Finally, a comparison study is performed on our proposed model and other existing works.

The remainder of the paper is organized as follows. Section 2 discusses related works published by other researchers, and Sect. 3 discusses the proposed model's architecture overview. Section 4 discusses the basic experimental setup, Sect. 5 analyzes the experimental results, and Sect. 6 concludes.

¹ <https://www.cisa.gov/uscert/report-phishing>.

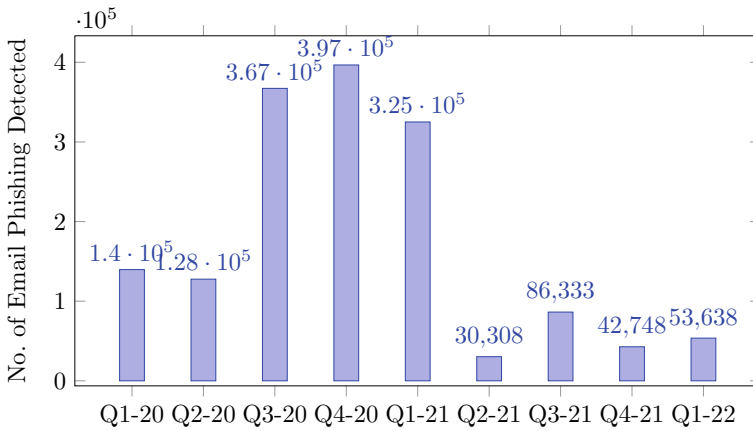


Fig. 1 APWG 2020-21 phishing email statistics

2 Related Works

According to Catal et al. [23], very limited number of papers published so far use body text and attachments of an email for classification of phishing emails. The objective of the current research is to work only on the body text content of an email. As of now, very few researchers contributed using only email body text [18–22], and most of the works are carried out using hybrid features. Existing works used ML, DL, and Natural Language Processing (NLP) techniques and algorithms to identify or classify phishing emails. In our previous work [17], we conducted literature reviews of the majority of the existing works. A brief discussion of some of the existing works which use only email body text is discussed in this section, and a summary of the related works is tabulated in Table 1.

Alhogail et al. [18] proposed a model that uses Graph Convolution Networks (GCN) and NLP over an email body text and achieved an accuracy of 98.2% with a false positive rate of 0.015. As is mentioned, GCN works well for text classification; the obtained result improvement is very nominal by using email body text. Castillo et al. [20] proposed email threat detection using distinct neural networks with email body text and word embedding and ML techniques. From their experimental analysis, the backpropagation method performed the best by achieving 95.68% accuracy to detect malicious emails. The obtained accuracy of their work is less compared to other existing works.

Bountakas et al. [19] proposed comparison models to classify phishing emails using NLP and machine learning algorithms. The models are developed using three NLP algorithms with the Chi-Square feature selection technique and five machine learning classifiers. According to the author, Word2Vec with Random Forest (RF) model achieves the best with an accuracy of 98.95% with a balanced dataset. An additional feature selection technique is used to achieve the best accuracy.

Table 1 Related work summary based on email body text content

Author(s)	Techniques	Datasets	Precision	Recall	F-score	Accuracy
Alhogail et al. [18]	GCN and NLP	CLAIR-ACL	0.985	0.983	0.985	0.982
Castillo et al. [20]	Word embedding and ML	Enron, APWG, and Private	–	–	–	0.9568
Bountakas et al. [19]	Word2Vec-RF	Enron and Nazario	0.9863	0.9931	0.9897	0.9895
Bountakas et al. [19]	BERT-LR	Enron and Nazario	0.85	0.8409	0.8454	0.8449
Hiransha et al. [21]	Word embedding and CNN	IWSPA-AP 2018	–	–	–	0.942
Ramanathan et al. [22]	phishGILLNET-NLP and ML	SpamAssassin, Phishing Corpus, Enron	0.997	0.997	0.997	0.977

Hiransha et al. [21] proposed deep learning-based phishing email detection and used Keras word embedding and Convolution Neural Networks (CNN) to build the model. The author used two datasets with header and without header (Body text content) of the IWASP-AP shared task committee. The model performed better with a hybrid dataset than the dataset with only the body of an email, the achieved accuracy with the only body of an email is 94.2% which is less than other works. Ramanathan et al. [22] proposed phishGILLNET, a multi-layered phishing email attach detection technique using NLP and ML algorithms. The author used three-stage processing of email text in a layered approach using different algorithms in each layer to efficiently classify the given text. The authors used 47 complex features that consume more memory and computation time and achieves an accuracy of 97.7%.

To summarize, the related works tabulated in Table 1 are executed using email body text content as an input parameter. The existing works used different algorithms, techniques, and datasets. The common parameter in all the related works is input email body text content. And also the performance of the existing works is measured using precision, recall, F-score, and accuracy. From the above works, it is observed that Bountakas et al. [19] obtained fairly good accuracy of 98.95%.

3 Proposed Work

The architecture and the functional parameters of the proposed work are discussed in this section. The architecture is designed to satisfy the objective of the proposed work as shown in Fig. 2. The objective of the proposed work is to classify the email as phishing or ham using only the text part of an email body. The proposed architecture

3.2 In-House Dataset Preparation

In-house datasets are prepared by analyzing the behavior of the recently collected emails, source code of the original emails, Google warning indicators, and using MxToolbox² online tool [17]. The prepared in-house Dataset-III and its size are tabulated in Table 2.

3.3 Open-Source Dataset Collection

The open-source datasets are collected from Nazario³ for phishing and ham data from SpamAssassin⁴ repositories. The Nazario dataset has emails from the years 2004 to 2017, and SpamAssassin datasets are collected for the period 2002 to 2004. These open-source datasets may not align with the timeframe in which they were collected, potentially rendering them susceptible to being learned by phishers. The open-source datasets are named Dataset-I and tabulated in Table 2. Dataset-II has legitimate emails from in-house corpus captured during the period 2004 to 2017 and phishing emails from Nazario's phishing corpus. The Dataset-II is prepared to overcome the problem of period mismatch.

3.4 Data Pre-processing

Initially, the input data should be pre-processed using Python scripts. Python scripts are written to process emails collected from open-source and in-house repositories as input. The developed Python scripts extract the body text of an email from MBOX files and remove unwanted tags, junk, and special characters. Cleaning the processed data involves removing inflectional endings and special characters from the email body text.

3.5 Training and Classification Using Transformers

A transformer is a deep learning model used majorly in the fields of NLP and computer vision. The transformer is introduced by Vaswani et al. [24] and designed to process sequential data for translation and text summarization by using an attention mechanism. Among many transformer models, Bidirectional Encoder Representations from Transformers (BERT) is one of the popular and efficient language trans-

² <https://mxtoolbox.com/Public/Tools/>.

³ <https://monkey.org/~jose/phishing/>.

⁴ <https://spamassassin.apache.org/old/publiccorpus/>.

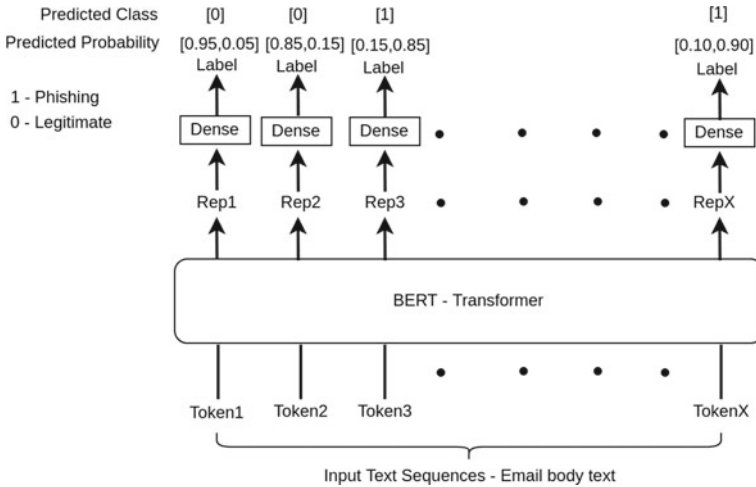


Fig. 3 BERT—transformer architecture

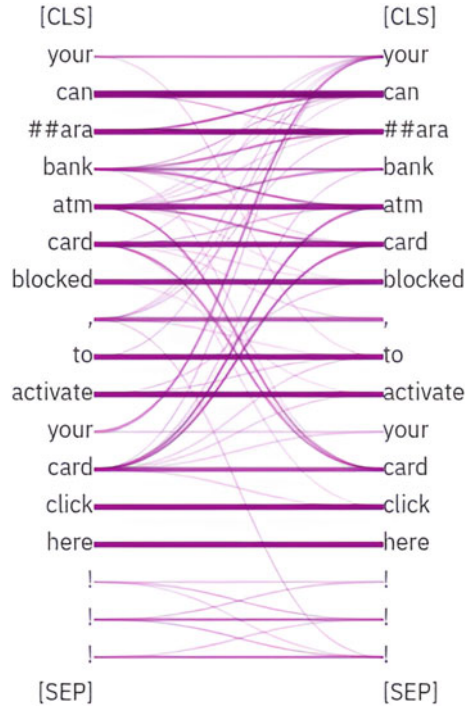
formation models proposed by Google and published by Devlin et al. [25] and his colleagues⁵ as a new language representation model.

In this work, we used the BERT model, a popular NLP model to classify emails by training only body text which is in the form of a natural language. BERT uses bidirectional context word learning in left to right and right to left contexts of email content. To do language processing and classification, we use a bert-base-uncased pre-trained model of Huggingface’s library called transformers to train and classify the given email body text as shown in Fig. 3. The model includes 12 transformer layers, 768 hidden sizes, 12 self-attention heads, and 110 million parameters in total. BERT has token classification by fine-tuning the model. It can be applied to tasks other than natural language processing. The model has two pre-trained tasks, Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). MLM model randomly masks a given sentence to 15% of the words in the input. The model predicts the masked words from the entire input masked sentences. During prediction, the NSP model concatenates two masked sentences as inputs. The model must then predict whether or not the two sentences followed each other.

During pre-processing, the texts are lowercase to vocabulary size. The masking procedures are followed for each sentence with a rate of 15% of the masked tokens, [MASK] replaces 80% of the masked tokens, random token replaces 10% of masked tokens, remaining 10% of masked tokens are left as it is. Thus, the embedding has special tokens called [CLS] at the beginning of each sentence, the token [SEP] to separate two sentences in a sequence and at the end of the sentence, and [MASK] to

⁵ <https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html>.

Fig. 4 BERT-base-uncased—example



mask any word in the sentence. An overview of the BERT model for a classification task is shown in Fig. 4. In the classification stage, the model classifies the given email as phishing or ham and also outputs the prediction accuracy.

4 Experimental Evaluation

The basic experimental setup, resources, and evaluation metrics that will be used to evaluate the proposed model are described in this section.

4.1 Experimental Resources and Datasets

In the proposed work, we used email as a primary resource. The required emails were obtained from two sources: one from a predefined open-source corpus and the other from an in-house generated corpus. Section 3 discusses dataset preparation procedures, and Table 2 lists prepared datasets with corpus sizes. Dataset-I contains 3976 legitimate and 4216 phishing emails from the open-source corpus, Dataset-II

Table 2 Used datasets

Dataset	Ham emails	Phish emails	Total
Dataset-I	3976	4216	8192
Dataset-II	12288	9391	21679
Dataset-III	12288	10639	22927

contains 12288 legitimate and 9391 phishing emails, and Dataset-III contains 12288 legitimate and 10639 open-source phishing emails.

4.2 Evaluation Metrics

The classification of phishing emails with high accuracy does not serve the purpose; instead, it should not classify a legitimate email as phishing. We use some of the basic measuring metrics to evaluate the results; those are True Positive Rate (TPR), True Negative Rate (TNR), Accuracy (Acc), Precision (P), F-score (F), and Matthews Correlation Coefficient (MCC). The confusion matrix is the one which is used to measure the performance of the machine learning and deep learning models.

•

$$TPR = \frac{TP}{(TP + FN)} * 100 \quad (1)$$

•

$$TNR = \frac{TN}{(TN + FP)} * 100 \quad (2)$$

•

$$Acc = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100 \quad (3)$$

•

$$P = \frac{TP}{(TP + FP)} * 100 \quad (4)$$

•

$$F = 2 * \frac{P * TPR}{P + TPR} \quad (5)$$

•

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TN + FN)(TN + FP)(TP + FN)(TP + FP)}} \quad (6)$$

All the above evaluation metrics are used in the proposed work and experiments are conducted using transformers, a deep learning model.

5 Experimental Results and Discussion

To assess the proposed model, we ran three different experiments with three different datasets, which are listed in Table 2. Table 3 summarizes the findings of all three experiments. The basic experimental setup required to carry out these experiments is provided below.

5.1 Basic Experimental Setup

To begin with the basic experimental setup, the programming language used is *Python*, and the libraries used are *Pandas*, *NumPy*, *Seaborn*, *NLTK*, and *Ktrans*. The tool used to develop the model is Jupyter Notebook, and the operating system used is Ubuntu-18.04.6 LTS. The hyper parameters used to develop the proposed model are as follows: the MODEL_NAME used is bert-base-uncased, MAXLEN size is set to 128, the batch size used is 32, learning rate is set to 5e-5, and the number of epochs used is 10. The data used for training is 75% of the total dataset size and testing of 25% of the dataset size for all three datasets and the SEED used is 2020 random state.

5.2 Results and Discussion

In this section, the experimental procedures are discussed. The model uses randomly selected data for training and testing with a ratio of 75:25% of the total dataset size. After removing unwanted characters and symbols, the data is fed to the BERT transformer model. The BERT base model has 12 transformers layers, 768 hidden sizes, and 12 self-attention heads. The transformer learns and selects parameters from the input data. For Dataset-I and II, the TFBertMainLayer is set to 109482240 parameters associated with dropout_37 and dense classifier. The model's total parameters and trainable parameters for Dataset-I are set to 109484547. The size and contents

Table 3 Model performance with all three datasets

Dataset	Learning rate = 5e-5 (0.0337)				
	Training accuracy	Training loss	Validation accuracy	Validation loss	Training time (s)
Dataset-I	0.9995	0.0023	0.9951	0.0251	951
Dataset-II	0.9903	0.0253	0.9856	0.0609	2504
Dataset-III	0.9902	0.0260	0.9897	0.0254	2699

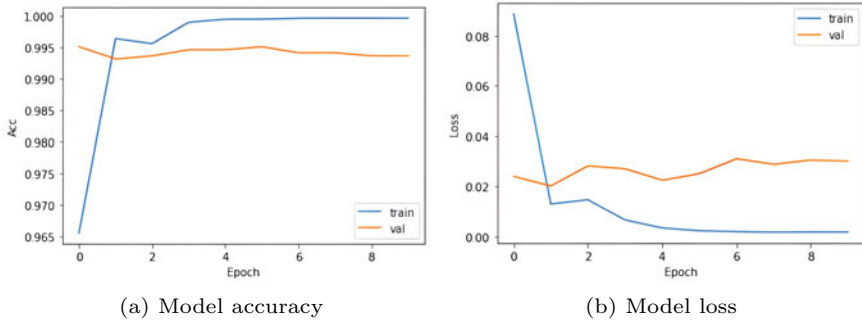


Fig. 5 Accuracy and loss charts for Dataset-I

of Dataset-III vary, the model parameters and layers also vary with respect to the complexity of the data.

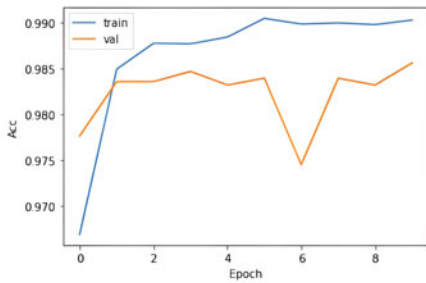
The objective of the proposed model is to classify the given input as either positive or negative (0 or 1) to indicate phishing or legitimate email. The model performance with Dataset-I, Dataset-II, and Dataset-III are tabulated in Table 3. The model performance with Dataset-I is 99.95% training accuracy, 99.51% validation accuracy, 0.0023 training loss, 0.0251 validation loss, and time taken to build the model is 951 s. The validation accuracy and validation loss graphs are shown in Fig. 5. The results of all evaluation metrics for the Dataset-I are tabulated in Table 4. According to Table 4, the precision is 99.20%, recall is 99.80%, and F-score is of 99.50%. The obtained results with open-source datasets are competitive and outperformed all other existing works.

Using in-house data Dataset-II, the model performance is analyzed with a training accuracy of 99.03%, validation accuracy of 98.56%, training loss of 0.0253, and validation loss of 0.254, and time taken to train and validate the model is 2504 s seconds. Also, all metrics are measured to evaluate the model, and the results of all metrics are tabulated in Table 4. According to the results, the obtained precision is 99.74%, recall is 97.76%, and F-score is of 98.74%. The model accuracy charts for the given input are shown in Fig. 6. The results of the proposed model with Dataset-II prove that the prepared dataset is appropriate and suites best for the phishing classification.

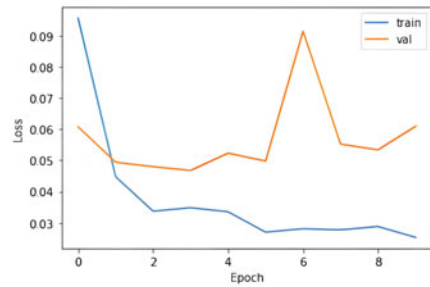
Dataset-III is a combination of in-house legitimate and Nazario’s phishing emails. The model performed equally well with the selected data and achieved a training accuracy of 99.02%, validation accuracy of 98.97%, training loss of 0.0260, and validation loss of 0.0254. The accuracy and loss charts are shown in Fig. 7, and performance metrics are tabulated in Table 4.

Table 4 Obtained results using transformers with all three datasets

Measure	Dataset-I	Dataset-II	Dataset-III
Sensitivity/recall	0.9980	0.9776	0.9813
Specificity	0.9925	0.9965	1.0000
Precision	0.9920	0.9974	1.0000
Negative prediction value	0.9981	0.9702	0.9777
False positive rate	0.0075	0.0035	0.0000
False discovery rate	0.0080	0.0026	0.0000
False negative rate	0.0020	0.0224	0.0187
Accuracy	0.9951	0.9856	0.9897
F-score	0.9950	0.9874	0.9905
MCC	0.9902	0.9709	0.9795

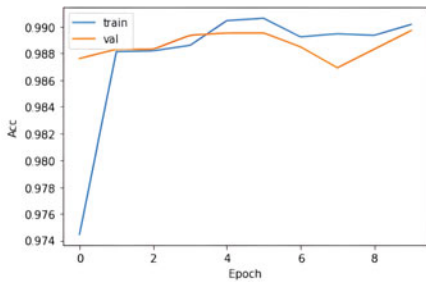


(a) Model accuracy

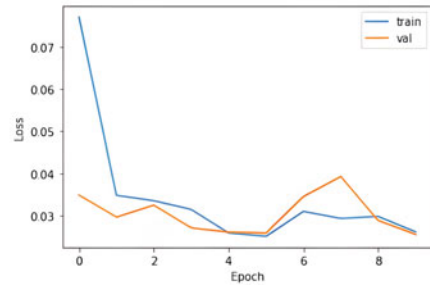


(b) Model loss

Fig. 6 Accuracy and loss charts for Dataset-II



(a) Model accuracy



(b) Model loss

Fig. 7 Accuracy and loss charts for Dataset-III

Table 5 Comparison study

Author(s)	Datasets	Precision	Recall	F-score	Accuracy
Alhogail et al. [18]	CLAIR-ACL	0.985	0.983	0.985	0.982
Castillo et al. [20]	Enron, APWG, and Private	–	–	–	0.9568
Bountakas et al. [19]	Enron and Nazario	0.9863	0.9931	0.9897	0.9895
Bountakas et al. [19]	Enron and Nazario	0.85	0.8409	0.8454	0.8449
Hiransha et al. [21]	IWSPA-AP 2018	–	–	–	0.942
Ramanathan et al. [22]	SpamAssassin, Nazario's Phishing Corpus and Enron	0.997	0.997	0.997	0.977
Proposed work	Dataset-I	0.9920	0.9980	0.9950	0.9951
	Dataset-II	0.9974	0.9776	0.9874	0.9856
	Dataset-III	1.0	0.9813	0.9905	0.9897

5.3 Comparison Study

In Table 5, we listed some recent related works which use only email body text contents as input to their models. Majority works used commonly available open-source datasets and achieved fair results. In the proposed work, Dataset-I consists of Nazario's phishing corpus and SpamAssassin ham datasets. The same dataset is used by Ramanathan et al. [22]. Using Dataset-I, the proposed model outperformed all other existing works with an accuracy of 99.51% at 1st and 6th epoch out of 10 epochs. The model also achieved the best precision, recall, and F-score of 99.20%, 99.80%, and 99.50%, respectively. The proposed model also performed well with Dataset-II and Dataset-III with an accuracy of 98.56% and 98.97%, respectively. From the experimental results, it is evident that the proposed model outperformed the other existing techniques.

6 Conclusion and Future Work

In this paper, we present a novel phishing email classification model based on BERT transformers. We also built an internal email dataset and validated it with our proposed model. For the open-source data, the proposed model with Dataset-I achieved the highest accuracy of 99.51%. Furthermore, the proposed work outperformed all other existing works using only the email body text feature for the identification or detection of phishing emails.

This work can be extended using minimum features of both email header and body text for the classification. And also, we may extend this work using different and advanced transformers.

References

1. Sharma H, Meenakshi E, Bhatia SK (2017) A comparative analysis and awareness survey of phishing detection tools. In: 2017 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT). IEEE, pp 1437–1442
2. APWG (2022) APWG 2022 phishing activity trends reports, first quarter 2022. https://docs.apwg.org/reports/apwg_trends_report_q1_2022.pdf. Accessed 07 June 2022
3. Salloum S, Gaber T, Vadera S, Sharan K (2022) A systematic literature review on phishing email detection using natural language processing techniques. IEEE Access
4. Hamid IRA, Abawajy J (2011) Hybrid feature selection for phishing email detection. In: International conference on algorithms and architectures for parallel processing. Springer, pp 266–275
5. Abu-Nimeh S, Nappa D, Wang X, Nair S (2009) Distributed phishing detection by applying variable selection using Bayesian additive regression trees. In: 2009 IEEE international conference on communications. IEEE, pp 1–5
6. Bagui S, Nandi D, Bagui S, White RJ (2019) Classifying phishing email using machine learning and deep learning. In: 2019 International conference on cyber security and protection of digital services (cyber security). IEEE, pp 1–2
7. Gansterer WN, Pölz D (2009) E-mail classification for phishing defense. In: European conference on information retrieval. Springer, pp 449–460
8. Harikrishnan N, Vinayakumar R, Soman K (2018) A machine learning approach towards phishing email detection. In: Proceedings of the anti-phishing pilot at ACM international workshop on security and privacy analytics (IWSPA AP) 2013, pp 455–468
9. Islam R, Abawajy J (2013) A multi-tier phishing detection and filtering approach. J Netw Comput Appl 36(1):324–335
10. Khonji M, Iraqi Y, Jones A (2012) Enhancing phishing e-mail classifiers: a lexical url analysis approach. Int J Inf Secur Res (IJISR) 2(1/2):40
11. Ma L, Ofoghi B, Watters P, Brown S (2009) Detecting phishing emails using hybrid features. In: 2009 Symposia and workshops on ubiquitous, autonomic and trusted computing. IEEE, pp 493–497
12. Nguyen M, Nguyen T, Nguyen TH (2018) A deep learning model with hierarchical lstms and supervised attention for anti-phishing. Preprint at [arXiv:1805.01554](https://arxiv.org/abs/1805.01554)
13. Ra V, HBa BG, Ma AK, KPa S, Poornachandran P, Verma A (2018) Deepanti-phishnet: applying deep neural networks for phishing email detection. In: Proceedings of the 1st AntiPhishing shared pilot 4th ACM international workshop security privacy analysis (IWSPA). Tempe, AZ, USA, pp 1–11
14. Smadi S, Aslam N, Zhang L (2018) Detection of online phishing email using dynamic evolving neural network based on reinforcement learning. Decis Support Syst 107:88–102
15. Toolan F, Carthy J (2009) Phishing detection using classifier ensembles. In: 2009 eCrime researchers summit. IEEE, pp 1–9
16. Valecha R, Mandaokar P, Rao HR (2021) Phishing email detection using persuasion cues. IEEE Trans Depend Secure Comput
17. Somesha M, Pais AR (2022) Classification of phishing email using word embedding and machine learning techniques. J Cyber Secur Mobil:279–320
18. Alhogail A, Alsabih A (2021) Applying machine learning and natural language processing to detect phishing email. Comput Secur 110:102414

19. Bountakas P, Koutroumpouchos K, Xenakis C (2021) A comparison of natural language processing and machine learning methods for phishing email detection. In: The 16th International conference on availability, reliability and security, pp 1–12
20. Castillo E, Dhaduvai S, Liu P, Thakur KS, Dalton A, Strzalkowski T (2020) Email threat detection using distinct neural network approaches. In: Proceedings for the first international workshop on social threats in online conversations: understanding and management, pp 48–55
21. Hiransha M, Unnithan NA, Vinayakumar R, Soman K, Verma A (2018) Deep learning based phishing e-mail detection. In: Proceedings of the 1st AntiPhishing shared pilot 4th ACM international workshop security privacy analysis (IWSPA). Tempe, AZ, USA
22. Ramanathan V, Wechsler H (2012) phishgillnet-phishing detection methodology using probabilistic latent semantic analysis, adaboost, and co-training. *EURASIP J Inf Secur* 2012(1):1–22
23. Catal C, Giray G, Tekinerdogan B, Kumar S, Shukla S (2022) Applications of deep learning for phishing detection: a systematic literature review. *Knowl Inf Syst*:1–44
24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. *Adv Neural Inf Process Syst*:30
25. Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding. Preprint at [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)

Vehicle Theft Detection and Tracking Using Surveillance Video for the Modern Traffic Security Management System



Charanarur Panem, Ashish Kamboj, Naveen Kumar Chaudhary,
and Lokesh Chouhan

Abstract The present paper is to demonstrate the identification of automobiles using an image, camera, or video clip by utilizing Python OpenCV. It is necessary to first download and then install OpenCV. The Python programming language is used for the development of the present system. This paper, focused on scenario analysis to detect and track the vehicles. Detailed instructions on how to do an analysis of video sequences obtained from an optical sensor in the paper on monitoring road sections were provided. These kinds of algorithms are able to identify road markers, count cars, and assess information about traffic flow. The proposed algorithm for vehicle recognition is built on top of an integrated platform of smart cameras, which is also utilized to test and validate the algorithm. The effectiveness of the algorithms and software has been shown via experimental testing. The findings demonstrate that the suggested algorithms make it possible to solve the problem in question in real-time and in a variety of observation settings, as was anticipated.

Keywords Traffic surveillance · Smart cameras · Video analysis · Image processing · Object detection · Background subtraction · Line detection

1 Introduction

The design and execution of several tasks in the area of transportation analytics are directly tied to the rising economic and social expenses associated with the ever-increasing number of autos on the road. These responsibilities consist of, but are not limited to, addressing the following: Poor traffic safety measures and traffic

C. Panem

School of Cyber Security and Digital Forensic, National Forensic Sciences University, Tripura Campus, Agarthala, Tripura, India

A. Kamboj (✉) · N. K. Chaudhary · L. Chouhan

School of Cyber Security and Digital Forensic, National Forensic Sciences University, Goa Campus, Ponda, Goa, India

e-mail: ashishkamboj1000@gmail.com

congestion are two of the most pressing issues facing today’s most populous cities across the globe. While there is clearly an increasing need for financial resources to meet these difficulties, the quantity now at hand is woefully inadequate.

Adaptive traffic control systems might be installed on city roadways as one solution. Delays and congestion have been reduced and other issues have been addressed thanks to traffic control technology.

Detection of accidents and cars parked in the improper spot. Keeping an eye out for compliance and making a note of any traffic violations.

Another crucial component is the collecting of statistics on traffic. This study aims to provide solutions to problems that arise in traffic management systems, such as identifying and counting the number of vehicles that are in motion. A technique for recognizing road markings is also provided in this article. These problems with a stationary video camera that is situated high above the street are being examined along with potential solutions. We took care to factor in the necessity for deployment on an embedded platform made up of smart video cameras while we designed these algorithms.

Figure 1 shows a schematic of the technology used to detect and tally the number of vehicles on the road-1.

The quantity of data that is sent to the traffic control center may be reduced thanks to the online video processing that has been included in the system. As a direct result of this, the processing power and bandwidth requirements for the server will be significantly reduced. However, it is important to remember that these cameras have limited processing power, and that advanced image processing needs access to powerful computers in order to account for a broad range of aspects. However, it is important to remember that these cameras do have a certain practical use.

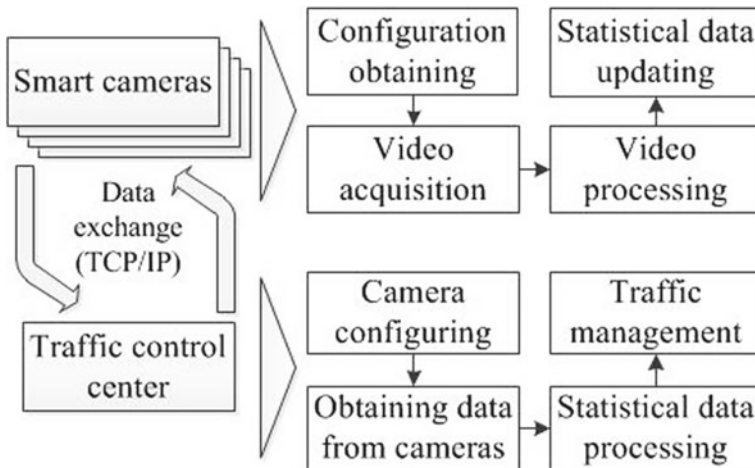


Fig. 1 Vehicle detection and counting system

The darkness of a scene may be affected by the shadows that are thrown by things like clouds, trees, and buildings.

The path that the sun takes across the sky produces ever-changing shadows and illuminates the ground below in a way that is mesmerizing.

- Light from incoming cars and streetlights that reflect off the asphalt surface of the road.

How the weather is now being (Rain, Snow, Fog).

- Multiple viewing points and views captured by the camera.
- Road segment designs and intersection arrangements that deviate from the standard.

2 Related Work

Various research initiatives are now being carried out with the purpose of improving techniques for detecting things based on the motion in which they are moving. In the field of video analytics, the usage of such approaches has reached an optimal point at which it may be implemented. There are two basic focuses of investigation in the academic world [1].

To begin, research efforts are concentrated on improving approaches for estimating optical flow in order to provide more accurate estimates of optical flow and cut down on the amount of processing that is necessary [2, 3].

The second step is to investigate and improve different methods of modeling the backdrop [4–7]. The objective of several ongoing investigations in this area is to develop more accurate statistical models of the brightness distribution of the background. In this section, we also explore several methods for predicting the time-variable changes in the parameters of these models. Background modeling as a result of matrix decomposition [8] and background modeling as a result of tensor decomposition [9] has been the focus of a number of research in the recent past.

Machine learning techniques may be used in many fields. Neural networks, decision trees, cascade classifiers, and other approaches are only some of the many that have been developed to handle these mathematical difficulties. In the field of machine learning, the availability of a wide variety of mathematical methods, the possibility of combining a number of different mathematical models, and the development of new heuristic processes all contribute to the development of one-of-a-kind discoveries in this area of research. It is now impossible to single out even a single piece, much less a number of works, that are of great quality. Although a great deal of work has been achieved, the produced algorithms are not yet widely used. Furthermore, the extent to which a training set is exhaustive and how well it adheres to typical operating parameters is a crucial factor in the success of the algorithms.

It should be brought to everyone's attention that the effectiveness of the industrial solutions that are now available for recognizing and counting moving items is

inadequate. Because the primary emphasis of academic research is on the development of theoretical methods, it may be challenging to determine whether or not these methods are effective across the full range of real-world contexts. In academic research, it is common practice to ignore the computational complexity of algorithms. This indicates that a significant number of the proposed approaches have a level of computational complexity that prevents them from being realistically applicable in the absence of specialized hardware.

3 Vehicle Detection and Counting Algorithm

Giving local transportation officials a reliable estimate of the number of vehicles using the city's roads, highways, and streets might help them better allocate funds and enhance the level of service they provide. Road traffic analysis systems have the ability to aid in both the real-time problem-solving (for adaptive traffic management) and the statistical study of traffic flow. These technologies have the potential to be useful. We are in a position to make recommendations regarding the implementation of practical solutions to improve the flow of passenger and cargo traffic as well as the safety of traffic as a result of our capability to process and analyze statistics. This ability is made possible by the fact that we are able to process and analyze statistics. The installation of traffic signals and the establishment of rules for their use, the alteration of the configuration of a road junction, and the building of extra lanes are all examples of potential solutions [10].

The vehicle counter will be incremented by one if the sensor determines that any of the following circumstances have been met.

There was an object in the region just before the threshold.

The object was seen in the area just before exiting the zone.

The object has now moved beyond the point of no return.

In the first stages of the algorithm, most of the focus is placed on doing background modeling in each area. The system has at this point determined which frames do not include any motion. To do this, a rough estimate of the total number of active locations in each frame is first calculated:

Figure out the difference between one frame and the next.

The thresholding approach is used to find out how many discrete events have occurred in total.

The area is considered to have motion when the number of moving points in the region is larger than a threshold, which changes based on the picture's quality. It depends on the answer to this question whether or not motion sensors can pick up activity there. No innovation can take place unless there is some kind of change or transition.

When a car departs the entrance zone, it is also recorded as having left the exit zone, since the two are time-synchronized. This must be done to prevent selecting a reference frame based on the time a vehicle came to a stop inside the zone. In order to avoid any misunderstandings, this is essential. A reference frame will be chosen for

use in the background estimate if there is no motion in the region during the period of the allotted time.

Following the selection of a reference frame, an analysis of the background stability over a predetermined period of time is performed (usually several seconds). In order to do this, we examine the ways in which the current frame deviates from the reference frame.

As soon as the phase of estimating the background is over, the zone will transition into its normal mode of operation, which will consist of the following processes:

Calculating the amount of time that has elapsed between background frames F and B by using the following steps [10]:

$$d_{x,y} = \begin{cases} F_{x,y} - B_{x,y}, & \text{if } F_{x,y} \geq B_{x,y}, \\ B_{x,y} - F_{x,y}, & \text{Otherwise} \end{cases} \tag{1}$$

where x, y —coordinates of the pixel.

Using thresholding to classify a set of dots as either an object, background, shadow, or excessive illumination. (For general consumption and for use on public transportation) and a few more.

For the created algorithm to successfully recognize and count vehicles, it is necessary to define unique zones of interest for the sensors in the picture. Each lane of the road receives a sensor. The sensors in each device are split into two halves. By specifying the distance between the zones, we can calculate an approximation of the passing vehicle’s speed and use this information to predict the direction of its progress.

$$b_{x,y} = \begin{cases} 1, & \text{if } d_{x,y} \geq t_{bin}, \\ 0, & \text{Otherwise} \end{cases} \tag{2}$$

$$l_{x,y} = \begin{cases} 1, & \text{if } F_{x,y} \geq B_{x,y}d_{x,y} \text{ and } b_{x,y} \equiv 1, \quad F_{x,y} > t_{light}, \\ 0, & \text{Otherwise} \end{cases} \tag{3}$$

$$s_{x,y} = \begin{cases} 1, & \text{if } F_{x,y} < d_{x,y} \text{ and } b_{x,y} \equiv 1, \quad F_{x,y} < t_{shadow}, \\ 0, & \text{Otherwise} \end{cases} \tag{4}$$

where “ $b_{x,y}$ ” indicates that the pixel is not part of the background, “ $l_{x,y}$ ” suggests that there is too much light, and “ $s_{x,y}$ ” indicates that there is too much darkness.

Whether a spot falls under the shadow (t_{shadow}) or is bathed in too much light, the defining thresholds are laid forth (t_{light}). These points must be filtered out to prevent false positives. The zone enters the “vehicle detected” stage when points associated with an item take up a significant portion of the zone’s available space. As was previously said, the choice to detect a vehicle is ultimately made at the sensor level.

4 RMD Algorithm

Pixels that have the coordinates $(b_{x,y})$ but are not in the background; pixels that have the coordinates $(l_{x,y})$ but have an excessive amount of light; and pixels that have the coordinates $(s_{x,y})$ but not enough amount of dark.

The lines that demarcate what it means to be in the shade (tshadow) and what it means to be dazzled by the sun are drawn (tlight). It is crucial to eliminate certain data points so as to prevent producing false positives in the analysis. After a certain number of points associated with an item have occupied a significant portion of the zone, the “vehicle detected” phase will start. As was said before, the sensors are the ones that make the choice as to whether or not a vehicle will be detected by them [11–15].

$$T(s, \alpha) = \int_{\sigma=s_{\min}}^s S(\sigma, \alpha) d\sigma, \quad s \in [s_{\min}, s_{\max}] \quad (5)$$

where $S(\sigma, \alpha)$ —VRT result; (s, α) —line parameters.

The following limitations, on the other hand, reduce its usefulness in dealing with the problem as it is described.

The Interactive Voice Responses Technology (I.V.R.T) is unable to establish the positions of the endpoints of the marking lines; it can only detect the straight paths along which the marking lines are positioned.

The I.V.R.T. is unable to recognize curving marking lines as separate entities. Certain video cameras generate footage that is considerably warped, which may dramatically alter marking lines that are normally straight. This may be a problem when trying to create accurate maps.

In order to get around these restrictions, it is recommended that the original picture be cut up into smaller pieces (blocks). In this scenario, we ignore the possibility that lines might be curved, and as a result, it is frequently difficult to precisely locate the ends of segments within these blocks.

The line detections made by the IVRT on a block-by-block basis need to be pieced together to create continuous curved or straight lines that may be read as road markings [16]. Each newly identified segment is connected to a proactive agent that is responsible for the ongoing search for connections with other agents in the area. Following a number of iterations of agent interaction, a demand-resource network (DR-network) of sufficient stability is formed. The structure of the network itself incorporates several lines of investigation [17].

5 Experimental Results

In order to test the efficacy of the created algorithms and software, researchers analyzed real-world footage captured by embedded smart cameras while monitoring various stretches of road in real-time.

Video was taken during various times of the day and night, and in a variety of climates. Images after processing are 1024 pixels wide by 768 pixels high, and the data set used for experimental study is the ImageNet classification task at half the resolution (224×224 input image) and then double the resolution for detection [10].

A. The Conclusions Reached by the Road Marking Detection Algorithm

The vehicle recognition and counting system's operational settings can only be set with the help of the road marker detection algorithm. Thus, the operator of the traffic control center needs the outcomes of this algorithm. Because of this, all experiments were conducted on a computer equipped with a 3.40 GHz Intel (R) Core (TM) i7-3770 CPU. Eleven videos worth of stills were used. Location and lighting conditions affect what is captured in photographs.

The findings of experimental study conducted on various photographs. According to these findings, the average detection accuracy of road markings is around 76%. Fig. 2 shows the road marking detection vehicles.

B. Results for Vehicle Detection and Counting Algorithm

The created vehicle recognition and counting algorithm will be optimized for use on an embedded platform of smart cameras. The AXIS M2026-LE Mk II network camera, which uses an Ambarella ARM CPU, was used to evaluate the algorithm. With five lanes of traffic under observation, it took 4 m/s to analyze a single picture.

Vehicle identification and counting algorithm process is shown in Fig. 3. Multiple camera angles were used to study three- and four-lane stretches of road.

Table 1 presents the findings of the various empirical investigations. Carried out on the algorithm for vehicle detection and counting for video sequences that were taken in a number of different observational settings.

5.1 Conclusions and Future Scope

The present paper's main contribution are Hybrid smart technology, Simple to deploy on all devices and systems, Algorithm improvements were also made to improve the accuracy and speed of searching, and detailed instructions on how to do an analysis of video sequences obtained from an optical sensor in the course of monitoring road sections were provided. These kinds of algorithms are able to recognize road markers, count cars, and analyze information about traffic flow. The proposed algorithm for



Fig. 2 Road marking detection

vehicle recognition is built on top of an integrated platform of smart cameras, which is also utilized to test and validate the algorithm.

The effectiveness of the algorithms and software has been shown via experimental testing. The findings demonstrate that the suggested algorithms make it possible to solve the problem in question in real-time and in a variety of observation settings, as was anticipated. The scope of the paper is to help traffic police and Maintaining records. Traffic surveillance control modern traffic management system.



Fig. 3 Process of vehicle detection

Table 1 Results of vehicle detection

Test sequence	Results		
	Vehicle count	Detection accuracy (%)	False alarm rate (%)
1	800	99.69	0.61
2	777	99.29	0.59
3	999	100	0.73
4	5467	98.52	0.65
5	999	96.96	1.35

References

1. Pandu Ranga HT, Ravi Kiran M, Raja Shekar S, Naveen kumar SK (2010) Vehicle detection and classification based on morphological technique. In: 2010 International conference on signal and image processing. pp 45–48. <https://doi.org/10.1109/ICSIP.2010.5697439>
2. Muslu G, Bolat B (2019) Nighttime vehicle tail light detection with rule based image processing. *Sci Meet Electr-Electron Biomed Eng Comput Sci (EBBT)* 2019:1–4. <https://doi.org/10.1109/EBBT.2019.8741541>
3. Mittal U, Potnuru R, Chawla P (2020) Vehicle detection and classification using improved faster region based convolution neural network. In: 2020 8th International conference on reliability, infocom technologies and optimization (trends and future directions) (ICRITO). pp 511–514. <https://doi.org/10.1109/ICRITO48877.2020.9197805>
4. Kul S, Eken S, Sayar A (2017) A concise review on vehicle detection and classification. *Int Conf Eng Technol (ICET)* 2017:1–4. <https://doi.org/10.1109/ICEngTechnol.2017.8308199>
5. Tan Q, Wang J, Aldred DA (2008) Road vehicle detection and classification from very-high-resolution color digital orthoimagery based on object-oriented method. In: *IGARSS 2008—IEEE international geoscience and remote sensing symposium*. pp IV-475–IV-478. <https://doi.org/10.1109/IGARSS.2008.4779761>
6. Momin BF, Mujawar TM (2015) Vehicle detection and attribute based search of vehicles in video surveillance system. In: 2015 International conference on circuits, power and computing technologies [ICCPCT-2015]. pp 1–4. <https://doi.org/10.1109/ICCPCT.2015.7159405>
7. George J, Mary L, Riyas KS (2013) Vehicle detection and classification from acoustic signal using ANN and KNN. In: 2013 International conference on control communication and computing (ICCC). pp 436–439. <https://doi.org/10.1109/ICCC.2013.6731694>
8. Baek JW, Lee E, Park M-R, Seo D-W (2015) Mono-camera based side vehicle detection for blind spot detection systems. In: 2015 Seventh international conference on ubiquitous and future networks. pp 147–149. <https://doi.org/10.1109/ICUFN.2015.7182522>
9. Chandrika RR, Ganesh NSG, Mummoorthy A, Raghunath KMK (2019) Vehicle detection and classification using image processing. In: 2019 International conference on emerging trends in science and engineering (ICESE). pp 1–6. <https://doi.org/10.1109/ICESE46178.2019.9194678>
10. Chen T, Chen Z, Shi Q, Huang X (2015) Road marking detection and classification using machine learning algorithms. In: 2015 IEEE intelligent vehicles symposium. pp 617–621
11. Dong Q, Zou Q (2017) Visual UAV detection method with online feature classification. In: 2017 IEEE 2nd information technology, networking, electronic and automation control conference (ITNEC). pp 429–432. <https://doi.org/10.1109/ITNEC.2017.8284767>
12. Seenouvong N, Watchareeruetai U, Nuthong C, Khongsomboon K, Ohnishi N (2016) Vehicle detection and classification system based on virtual detection zone. In: 2016 13th International joint conference on computer science and software engineering (JCSSE). pp 1–5. <https://doi.org/10.1109/JCSSE.2016.7748886>
13. Pandya HA, Bhatt MS (2015) A novel approach for vehicle detection and classification. In: 2015 International conference on computer communication and informatics (ICCCI). pp 1–5. <https://doi.org/10.1109/ICCCI.2015.7218064>
14. Liu X, Dai B, He H (2011) Real-time on-road vehicle detection combining specific shadow segmentation and SVM classification. In: 2011 Second international conference on digital manufacturing and automation. pp 885–888. <https://doi.org/10.1109/ICDMA.2011.219>
15. Shi K, Bao H, Ma N (2017) Forward vehicle detection based on incremental learning and fast R-CNN. In: 2017 13th International conference on computational intelligence and security (CIS). pp 73–76. <https://doi.org/10.1109/CIS.2017.00024>

16. Kalyan SS, Pratyusha V, Nishitha N, Ramesh TK (2020) Vehicle detection using image processing. In: 2020 IEEE international conference for innovation in technology (INOCON). pp 1–5. <https://doi.org/10.1109/INOCON50539.2020.9298188>
17. Roh HC, Sung CH, Chung MJ (2012) Fast vehicle detection using orientation histogram and segmented line projection. In: 2012 9th International conference on ubiquitous robots and ambient intelligence (URAI). pp 44–45. <https://doi.org/10.1109/URAI.2012.6462926>

Resilient Risk-Based Adaptive Authentication and Authorization (RAD-AA) Framework



Jaimandeep Singh, Chintan Patel, and Naveen Kumar Chaudhary

Abstract In recent cyber attacks, credential theft has emerged as one of the primary vectors of gaining entry into the system. Once attacker(s) have a foothold in the system, they use various techniques including token manipulation to elevate the privileges and access protected resources. This makes authentication and token-based authorization a critical component for a secure and resilient cyber system. In this paper, we discuss the design considerations for such a secure and resilient authentication and authorization framework capable of self-adapting based on the risk scores and trust profiles. We compare this design with the existing standards such as *OAuth 2.0*, *OIDC*, and *SAML 2.0*. We then study popular threat models such as *STRIDE* and *PASTA* and summarize the resilience of the proposed architecture against common and relevant threat vectors. We call this framework *Resilient Risk-based Adaptive Authentication and Authorization (RAD-AA)*. The proposed framework excessively increases the cost for an adversary to launch and sustain any cyber attack and provides much-needed strength to critical infrastructure. We also discuss the machine learning (ML) approach for the adaptive engine to accurately classify transactions and arrive at risk scores.

Keywords Federated authentication · Delegated authorization · Cyber resilience · Adaptive engine · Identity management systems · Threat models · Secure architecture and framework · OAuth 2.0 · OpenID connect · SAML 2.0

1 Introduction

As per the July-2022 report by IBM [19], one of the most frequent reasons for a data breach is using stolen or compromised credentials. The primary attack vector in 19% of incidents was stolen or exposed credentials. Stolen credentials can also lead to

J. Singh (✉) · N. K. Chaudhary
National Forensic Sciences University, Gandhinagar, Gujarat, India
e-mail: jaimandeep.phdcs21@nfsu.ac.in

C. Patel
The University of Sheffield, Sheffield, UK

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024
S. J. Patel et al. (eds.), *Information Security, Privacy and Digital Forensics*,
Lecture Notes in Electrical Engineering 1075,
https://doi.org/10.1007/978-981-99-5091-1_27

371

significant damage to the complete ecosystem as was seen in the case of ransomware attack on the colonial pipeline system in the United States [23]. Nowadays hackers do not break into the system but log into it, making the authentication and authorization systems a core design feature of any security architecture. With the help of authentication, we decide “who can enter” and with the help of authorization, we decide “up to what level he or she can access the system”.

Risk and trust-based adaptive approach for authentication and authorization (RAD-AA) is required for a secure and resilient cyber system. Stringent security requirements are known to increase the friction in user experience. The adaptive approach also helps in reducing this friction by adjusting security requirements based on the context and the flow. In a traditional authentication system the credentials of a user are verified in a sequential manner by employing various means such as user name, password, one-time password, and biometrics. However, in the RAD-AA approach the credential verification requirements are dependent on the *risk score* of the transactions and the *trust relationship* between different entities of the ecosystem. Example: For a high risk transaction, the system may ask for MFA such as approving the notification received over the device or answer the security questions, while at low risk, the system may continue based on just user name and password. RAD-AA-based framework decides *risk score* based on the user’s or system’s state. For example: If a user tries to log in from a different geographic location, device, vulnerable application, or browser, then the system increases the *risk score* for the transaction. Sometimes, if there is an attack on the complete critical infrastructure of the organization, then the RAD-AA framework-based system increases the *risk score* for the complete system and expects that each user must pass through the high-security verification. Similarly, for access control or authorization, based on *risk score*, the RAD-AA system can upgrade, degrade, or revoke the rights given to a user or group of users. In this paper, we propose a novel Resilient Risk-based Adaptive Authentication and Authorization (RAD-AA) framework that is attack-resilient and highly adaptive to the underlying risk factors.

The remainder of the paper is as follows: Sect. 2 presents related work. In this section, we discuss the existing features of *OAuth 2.0*, *OpenID connect*, and *SAML 2.0* standards. The most common threat models are discussed in Sect. 3 followed by design considerations for RAD-AA framework in Sect. 4. The detailed architecture and threat matrix of the RAD-AA framework is discussed in Sect. 5. In Sect. 6, we discuss the ML-based approach for adaptive engine. This is followed by conclusion and future work in Sect. 7.

2 Related Work

There are several authentication and authorization frameworks available and adopted by the industries. *OAuth 2.0* [8] protocol provides authorization for mobile, web, and desktop applications and other smart devices. The *OAuth 2.0* framework consists of four essential components: *resource owner*, *resource server*, *client*, *authorization*

framework. The RFC 6819 [11] defines the attacker capabilities and threat model for the *OAuth 2.0* standard such as obtaining client secrets, obtaining refresh tokens, obtaining access tokens, phishing of end-user credential, open redirectors on client, and password phishing.

Since *OAuth 2.0* provides only an authorization framework, *OpenID connect* [18] integrates an identity layer to it and enables the authentication of the end user for the client-side applications. An *OpenID connect* permits all types of clients such as JSCient, web client, and mobile client.

An open XML-based *SAML 2.0* (standard called Security Assertion Markup Language) [4, 7] is frequently used to exchange authentication and authorization (AA) data among federated organizations. With the help of *SAML 2.0*, the end user can log into multiple web-applications using the same credentials. *SAML 2.0* enables single sign-on (SSO) facilities to access several independent applications with the help of an identity provider. The *SAML 2.0* provides numerous advantages, such as users not having to remember multiple user names and passwords, which also reduces access time, cost reduction, and labor cost reduction.

There are other several authentication protocols such as *Lightweight Directory Access Protocol* (LDAP), *Kerberos*, and *RADIUS* but considering industry adoption and need for a resilient and adaptive authentication and authorization framework, we have compared the proposed framework with *OAuth 2.0*, *OpenID connect*, and *SAML 2.0* in Sect. 4.

3 Threat Models

This section presents widely used threat models which are relevant to our proposed framework. Threat modeling is the security procedure used to identify, classify, and examine potential risks. Threat modeling can be carried out either proactively during design and development or resolutely after a product has been released. In either situation, the method identifies the possible harm, the likelihood that it will happen, the importance of the issue, and the ways to remove or lessen the threat. Threats and vulnerabilities are frequently paired in order to find the risk to an organization or a system.

In this paper, we have adopted the relevant portions of the existing threat models and tailored them based on the characteristics of different entities of the framework and their interactions with internal and external entities. The first threat model we have considered is STRIDE, developed by Microsoft [20]. The STRIDE model discusses six different types of threats. The first threat is *Spoofing* where the adversary spoofs user identity and tries to access the resources with *secure authentication* as a desirable property. The next threat is *Tampering* where an adversary tampers the messages communicated over the open channel, and the desired property is *integrity*. The third threat is *repudiation* where attacker performs some illegal activities and denies performing those activities. The desirable property to tackle this threat is *non-repudiation*, achievable using secure signature methods. The next threat is *informa-*

tion disclosure where the adversary tries to read or access the secret information. The desirable property is the *confidentiality* that can be achieved using secure encryption/decryption system. The next threat is *denial of service attack* where an adversary tries to prevent an end-user from accessing services where the desirable property is *availability* that is achieved through intrusion detection and intrusion prevention system. According to the STRIDE threat model, the last threat is *elevation of privilege* where an attacker is either an insider or somehow became the trusted insider with the high privilege to destroy the system. In this paper, we consider a number of threats from the STRIDE model such as *spoofing identity* and *tempering*.

The *Process for Attack Simulation and Threat Analysis* (PASTA) is a risk-centric threat modeling technique that incorporates risk analysis and context into the complete security of critical infrastructure from the start [24]. PASTA allows threat modeling linearly through the interdependent seven stages. In the *first stage*, PASTA defines the *objectives for risk analysis* and provides well-defined business objectives and analysis reports. In *stage two*, PASTA defines the *technical scope* and tries to understand possible attack surface components and provides a detailed technical report on all attack surface components. In the *third stage*, PASTA performs *application decomposition and analysis* and provides a data-flow diagram, interfaces list with trust level, asset list, and access control matrix. In the *fourth stage*, PASTA performs *threat analysis* and generates a list of threat agents, attack vectors, and incident event reports. In *stage five*, it performs *vulnerability assessments* and provides scoring based on Common Vulnerability Scoring System (CVSS). In stage six, PASTA performs *attack modeling* and simulation of various well-known threats. As an outcome of stage six, it provides attack trees and possible attack paths. In the last (*seventh*) stage, it performs *risk analysis and management* where it outcomes risk profile, risk mitigation strategy, and threat matrix. We have considered the PASTA framework wherever required during the framework design.

4 Design Considerations for RAD-AA Framework

In this paper, we propose a Risk-based Adaptive Authentication and Authorization (RAD-AA) framework which is secure by design and excessively increases the cost for the attacker. The design considerations include adaptive ecosystem which is capable of modifying the security requirements and the access rights to the protected resources based on the risk score of the transactions and interactions between different entities of the ecosystem. Table 1 presents a comparison of the proposed framework with the existing standards.

The adaptive design enables the framework to anticipate the attack. In case of a breach, the adaptive design is able to *withstand* and *constrain* the level of damage by revoking or restricting the access rights granted or by de-authenticating already authenticated users or increasing the security requirements of the transactions. The design considerations of RAD-AA are given below.

Table 1 Comparison of existing standards with proposed design considerations of RAD-AA

Design considerations	OAuth 2.0 [8]	OpenID connect [18]	SAML 2.0 [7]	Proposed framework
Authentication	No	Yes	Yes	Yes
Adaptive engine for cyber resilience	No	No	No	YES
Federated authentication	No	SSO Only	SSO Only	YES
Delegated authorization	Yes	Yes	Yes	Yes
Decoupling authentication and authorization	Yes	Yes	Yes	Yes
Out of the box support for confidentiality and Non-repudiation of claims	No	No	No	Yes
Audience binding	No	No	No	Yes
Trust relationships	No	No	No	Yes
Time-limited validity of claims	Yes	Yes	Yes	Yes
Ability to revoke issued tokens	No	No	No	Yes
Support for REST API architecture	Yes	Yes	Yes	Yes
Extensible to ML-based classification	No	No	No	Yes

- **Adaptive Engine for Cyber Resilience.** The authentication and authorization transactions and trust level between different entities in the ecosystem should be able to adapt based on the risk score such as Geo-location, impossible travel, IP reputation, and device information. The system should be able to enhance, reduce, or completely revoke the entry into the ecosystem or access to the protected resources based on the risk score.
- **Federated Authentication.** It is a system in which two parties trust each other to authenticate the users and authorize access to resources owned by them. In a Federated Authentication or Identity management system, the identity of a user in one system is linked with multiple identity management systems. It extends beyond a single organization wherein multiple organizations can agree to share the identity information and join the federation. Once the users login into their

organizations, they can use this federated identity to access resources in any other organization within the federation.

- **Delegated Authorization.** A delegated authorization system can delegate access rights to the services or processes in the form of assertions or claims. Examples include end-user authorization delegation to a web or native application.
- **Decoupling Authentication and Authorization.** Decoupled authentication and authorization systems allow for modularity and provide necessary interfaces to replace one system with another.
- **Confidentiality and Non-repudiation of claims.** The confidentiality will ensure that the contents of the claims are not revealed to unauthorized parties. Non-repudiation will assure the recipient that it is coming from the identified sender and that the data integrity is assured.
- **Audience Binding.** The primary aim of binding the token to the audience or the client to which it is issued is to prevent unauthorized entities from using the leaked or stolen tokens. The token, when issued, is bound to a public key of the audience or the client to which it is issued. The client now needs the corresponding private key to use the token further. This will protect against the token's misuse by unauthorized parties that do not possess the private key. It provides an added assurance to the receiver that such a claim token has been sent by the sender authorized to use it.
- **Trust Relationships.** The system should be capable of establishing trust relationships between different entities of the ecosystem. The trust relationship will also depend on how the two entities identify each other and what the relationship between the entities is. The risk score of the client applications can be taken into consideration while deciding the trust relationship level.
- **Propagation of identity and claims between different ecosystems.** The requested resource or a subset of the resource may be required to be fetched from a different ecosystem. This necessitates the propagation of identity and claims across different ecosystems. When transitioning from one ecosystem to another, properties of confidentiality, non-repudiation, limiting the information contained in the claim, and assurance of use by the authorized sender should be maintained.
- **Time-Limited Validity and Revoking Issued Claims.** The lifetime of the issued tokens should be limited to avoid misuse of the stolen tokens [8]. The ability to revoke the tokens should be in-built into the design.
- **Support for REST API architecture.** Most of the developers have now moved from WS-* to REST as conceptualized by Fielding in his seminal PhD dissertation [6] and stateless APIs [12]. The system design should therefore support the REST API architecture. The REST can simply be described as HTTP commands pushing JSON packets over the network as defined in RFC 8259 [2].

5 Architecture and Threat Matrix for RAD-AA Framework

In this section, we describe the detailed architecture of RAD-AA based on the various existing standards [2, 7, 8, 10, 18], threat considerations [11], best current practices [3, 9], and augmented features of OAuth 2.0 [21]. The architecture, entities, protocol flow, and their interactions are given in Fig. 2.

We have analyzed various threats that can manifest in any authentication and authorization protocol. We have then brought out the features in the proposed framework that can mitigate the effect of these threats by anticipating and adapting themselves in the face of the changing operating environment.

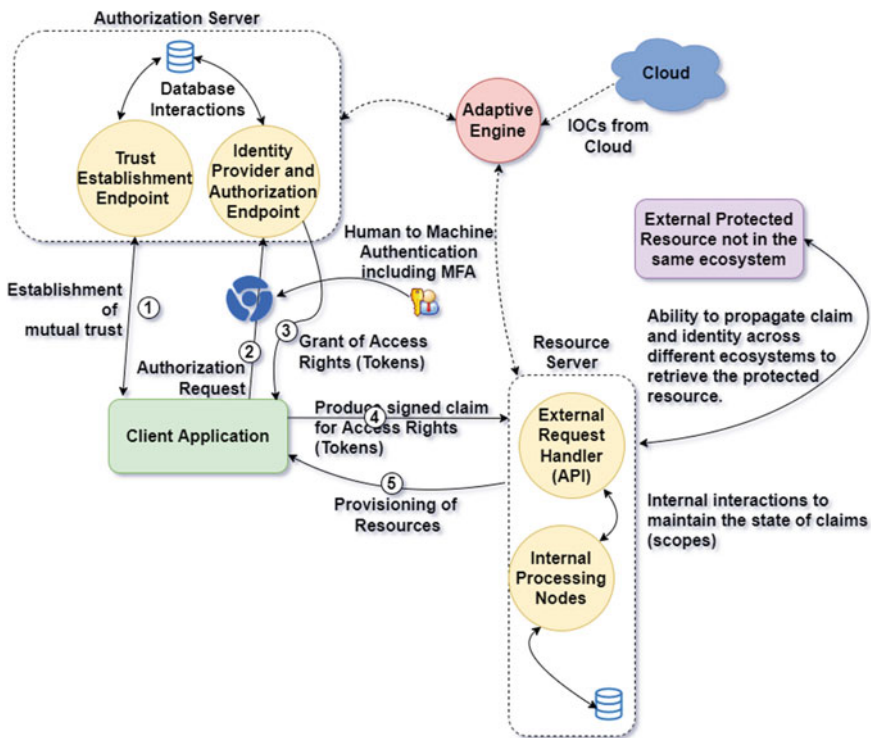


Fig. 1 Architecture of RAD-AA framework

5.1 Entities

The RAD-AA framework consists of following entities:

- **Resource Owner.** It is an entity which is capable of granting access to a protected resource.
- **Resource Server (RS).** This server hosts the protected resources and is capable of accepting and responding to protected resource requests based on the access tokens and scopes.
- **Client.** This is an application which makes a request for accessing protected resource requests on behalf of the resource owner after due authorization by the resource owner.
- **Authorization Server (AS).** This server issues access tokens to the client after successfully authenticating the resource owner and obtaining authorization.
- **Adaptive Engine.** This engine analyzes various transactions in the protocol flow. It then assigns the risk score which is used by various entities like authorization server and resource server to modify their behavior by making the security requirements more stringent or to limit or deny access to the protected resources.

5.2 Protocol Flow and Threat Matrix

The summary of common threat vectors and resilience features of RAD-AA framework is given in Table 2.

5.2.1 Risk Score-Based Adaptive Engine

The adaptive engine will determine the risk score of each transaction in the protocol flow based on AI/ML models. It will provide its inputs to all the entities in the ecosystem such as authorization server and the protected resource server. The engine will take into consideration various parameters such as the level of trust relation between the sender and the receiver, Geo-location, impossible travel, IP reputation, and device information and then classify the transaction into *LOW*, *MEDIUM*, and *HIGH* risks.

The authorization server and the protected resource server based on the classification of the transaction can do either of the following:

- **Raise the security requirements.** The security requirements of the transaction can be made more stringent by demanding the sender to provide additional verification details.
- **Accept or limit/Reject the Request.** The AS/RS based on the risk score of the transaction can either accept the transaction or lower the authorization of the scopes or completely reject the transaction as a high risk transaction.

Table 2 Summary: threat matrix for RAD-AA framework

Threat vectors	RAD-AA cyber resilience features
Client impersonation	Mutual authentication using mTLS [3] or DPOP [5]
Cross-site request forgery (CSRF)	AS supports proof key for code exchange (PKCE) [16]
Authorization Server (AS) Mix-Up attack	Validation of the issuer of the authorization response
Cross-Origin Resource Sharing (CORS)	<ul style="list-style-type: none"> – Establish trust relationships through mutual authentication – Calculate risk score before allowing transactions – Information in HTTP request is assumed fake
Cross-Site Scripting (XSS)	<ul style="list-style-type: none"> – CSP headers – Input validation
DDoS on AS	<ul style="list-style-type: none"> – Check HTTP request parameters are not pointing to unexpected locations, RFC 9101 [17] – Adaptive engine to thwart malicious requests based on the risk score and the trust relations between the client applications and the authorization server
Access token injection	Audience restricted token binding
Access token replay	Sender-constrained and audience-restricted access tokens

The authorization and access to the protected resources will still be governed by the requirements specified in the protocol. The engine will only classify the transaction cost as an additional security layer. The engine classification at no time will bypass the requirements of other validation requirements as given in the specifications. The common risks associated with the adaptive engine which are ordinarily associated with AI/ML engines must be taken into consideration [13].

5.2.2 Establishment of Mutual Trust

The client applications like native desktop or mobile app, JavaScript-based Single Page App, or the web server-based web apps need to establish a level of trust while authenticating their identities with the AS/RS. Each such client authentication will be assigned a trust assurance level which will regulate the ability of the client application to acquire elevated authorization rights or scopes for an extended duration.

- **Trust Assurance Level 0.** This is the minimum trust assurance level. The application does not have means of mutual identification and authentication. The delegated authorization rights would be restricted to a minimum level of permissions

- (scopes). The lifespan of the delegated authorization rights (in terms of access and refresh tokens) will be minimal or for one-time use only.
- **Trust Assurance Level 1.** Client application has the means of establishing mutual identification and authentication by using mTLS [3] or DPop [5]. The delegated authorization rights/scopes will be more permissive in nature with an extended lifetime validity.

Threat Vector: Client Impersonation.

Threat Description and Artifacts: The identity of the client application can be spoofed and a malicious client can impersonate as a genuine client.

Mitigating Features: In the proposed framework, the client and authorization server need to establish a trust relation by mutually identifying and authenticating each other by using mTLS [3] or DPoP [5] or similar such standards.

5.2.3 Authorization Request

The client application will initiate the request to the authorization server for grant of access rights to acquire the protected resources owned by the resource owner. The resource owner will authorize the scopes that the authorization server should grant to the requesting client. The access rights of the client will be constrained and restricted to the scopes authorized by the resource owner. The authorization server will first verify the identity of the resource owner by passing the request to the federated identity management system [15]. The Pushed Authorization Requests (PAR) as defined in RFC 9126 [10] will be used to initiate the authorization request by the client.

Threat Vector: Cross-Site Request Forgery (CSRF).

Threat Description and Artifacts: Clients Redirect/Request URIs are susceptible to CSRF.

Mitigating Features: The framework prevents CSRF attacks against client's redirect/request URIs by ensuring that the authorization server supports Proof Key for Code Exchange (PKCE) [16]. The CSRF protection can also be achieved by using “nonce” parameter, or “state” parameter to carry one-time-use CSRF tokens.

Threat Vector: Authorization Server (AS) Mix-Up attack.

Threat Description and Artifacts: The AS attack can manifest when multiple AS are used in which one or more is a malicious AS operated by the attacker.

Mitigating Features: The framework prevents mix-up attacks by validating the issuer of the authorization response. This can be achieved by having the identity of the issuer embedded in the response claim itself, like using the “iss” response parameter.

Threat Vector: Cross-Origin Resource Sharing (CORS) attacks.

Threat Description and Artifacts: CORS allows web applications to expose its resources to all or restricted domains. The risk arises if the authorization server allows for additional endpoints to be accessed by web clients such as metadata

URLs, introspection, revocation, discovery, or user info endpoints. These endpoints can then be accessed by web clients.

Mitigating Features: In the proposed framework, we establish trust and calculate risk score before allowing any transactions. It is accepted that access to any of the resources as all the information contained in the HTTP request can be faked.

Threat Vector: Cross-Site Scripting (XSS) attacks.

Threat Description and Artifacts: Cross-Site Scripting (XSS) attack risk arises when the attacker is able to inject malicious scripts into otherwise benign and trusted websites.

Mitigating Features: Best practices like injecting the Content-Security-Policy (CSP) headers from the server are recommended, which is capable of protecting the user from dynamic calls that will load content into the page being currently visited. Other measures such as input validation are also recommended.

Threat Vector: DDoS Attack on the Authorization Server.

Threat Description and Artifacts: A large number of malicious clients can simultaneously launch a DoS attack on the authorization server by pointing to the “request_uri” as defined in PAR RFC 9126 [10].

Mitigating Features: To mitigate the occurrence of such an attack, the server is required to check that the value of the “request_uri” parameter is not pointing to an unexpected location as recommended in the RFC 9101 [17]. Additionally, the framework employs an adaptive engine to thwart such requests based on the risk score and the trust relations between the client applications and the authorization server.

5.2.4 Delegation of Access Rights

Once the resource owner has authorized the request, authorization grant has been received by the resource owner, and the request received from the client application is valid, the authorization server issues access tokens and optional refresh tokens as defined in RFC 6749 [8].

Threat Vector: Access Token Injection.

Threat Description and Artifacts: In this attack, the attacker attempts to utilize a leaked access token to impersonate a user by injecting this leaked access token into a legitimate client as defined in draft OAuth 2.0 Security Best Current Practice, 2022 [9].

Mitigating Features: In the proposed framework, the token is issued after binding to the client application.

5.2.5 Accessing Protected Resources

The client application can access protected resources by presenting the access token to the resource server as defined in OAuth 2.0 RFC 6749 [8]. The resource server will

then check the validity of the access token. It will also ensure that the token has not expired and that the requested resource is conformance with the scopes authorized in the token.

Threat Vector: Access Token Replay Attack.

Threat Description and Artifacts: An attacker can attempt to replay a valid request to obtain or modify/destroy the protected resources as defined in OAuth 2.0 threat model RFC 6819 [11].

Mitigating Features: The proposed framework mandates sender-constrained and audience-restricted access tokens as defined in draft OAuth 2.0 Security Best Current Practice, 2022 [9]. In addition, the resource server may reduce the scope or completely deny the resource based on the risk score of the client application from where the request has materialized.

6 ML-Based Classification Approach for Adaptive Engine

The adaptive engine can employ various ML algorithms or models to classify the transaction data into *HIGH*, *MEDIUM*, or *LOW* risk on a real-time basis. Though there is a high chance that data received by the adaptive engine can be noisy, the K-Nearest Neighbor (KNN) [14] or Random Forest (RF) [1] can be adopted based on their capability to work with multidimensional feature sets. RF requires a large training dataset that can be gradually built into the cloud-based system and later used. Adoption of *Incremental Machine Learning (IML)* [22] approaches over the adaptive engine can provide more accurate classification as it continuously trains the model and over a period of time (Fig. 2).

The output of the adaptive engine is taken into consideration by various systems to grant or deny permissions or accept/reject or terminate the transaction, and restart the process. Some important features that can provide a high degree of correlation to classify the transaction data are tabulated in Table 3.

The Network Intrusion Detection System (NIDS) will take Netflow data as an input stream. The output of NIDS will become one of the input features for the adaptive engine. Trust Assurance Level is an important feature based on the client applications' verifiable identity and authentication. A client application that can authenticate itself by using various technologies/protocols like mutual-TLS will have a higher trust assurance level than client applications that cannot authenticate themselves. The model or the algorithm can be fine-tuned by adding other features, hyper-parameters or removal/addition of necessary weights or biases for some features over others.

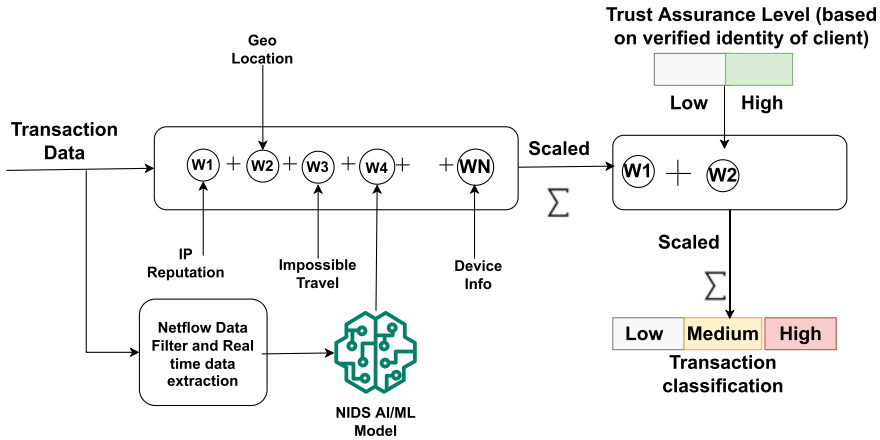


Fig. 2 Use of ML-based classification in adaptive engine

Table 3 Classification features

Features	Descriptions
IP reputation	The IP reputation will let the system know if the request is coming from an IP with a bad reputation
Geo-location	This will identify the geo-location of the transaction
Impossible travel	This feature will indicate if the same user has connected from two different countries or geographical locations, and the time between these locations can't be covered through conventional travel means
Device info	Device information will provide the necessary data about the device being used by the client application
NIDS output	It will take Netflow data as input stream and classify data as malicious or benign
Trust assurance level	Trust Assurance Level will determine the degree of trust that server can have on the client application based on the verifiable client identity

7 Conclusion and Future Work

This paper presents the design considerations and architecture for a Resilient Risk-based Adaptive Authentication and Authorization (RAD-AA) Framework. The proposed framework achieves higher level of resilience than the existing frameworks such as *OAuth 2.0*, *OpenID Connect*, and *SAML 2.0* with the help of an adaptive

engine based on risk score and trust relationship. In future, we aim to continue developing this framework and consider further aspects related to its deployment on different platforms using ML.

References

1. Biau G, Scornet E (2016) A random forest guided tour. *Test* 25(2):197–227
2. Bray T (2017) The JavaScript Object Notation (JSON) Data Interchange Format. RFC 8259 (Dec 2017), <https://www.rfc-editor.org/info/rfc8259>
3. Campbell B, Bradley J, Sakimura N, Lodderstedt T (2020) OAuth 2.0 Mutual-TLS client authentication and certificate-bound access tokens. RFC 8705, <https://www.rfc-editor.org/info/rfc8705>
4. Cantor S, Moreh J, Philpott R, Maler E (2005) Metadata for the oasis security assertion markup language (saml) v2. 0
5. Fett D, Campbell B, Bradley J, Lodderstedt T, Jones M, Waite D (2022) OAuth 2.0 demonstrating proof-of-possession at the application layer (DPoP). Internet-Draft draft-ietf-oauth-dpop-10, Internet Engineering Task Force, <https://datatracker.ietf.org/doc/draft-ietf-oauth-dpop/10/>, work in Progress
6. Fielding RT (2000) Architectural styles and the design of network-based software architectures. University of California, Irvine
7. Groß T (2003) Security analysis of the SAML single sign-on browser/artifact profile. In: Proceedings of the 19th Annual computer security applications conference. IEEE, pp 298–307
8. Hardt D (2012) The OAuth 2.0 authorization framework. RFC 6749. <https://www.rfc-editor.org/info/rfc6749>
9. Lodderstedt T, Bradley J, Labunets A, Fett D (2022) OAuth 2.0 Security best current practice. Internet-Draft draft-ietf-oauth-security-topics-20, Internet Engineering Task Force, <https://datatracker.ietf.org/doc/draft-ietf-oauth-security-topics/20/>, work in Progress
10. Lodderstedt, T., Campbell, B., Sakimura, N., Tonge, D., Skokan, F.: OAuth 2.0 Pushed Authorization Requests. RFC 9126 (Sep 2021), <https://www.rfc-editor.org/info/rfc9126>
11. Lodderstedt T, McGloin M, Hunt P (2013) OAuth 2.0 threat model and security considerations. RFC 6819, <https://www.rfc-editor.org/info/rfc6819>
12. Masse M (2011) REST API design rulebook: designing consistent RESTful web service interfaces. O'Reilly Media, Inc
13. McLean S, Read GJ, Thompson J, Baber C, Stanton NA, Salmon PM (2021) The risks associated with artificial general intelligence: a systematic review. *J Expe Theor Artif Intell*:1–15
14. Peterson LE (2009) K-nearest neighbor. *Scholarpedia* 4(2):1883
15. Hedberg R, Jones M, Solberg A, Bradley J, Marco GD (2022) Openid connect federation 1.0—draft 20. The OpenID Foundation, https://openid.net/specs/openid-connect-federation-1_0.html
16. Sakimura N, Bradley J, Agarwal N (2015) Proof key for code exchange by OAuth public clients. RFC 7636, <https://www.rfc-editor.org/info/rfc7636>
17. Sakimura N, Bradley J, Jones M (2021) The OAuth 2.0 authorization framework: JWT-secured authorization request (JAR). RFC 9101, <https://www.rfc-editor.org/info/rfc9101>
18. Sakimura N, Bradley J, Jones M, De Medeiros B, Mortimore C (2014) Openid connect core 1.0. The OpenID Foundation, S3p. https://openid.net/specs/openid-connect-core-1_0.html
19. Security I (2022) Cost of a data breach report. Technical Report. <https://www.ibm.com/security/data-breach>
20. Shostack A (2008) Experiences threat modeling at microsoft. MODSEC@ MoDELS 35
21. Singh J, Chaudhary NK (2022) Oauth 2.0: architectural design augmentation for mitigation of common security vulnerabilities. *J Inf Secur Appl* 65:103091

22. Solomonoff RJ (2002) Progress in incremental machine learning. In: NIPS workshop on universal learning algorithms and optimal search. Citeseer, Whistler, BC
23. Su Y, Ahn B, Alvee SR, Kim T, Choi J, Smith SC (2021) Ransomware security threat modeling for photovoltaic systems. In: 2021 6th IEEE workshop on the electronic grid (eGRID). IEEE, pp 01–05
24. UcedaVelez T, Morana MM (2015) Risk centric threat modeling: process for attack simulation and threat analysis. John Wiley & Sons

Survey on Blockchain Scalability Addressing Techniques



B. S. Anupama and N. R. Sunitha 

Abstract Decentralized, blockchain-based cryptocurrencies have received a lot of interest and have been used extensively in recent years. Blockchain technology, which is rapidly establishing itself as the most revolutionary technology of recent years is attracting the interest of both the private and public sectors. This blockchain technology has been frequently used in decentralized cryptocurrencies like Bitcoin and Ethereum. Blockchain is being widely used for purposes other than cryptocurrencies. Despite these advantages, scalability remains a major barrier to the widespread use of blockchain and thus a problem worth addressing. In this paper, initially, we discuss blockchain's evolution in four stages, challenges of blockchain technology, and scaling quadrilemma and trilemma. Later we address the challenges associated in achieving the scalability in the blockchain network and review the approaches for addressing blockchain scalability challenges in different layers with different techniques. We envision this paper as a reference for analyzing and conducting research on the scalability of blockchains.

Keywords Scalability · Consensus · Blockchain

1 Introduction

The ability of a blockchain to handle an increasing number of transactions is referred to as its scalability. The primary cause of the scalability problem is that for a transaction to be accepted as genuine, consensus needs to be arrived among stake holders. Currently, Bitcoin can process seven transactions per second (TPS), which is known as its throughput. With about 30 TPS, Ethereum is occupying a marginally higher position. These figures don't seem too bad at first, but they are nothing compared to Visa's throughput, which may reach up to roughly 1,700 TPS. Therefore, there is a

B. S. Anupama (✉) · N. R. Sunitha

Department of Computer Science and Engineering, Siddaganga Institute of Technology, Affiliated to Visvesvaraya Technological University, Belagavi 590018, Tumkur, Karnataka, India
e-mail: anupamabs@sit.ac.in

need for research to be done to improve the number of TPS in blockchain technology. The various techniques available in literature addressing the scalability problem is herewith discussed in the following sections.

1.1 Blockchain’s Evolution

The evolution of blockchain technology is examined by versioning from 1.0 to 4.0 as shown in Fig. 1.

1.1.1 Blockchain 1.0—Cryptocurrency

The first application of distributed ledger technology (DLT) is cryptocurrency. It enables the execution of transactions in the financial field based on DLT or blockchain technology, with Bitcoin serving as the most notable example in this sector.

1.1.2 Blockchain 2.0—Smart Contracts

Smart contracts are programs that are deployed in the blockchain. They are self-sufficient programs that run automatically under predetermined conditions such as the facilitation, verification, or enforcement of contract fulfillment. The main advantage of smart contracts is they cannot be tampered or hacked. So, it reduces the cost of verification and execution. Smart contracts are mostly deployed in the Ethereum blockchain.



Fig. 1 Evolution and versioning of blockchain

1.1.3 Blockchain 3.0—Decentralized Application (DApps)

Communication and storage are decentralized in DApps. So, most of the DApps backend code run in a decentralized p2p network. DApps frontend code can be in any language which makes a call to its backend. DApps frontend is hosted on decentralized storage.

1.1.4 Blockchain 4.0—Blockchain in the Industry (4.0)

Previous versions laid the foundation for this version. Blockchain 4.0 describes concepts and solutions that adapt blockchain technology to meet business needs. An increase in the degree of privacy protection and trust are the demands of the industrial revolution.

1.2 Challenges in Blockchain Technology

1.2.1 Inefficient Technological Design

Despite its many benefits, blockchain technology still has a number of technical restrictions. One of the essential components of this is a programming error or bug.

1.2.2 The Criminal Connection

The experts and criminals are attracted to blockchain technology due to its anonymous features. Due to the decentralized nature of the network, it is very difficult to know the true identity of the node.

1.2.3 Scalability

Scalability is the main challenge in implementing blockchain technology. The transitions take more time to process as the number of users increases in the network.

1.2.4 High Energy Consumption

Proof of work is used as a consensus protocol in most of the blockchain networks. Mining in Proof of Work requires a lot of computational power.

1.2.5 No Regulation

Most of the organizations use blockchain for transactions. There are no specific rules to be followed in the blockchain.

1.2.6 Lack of Skills

To handle blockchain, one needs to recruit a skilled person in addition to software and hardware.

1.2.7 Slow

Transaction processing requires more time due to the complex nature of the blockchain technology. Even the encryption makes it very slow.

1.2.8 Lack of Awareness

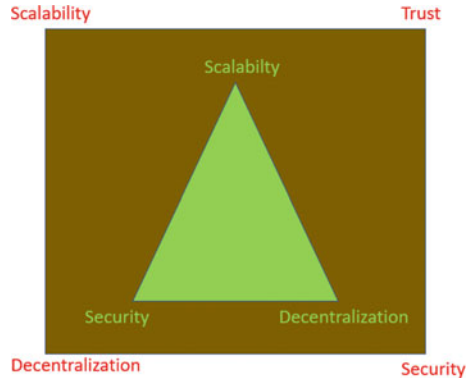
Many people are not aware of the existing blockchain technology.

1.3 Scaling Quadrilemma and Trilemma

According to the scalable trilemma, as shown in Fig. 2 (inner diagram in Fig. 2), all three cannot entirely coexist without sacrificing any two. Trust is very important in the scalability of blockchain. Blockchains with trustworthy parties may use simpler consensus, communications, and computations to increase scalability. Due to this Blockchain scalability is extended from trilemma to quadrilemma (outer square in Fig. 2) [1].

The “blockchain scalability quadrilemma” refers to the trade-off between scalability, decentralization, security, and trust in the current blockchain systems on top of the scalability trilemma. In the current blockchain technology, achieving all the 4 properties at the same time is very difficult. For instance, consortium and private blockchains, which are fully or partially centralized yet contain completely trusted partners, achieve security and scalability. DAG-based blockchains, which are less trustworthy and more scalable, accomplish decentralization and scalability. Public blockchains on the other hand have weak scalability but strong security and decentralization. Therefore, ideal degrees of quadrilemma must be established to develop an ideal blockchain system.

Fig. 2 Scaling quadrilemma and trilemma



2 Challenges in Achieving Scalability

Factors affecting the Scalability issues in a blockchain network as shown in Fig. 3 [2, 3].

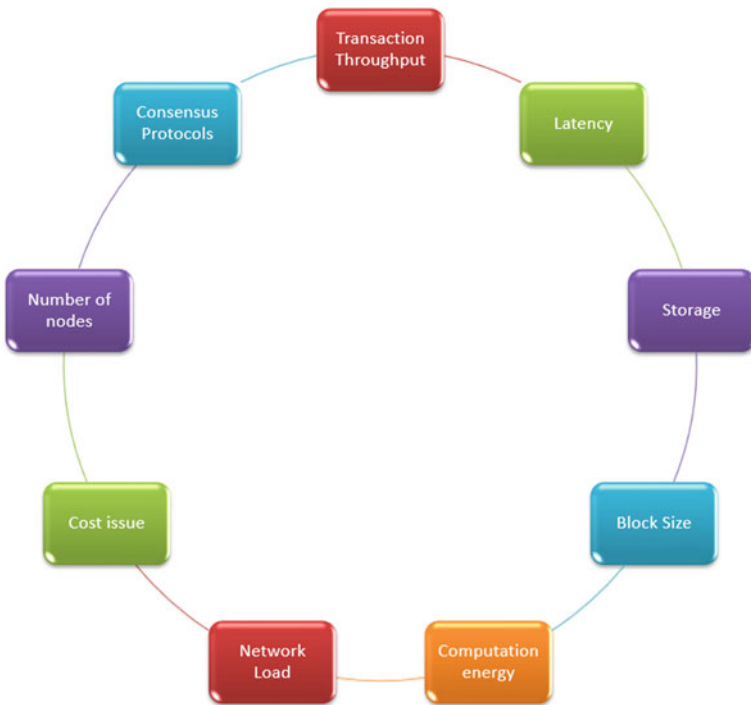


Fig. 3 Factors affecting scalability

Latency: This is related to the length of time required to reach a consensus after a transaction is launched.

Transaction throughput: It refers to the maximum number of transactions that are handled in one second.

Storage: Refers to the total space required by the blockchain network.

Block Size: The maximum number of transactions a block can store. If the storage capacity exceeds then the block will be rejected by the network.

Computation Energy: Energy required by the protocol to mine the blocks in the network.

Network Load: Number of transactions that can be handled by the network.

Cost issue: Total cost required to verify the transactions in the blockchain network.

Number of nodes: Number of users present in the blockchain network.

Consensus Model: It is used to approve and validate the transactions in the blockchain network.

3 Approaches for Addressing Blockchain Scalability Challenges

3.1 Layer 0—Approaches with Propagation of Protocols

It is used for communication between the blocks in the blockchain. To improve the throughput, data enhancement and optimization are required.

bloXroute [4]: Increasing the block size while reducing the time between blocks is the foundation of the design of the network. It avoids fork and enables fast propagation.

Velocity [5]: Fountain code, a type of erasure coding is used by velocity to decrease the amount of data that needs to be propagated, which also improves block propagation. It increases the throughput of the network.

Kadcast [6]: It is based on the Kademlia Architecture and functions like the process used for enhanced broadcasting with movable overhead. It enables secure and fast transmission.

Erlay: Erlay improves the transaction relay technology in Bitcoin to use less bandwidth overall while using more propagation latency. The protocol reduces costs while enhancing network connectivity.

Layered approaches to Scalability are shown in Fig. 4. Scalability solutions in different layers are shown in Table 1.

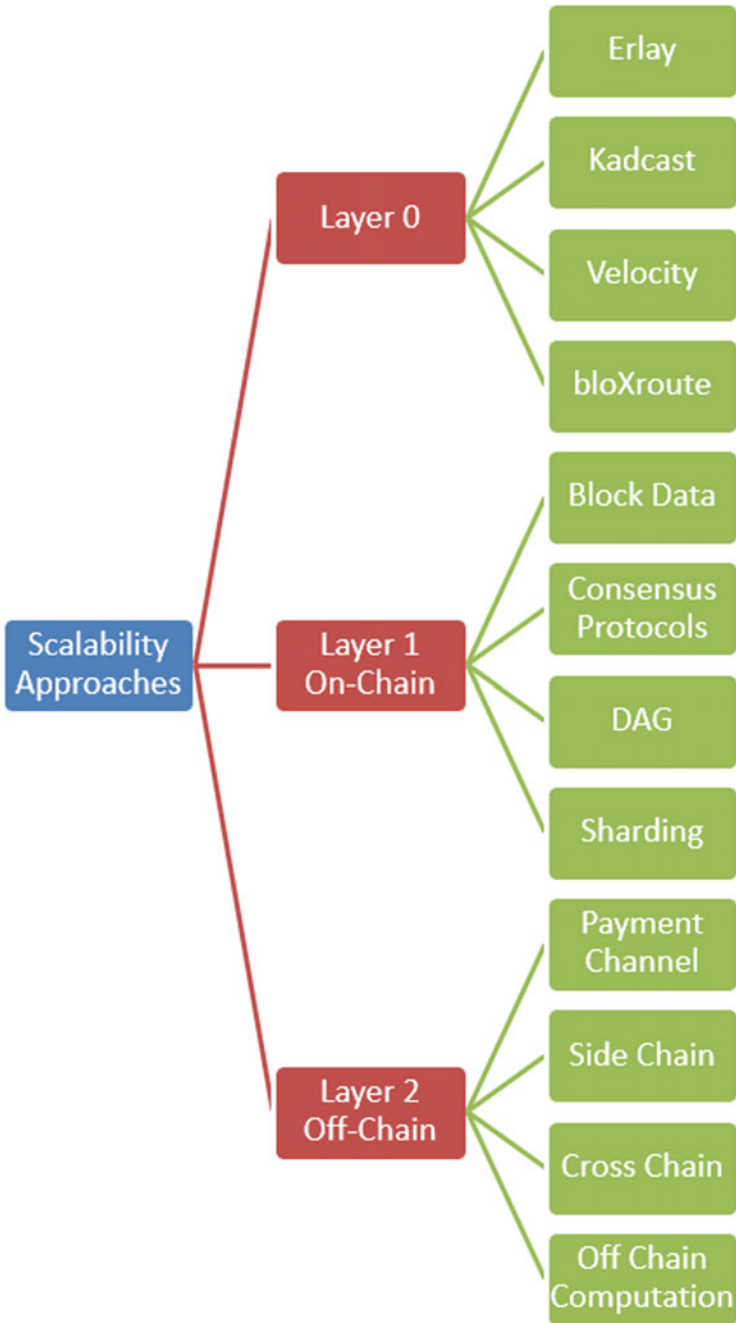


Fig. 4 Layered approaches of scalability

Table 1 Scalability solutions in different layers

Layer	Techniques	Scalability solutions
Layer 0	Data Propagation	Erlay, Kadcast, velocity, bloXroute
Layer 1 on-chain	Block data	Bitcoin-cash, CUB, compact block relay, Jidar, SegWit, Txilm
	Consensus protocols	Proof of work, proof of stake, delegated proof of stake, proof of authority, practical byzantine fault tolerance, bitcoin-NG, Algorand, Snowwhite, Ouroboros, Casper, Permacoin, Sawtooth, scp, Spacemint
	Directed acyclic graph (DAG)	SPECTRE, Conflux, IOTA, Nano, Aleph, Avalanche, Blockclique, Blockmania, CAPER, CDAG, Chainweb, DEXON, Dlattice, Dagcoin, Eunomia, GHOST, Graphchain, Haoitia, Byteball, Hashgraph, JointGraph, Lachesis, Meshcash, Obyte, OHIE, Inclusive, PARSEC, PHANTOM, Prism, StreamNet, Vite
	Sharding	Elastico, OmniLedger, RapidChain, Monoxide, Anon, Chainspace, Harmony, Logos, Ostraka, SSChain, Stegos, Zilliqa
Layer 2 off-chain	Payment channel	Lightning network, DMC, Raiden network, sprites, trinity, OUTPACE, A2L, Bolt, TumbleBit, Perun
	Side chain	Pegged Sidechain, Plasma, Liquidity network, DAppChains, Matic, Liquid, SKALE, NOCUST, Polygon’s PoS, Cartesi, Fuel Network, Off-chain Labs, OMG Network, Optimism, Aztec 2.0, Loopring, Starkware, zkSync, ZK-Rollup
	Cross-chain	Cosmos, Polkadot, Celer, Connnext, DeXTT, Loom, PoA, Waterloo, OX, Tokrex, AION, ARK, Komodo, Geeq
	Off-chain computation	Truebit, Arbitrum, BITE, Teechain, Teechan, Tesseract, ZLite

The advantages and disadvantages of Layer 0 scalable techniques are shown in Table 2.

Table 2 The advantages and disadvantages of Layer 0

Scalable techniques	Advantages	Disadvantages
Erlay	(1) Cost is affordable (2) Transmission is private	Less Secure
Kadcast	(1) Secure transmission is enabled (2) Fast propagation is enabled	Design complexity
Velocity	Throughput is high	Less consistent
bloXroute	(1) Fast propagation is enabled (2) Forks are avoided	Privacy is less

3.2 *Layer 1—Approaches Within the Blockchain Network (on Chain)*

It is used for implementing blockchain. So, it is also called as blockchain layer. In this layer, blocks are created and broadcasted in the network. Scaling the network is hard at this layer.

3.2.1 **Block Data**

The performance constraint of blockchain systems will be the restricted bandwidth and storage of the individual nodes. The system needs to communicate significantly more block data as TPS increases, which could make the congestion issues worse.

Segregated Witness (SegWit): The block size is maintained while new transactions are added to the block using the SegWit mechanism. To successfully create more room for new transactions, this strategy seeks to separate the signature information from the transaction and store it elsewhere. To accomplish the objectives, the transaction is split into two pieces, the unlocking signatures are removed from the first transaction hashes, and both the scripts and signatures are added to the new Witness structure.

Bitcoin-Cash: The scalability issue caused a hard fork in 2017 that resulted in the creation of the Bitcoin and Bitcoin-Cash blockchain branches. The size of the block is 1MB in bitcoin whereas the size of the block in bitcoin-cash has increased to 8MB. Further, the bitcoin-cash upgraded the size of the block to 32MB. The average block interval time is 10 min in Bitcoin.

Compact Block relay: The block compression approach was used in the design and implementation of the compact block relay. It is based on reducing the transaction header data and altering the original Bitcoin block's data structure.

Txilm [7]: The method of Txilm is built on the same idea of block compression. These methods are vulnerable to hash collisions. To conserve network bandwidth, the Txilm protocol is based on BIP152 and compresses the transactions within each block.

CUB [8]: A Consensus Unit is created by CUB using a plan that groups various nodes. Each node in a unit contains a portion of the block data. To reduce the overall query cost, the unit's nodes are given the blocks of the entire chain.

Jidar [9]: Jidar is a method of data compression for the Bitcoin system. Jidar's major goal is to relieve each node's storage pressure by letting users only keep the information they need. Each node only keeps a small portion of the full contents of a new block, including important transactions and Merkle branches. Jidar uses a bloom filter to determine whether a transaction's input has been used.

3.2.2 Consensus Protocols

A Blockchain system’s long-term stability depends on consensus processes. Consensus is the cornerstone of blockchain technology. Consensus protocols [10] are created to increase Blockchain’s effectiveness while also addressing the unique requirements of various application domains. The consensus mechanism boosts the system’s throughput, which improves the efficiency of the blockchain. Figure 5 shows the scalable consensus protocols.

Proof of Work (PoW): of Work (PoW)

Bitcoin was the first cryptocurrency to be used. The blockchain transactions are verified and approved by the consensus process. Transactions in PoW are verified through mining. Network miners attempt to crack the cryptographic conundrum when new transactions are made. The miner who solves the challenge first is responsible for producing a block and posting it for verification by blockchain nodes.

Proof of Stake (PoS): In PoS, validators who will mine a new block are chosen according to their stake size and stake age. By altering the initial protocol, other PoS variations are possible. The way the versions avoid concerns with duplicate expenditure and network centralization determines how they differ from one another.



Fig. 5 Scalable consensus protocols

Delegated Proof of Stake (DPoS): In DPoS, the PoS is utilized. In contrast to PoS, which is directed democratic, this method is representational democratic. This indicates that some nodes are chosen by a vote of all parties to serve as witnesses and delegates. Witnesses earn rewards for introducing new blocks and are penalized for doing so.

Practical Byzantine Fault Tolerance (PBFT): The PBFT uses replication among the known parties and can withstand up to a one-third failure rate. Byzantine Generals Problem (BGP), which results in Byzantine fault when an agreement cannot be reached, is resolved by PBFT. The selected leader creates an ordered list of the transactions and broadcasts it for verification.

Proof of Authority (PoA): A version of the PoS is the PoA. To validate the transactions in the block, validators are selected. The network has a high throughput and is extremely scalable with nearly no processing costs because there aren't many validators. Unlike with PoS, a validator simply needs to stake its reputation rather than any of its assets.

3.2.3 Directed Acyclic Graph (DAG)

Block stores transactions that are arranged in a single chain structure on a typical blockchain.

Inclusive [11]: For choosing the primary chain of the created DAG, an inclusive rule is suggested. Additionally, the ledger may also contain the data of off-chain blocks if they don't contradict with earlier blocks. The system can attain a greater throughput with the suggested protocol.

SPECTRE [12]: SPECTRE uses the DAG structure in order to establish the partial order between each pair of blocks, which cannot be extended to a total order across all transactions.

PHANTOM [13]: To accelerate block production and increase transaction throughput, PHANTOM also uses blockDAG. PHANTOM can handle smart contracts and suggests a greedy mechanism to rank transactions encoded in blockDAG.

Conflux: A blockchain system based on DAG called Conflux is quick and scalable. They suggested two distinct types of block edges in Conflux (parent and reference edges). A selection algorithm is used to choose a pivot chain made up of parent edges.

Dagcoin [14]: Each transaction is treated as a block in the DAG-based cryptocurrency Dagcoin, which prioritizes quicker security confirmations and higher throughput.

IOTA [15]: IOTA does not use blocks, miners, or transaction fees. After completing a particular computing activity, each node is free to make transactions and select two earlier transactions for validation and approval.

3.2.4 Sharding

Large commercial databases were the main focus of the early introduction of the traditional technology known as “Sharding” in the database business.

Elastico [16]: The first permissionless blockchain sharding mechanism is called *Elastico*. Participants in *Elastico* must complete a PoW puzzle to determine the consensus committee for each consensus epoch. Each committee operates PBFT as a shard to achieve consensus, and the outcome is approved by a leader committee, which is in charge of making the final judgments regarding the consensus outcomes of other shards. To update additional shards, the final value will then be transmitted back.

OmniLedger [17]: *Elastico*'s issues are attempted to be fixed by the more modern distributed ledger *OmniLedger*, which is based on the Sharding approach. For the shard assignment, it employs a bias-resistant public-randomness protocol that combines *RandHound* and *Algorand*. *Atomix*, a two-phase client-driven “lock/unlock” method, is introduced by *OmniLedger* to guarantee the atomicity of cross-shard transactions. *OmniLedger* also uses the blockDAG data structure to implement block commitment in parallel and improve transaction efficiency with Trust-but-Verify Validation.

RapidChain [18]: *RapidChain* is a Sharding-based public blockchain technology that is more resistant to Byzantine errors than *OmniLedger* is, up to a 1/3 fraction of participants. According to *RapidChain*, earlier Sharding-based protocols' transaction speed and latency are significantly hampered by the communication overhead associated with each transaction.

Monoxide [19]: Asynchronous Consensus Zones are proposed by the scale-out blockchain known as *Monoxide*, which also maintains the system's security and decentralization to a large extent. The entire *Monoxide* network is divided into numerous parallel zones, each of which is entirely responsible for itself.

Table 3 shows the advantages and disadvantages of Layer 1 scalable techniques.

3.3 Layer 2—Approaches off the Blockchain (Off-Chain)

It is used to handle the issues like security and scalability in the blockchain. It stores the final data of settlement between the parties in the blockchain.

3.3.1 Payment Channel

One of the main scaling objectives has been to enable blockchains to facilitate (micro-) payments with almost instantaneous confirmation, fewer on-chain transactions, and lower fees. For applications that are particular to payments, payment channels customize state channels.

Table 3 The advantages and disadvantages of Layer 1

Scalable techniques	Advantages	Disadvantages
Block data	(1) Number of transactions per block is more (2) Solved bitcoin malleability	Limits the increase in throughput
Sharding	(1)The overhead of the communication is less (2) Scalability in storage	(1) 1% attack exists (2) Complexity of the design
DAG	(1) High throughput and less confirmation time (2) Creation of blocks is parallel	(1) Consistency is weak (2) Issues in Security
Consensus protocol	(1) Scalability is high (2) Different options are pluggable	(1) Introduces issues in security (2) Energy consumption is high in PoW

Lightning Network [20]: It was suggested that the Bitcoin Lightning network increase the currency’s throughput. It makes it possible for two or more users to establish a secure channel apart from the primary chain and conduct immediate transactions through it without the need for a middleman.

Raiden-Network: The Ethereum off-chain solution for bi-directional multi-hop transfer is known as Raiden-Network. The Raiden Network is similar to the Lightning Network for Bitcoin, with the exception that Lightning Network only supports Bitcoin, whereas the Raiden Network also supports Ethereum’s ERC20 tokens.

Trinity Network: On the NEO blockchain platform, Trinity Network provides a low-cost, high throughput payment channel for rapid transactions. Trinity Network is an all-purpose, open-source off-chain solution. To achieve high throughput, state channel technology is employed.

3.3.2 Sidechain

The side chain is connected to the main chain in order to improve scalability and enable the gradual transfer of main chain assets to the side chain.

Plasma [21]: By employing a smart contract as its base, Vitalik Buterin’s Ethereum Plasma offers sidechains framework connected to Ethereum. To minimize the load on Ethereum and increase its throughput, Plasma is utilized for smart contract transactions instead of Ethereum. The main chain periodically posts the sidechain’s block headers for verification.

ZK- Rollup [22]: Instead of plasma, ZK-rollup is suggested by Vitalik Buterin as a scalability solution for Ethereum. Using ZK-proof, relayers combine multiple transactions transmitted to a single transaction known as ZK-SNARK proof, and store it

in the main chain. The weight of a single transaction is less than the transactions that are bundled.

Liquid network [23]: Blockstream, the original developers of the two-way pegged sidechains, built the Liquid Network, a network of Bitcoin sidechains. The main purpose of the sidechain is to swap cryptocurrencies and other digital assets for private and quick Bitcoin transactions (participants include crypto exchanges and dealers).

RootStock [24]: Bitcoin-based smart contracts are operated using Rootstock (RSK). A Federated two-way peg is used by the Bitcoin sidechain. It uses PoW consensus and miners are rewarded for mining. It uses merged mining to offer the same level of security as Bitcoin.

3.3.3 Cross-Chain

Solutions of Crosschain link various blockchains together for scalability and interoperability. The only difference between cross-chain and sidechain building is that cross-chain members are independent and pre-existing blockchains.

Cosmos: Zones make up the Cosmos network of separate blockchains. It links many blockchain systems that are running independently in parallel and engaging with one another for scalability and interoperability. The Cosmos Hub, the first zone, extends to several other zones (blockchains).

Geeq: Another cross-chain blockchain, Geeq, links several Geeq instances (Geeqchains) to facilitate asset exchange, scalability, and cooperation. The evidence of Honesty, a new consensus used by Geeq, is (PoH). Geeq can reach 99% Byzantine fault tolerance thanks to the consensus, as opposed to the lower number in other BFT-based blockchains. Scalability and very speedy finality are further benefits of the blockchain. The connected Geeq instances can effortlessly swap the Geeq token and extend the same genesis blockchain.

Polkadot [25]: Polkadot is a Parachain that provides security, scalability, and interoperability among various blockchains. Relay-Chain architecture is used in Polkadot. A decentralized network called Relay-Chain provides security services and acts as a bridge for several connected blockchains (Para-chains). Relaychain behavior is independent of the internal structure or operations of its Parachain.

3.3.4 Off-Chain Computation

To reduce unnecessary computations and scale the blockchain, just the off-chain node performs the computations rather than each node individually.

TrueBit [26]: Using a verifiable computation mechanism, TrueBit enables complicated smart contract computations to be performed apart from the Ethereum main

Table 4 The advantages and disadvantages of Layer 2

Scalable techniques	Advantages	Disadvantages
Payment channels	(1) Privacy and throughput are high (2) Transaction fees are less	(1) Less secure and limits itself to cryptocurrencies (2) Coins are deposited and will be locked
Off-chain computation	(1) Scalability is better (2) Parallel computation of the tasks	(1) Issues in privacy (2) Issues in security
Sidechains	(1) Interoperability is allowed (2) Parent-chain security issues are not related to child-chain security issues	(1) It is not user friendly (2) The main chain frequently checks child-chain
Cross-chain	(1) Scalability is better (2) Interoperability is allowed	(1) Complexity of design (2) Has privacy issues

chain. It expanded the concept of Ethereum’s compute markets and introduced solvers and oracle-based computation verification.

Arbitrum [27]: Scalability is increased with the off-chain verification of smart contracts provided by the cryptocurrency system called Arbitrum. The participants in a transaction can create a virtual machine using arbitrum and define how it should act while executing a smart contract.

Table 4 shows the advantages and disadvantages of Layer 2 scalable techniques.

4 Conclusion

Blockchain technology has developed quickly over the last few years and will soon be used for more applications in a variety of industries. The number of users has constantly expanded as blockchain technology is being adopted more widely. However, the persistent network congestion issue has compelled people to carefully consider how to address the scalability problem with blockchains. In this paper, we addressed the scalability issues that arise when blockchain technology is used in various applications. We classified the Scalability Solutions in different layers with different Technologies. We discussed various challenges associated with achieving scalability. Layer 1 solutions improve the architecture of the blockchain whereas Layer 2 solutions are used to build third-party blockchain. Layer 1 scalable solutions are used to improve protocols of large scale. Scalability can be enhanced faster in Layer 2 when compared to Layer 1. Layer 2 provides more network security and efficiency than Layer 1. Sidechains and Cross-Chains support interoperability along with scalability. Through this survey, we hope that our categorization of the available solutions may stimulate further academic research aimed at enhancing the scalability of blockchains.

References

1. Li S, Yu M, Yang CS, Avestimehr AS, Kannan S, Viswanath P (2021) PolyShard: coded sharding achieves linearly scaling efficiency and security simultaneously. *IEEE Trans Inf Forensics Secur* 16(c):249–261. <https://doi.org/10.1109/TIFS.2020.3009610>
2. Gao Y, Kawai S, Nobuhara H (2019) Scalable blockchain protocol based on proof of stake and sharding. *J Adv Comput Intell Intell Inform* 23(5):856–863. <https://doi.org/10.20965/jaciii.2019.p0856>
3. Bugday A, Ozsoy A, Öztaner SM, Sever H (2019) Creating consensus group using online learning based reputation in blockchain networks. *Pervasive Mob Comput* 59:101056. <https://doi.org/10.1016/j.pmcj.2019.101056>
4. Klarman U, Basu S, Kuzmanovic A (2018) bloXroute: a scalable trustless blockchain distribution network W. Bloxroute.Com. pp 1–12. [Online]. <https://bloxroute.com/wp-content/uploads/2018/03/bloXroute-whitepaper.pdf>
5. Chawla N, Behrens HW, Tapp D, Boscosovic D, Candan KS (2019) Velocity: scalability improvements in block propagation through rateless erasure coding. In: *ICBC 2019—IEEE international conference on blockchain cryptocurrency*. pp 447–454. <https://doi.org/10.1109/BLOC.2019.8751427>
6. Rohrer E, Tschorsch F (2019) KadCast: a structured approach to broadcast in blockchain networks. *AFT 2019—Proceedings of 1st ACM Conference on Advances in Financial Technologies*. pp 199–213. <https://doi.org/10.1145/3318041.3355469>
7. Ding D, Jiang X, Wang J, Wang H, Zhang X, Sun Y (2019) Txilm: lossy block compression with salted short hashing. pp 1–5. [Online]. <http://arxiv.org/abs/1906.06500>
8. Xu Z, Han S, Chen L (2018) CUB, a consensus unit-based storage scheme for blockchain system. In: *Proceedings of IEEE 34th International Conference on Data Engineering ICDE*. pp. 173–184. <https://doi.org/10.1109/ICDE.2018.00025>
9. Dai X, Xiao J, Yang W, Wang C, Jin H (2019) Jidar: a jigsaw-like data reduction approach without trust assumptions for bitcoin system. In: *Proceedings of the international conference on distributed computing systems*. pp 1317–1326. <https://doi.org/10.1109/ICDCS.2019.00132>
10. Anupama BS, Sunitha NR (2022) Analysis of the consensus protocols used in blockchain networks—an overview. In: *2022 IEEE international conference on data science and information system (ICDSIS)*. pp 1–6. <https://doi.org/10.1109/ICDSIS55133.2022.9915929>
11. Lewenberg Y, Sompolinsky Y, Zohar A (2015) Inclusive block chain protocols. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, vol 8975. pp 528–547. https://doi.org/10.1007/978-3-662-47854-7_33
12. Zohar A, Sompolinsky Y, Lewenberg Y, Zohar A (2017) SPECTRE: serialization of proof-of-work events: confirming transactions via recursive elections. *Medium*. p 66 [Online]. <http://diyhlpl.us/~bryan/papers2/bitcoin/Serialization%20of%20proof-of-work%20events:%20Confirming%20transactions%20via%20recursive%20elections%20-%202016.pdf>
13. Sompolinsky Y, Wyborski S, Zohar A (2021) PHANTOM GHOSTDAG: a scalable generalization of Nakamoto consensus: September 2, 2021. In: *AFT 2021 ACM conference on advances in financial technologies*. pp 57–70. <https://doi.org/10.1145/3479722.3480990>
14. Lerner SD (2015) DagCoin Draft
15. Silvano WF, Marcelino R (2020) Iota tangle: a cryptocurrency to communicate internet-of-things data. *Futur Gener Comput Syst* 112:307–319. <https://doi.org/10.1016/j.future.2020.05.047>
16. Loi Luu PS, Narayanan V, Zheng C, Baweja K, Gilbert S (2016) A secure sharding protocol for open blockchains. *ACM* 2(4). <https://doi.org/10.1186/1757-1626-2-6640>
17. Kokoris-Kogias E, Jovanovic P, Gasser L, Gailly N, Ford B (2017) OmniLedger: a secure, scale-out, decentralized ledger. *IACR Cryptol ePrint Arch* 406. [Online]. <https://eprint.iacr.org/2017/406>
18. Zamani M, Movahedi M, Raykova M (2018) RapidChain: scaling blockchain via full sharding definity Palo Alto, CA. In: *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. pp 1–38. [Online]. <https://eprint.iacr.org/2018/460.pdf>

19. Wang J, Wang H (2019) Monoxide: scale out blockchain with asynchronous consensus zones. In: Proceedings of 16th USENIX symposium on networked systems design implementation, NSDI. pp 95–112
20. Poon J, Dryja T (2016) The bitcoin lightning network: scalable off-chain instant payments. *Percept Psychophys* 18(3):205–208. <https://doi.org/10.3758/BF03205969>
21. Poon J, Buterin V (2017) Plasma: scalable autonomous smart contracts. Whitepaper. pp 1–47. [Online]. <https://plasma.io/plasma.pdf>
22. Lavour T, Lacan J, Chanel CPC (2022) Enabling blockchain Services for IoE with Zk-Rollups. *Sensors* 22(17):6493. <https://doi.org/10.3390/s22176493>
23. Nick J, Poelstra A, Sanders G (2020) Liquid: a bitcoin sidechain
24. Lerner SD (2019) RSK—RootStock platform. White Pap. pp 1–20. [Online]. <https://www.rsk.co/Whitepapers/RSK-White-Paper-Updated.pdf>
25. Wood G (2017) Polkadot white paper. pp 1–21. [Online]. <https://polkadot.network/PolkaDotPaper.pdf>
26. Teutsch J, Reitwießner C (2019) A scalable verification solution for blockchains. pp 1–50. [Online]. <http://arxiv.org/abs/1908.04756>
27. Kalodner H, Goldfeder S, Chen X, Weinberg SM, Felten EW (2018) Arbitrum: scalable, private smart contracts. In: Proceedings of 27th USENIX Security Symposium. pp 1353–1370

Anti-money Laundering Analytics on the Bitcoin Transactions



Rajendra Hegadi, Bhavya Tripathi, S. Namratha, Aqtar Parveez, Animesh Chaturvedi, M. Hariprasad, and P. Priyanga

Abstract Bitcoin is a popular cryptocurrency widely used for cross-border transactions. Anonymity, immutability, and decentralization are the major features of Bitcoin. However, criminals have taken advantage of these very features, resulting in the rise of illegal and fraudulent activities using the innovative technology of blockchain. This paper investigates the behavioral patterns of illicit transactions in the Bitcoin dataset and applies Machine Learning (ML) techniques to see how well they detect these transactions. The aim is to provide an insight into how ML techniques can support the proposed *Anti-Money Laundering Analytics* on the *Bitcoin Transactions*. The motivation behind this work stems from the recent COVID-19 pandemic, which saw a significant spike in various cybercrimes, particularly cybercrimes involving cryptocurrencies.

Keywords Cryptocurrency · Bitcoin · Cybercrime · Blockchain · Artificial intelligence · Machine learning · Money laundering · Crypto crime

R. Hegadi (✉) · B. Tripathi · S. Namratha · A. Parveez · A. Chaturvedi · M. Hariprasad
Indian Institute of Information Technology Dharwad, Dharwad, Karnataka, India
e-mail: rajendra.hegadi@gmail.com

B. Tripathi
e-mail: 18bcs019@iiitdwd.ac.in

S. Namratha
e-mail: 18bcs083@iiitdwd.ac.in

A. Parveez
e-mail: 18bcs010@iiitdwd.ac.in

A. Chaturvedi
e-mail: animesh@iiitdwd.ac.in

M. Hariprasad
e-mail: hariprasad@iiitdwd.ac.in

P. Priyanga
Global Academy of Technology, Bangalore, Karnataka, India

1 Introduction

Money laundering is one of the many ways for criminals to clean their illegitimate gains. It is the process of converting large amounts of money earned from criminal activity—such as drug trafficking and terrorism—into legitimate funds. Transactions using cryptocurrencies have become popular as they provide a safe way to money laundering.

Initially proposed by Satoshi Nakamoto [1], Bitcoin has become the most prominent and valuable cryptocurrency, with a market cap upwards of 865 billion USD as of March 2022. This has led Bitcoin to become a high-value target for hackers and scammers, thus leading to a record rise of bitcoin-associated cybercrimes such as darknet transactions for illegal goods, Ponzi schemes, ransomware attacks, blackmail exploits, denial-of-service (DOS) attacks, phishing, money laundering, and unlawful gambling. This has caused Bitcoin to be viewed in a negative light by the media and various governments leading to fear among the common people in regard to the whole Blockchain technology.

In 2022, the crypto crime trend that showed the total volume of cryptocurrency usage has seen a quantum jump of 567% in 2021 compared to that in the year 2020 [2]. A substantial amount of these transactions are by illicit means. The two main categories of these illicit transactions are stolen funds and scams through Decentralized-Finance-related thefts. The report mentions that there is an increase of 82% in scamming revenue in 2021 and a 516% increase in the cryptocurrency that was stolen in 2021.

Generating transactions is how Bitcoin payments and transfers are made. Transactions are carried by using Bitcoin addresses. Bitcoin transactions can be traced more easily when the same addresses are used repeatedly; therefore, offenders often use disposable addresses. The data (<https://bitaps.com/>) shows among over 1 billion addresses in Bitcoin, 95% of them are empty.

This paper describes *Anti-Money Laundering Analytics* on the *Elliptic Bitcoin Dataset* [3, 4], and the paper is divided into the following two categories of Elliptic Bitcoin Dataset analytics:

- **Analyzing the Elliptic Dataset:** Ismail Alarab et al. [5] worked on the importance-sampling for balancing the data. Oliveira et al. [6] used random walk-on transaction graphs and added extra features based on hitting an illicit node, thereby showing the improvement in the detection.
- **Employing machine learning on the Elliptic Dataset:** Ismail Alarab et al. produced a sequence of works in this direction. Alarab et al. [7] implemented several supervised learning algorithms and improved the detection using ensemble learning. Alarab et al. [8] also presented that graph-based convolutional networks were used with linear node embedding. Further, Alarab et al. [9] used graph-based Long Short-Term Memory (LSTM) and Convolutional Neural Nets (CNN) combined with Monte Carlo dropouts.

2 Literature Survey

This section presents current research works on ML techniques for Anti-Money Laundering solutions. Due to the intricacy of Bitcoin data, several approaches for investigating various activities on the Bitcoin ledgers have been intensively researched. Dealing with many nodes (transactions) and edges (flow of Bitcoins) in big data is difficult, especially when the goal is to identify unlawful services in the network for further investigation. As technology advances, criminals find different ways to exploit it for illicit services. It highlights the numerous ways criminals utilize Bitcoin to launder money [10].

Using intelligent machine learning algorithms to analyze the Bitcoin network based on historical data to spot illegal activities has yielded encouraging results. Weber et al. [4] use Supervised Machine Learning methods to predict the type of a yet-to-be-identified entity in the Bitcoin blockchain network. This study has prepared the path for future studies in which more data and different classification algorithms may yield better results.

Currently, we are employing Elliptic data in our experiment, which is based on earlier research [4]. This information has been made public and labeled as licit or illicit node transactions on the Bitcoin network. The fundamental contribution of this data collection in the Anti-Money Laundering use case was supplied by the original study in [4], which used multiple Machine Learning approaches to forecasting licit or illicit transactions on the Elliptic Dataset. Depending on the features, the Elliptic Dataset [4] is divided into three groups: The first 94 features in the Local Features (LF) group, all 166 features in the All Features (AF) group, and the third group is all features concatenated with node embedding features acquired from the Graph Convolutional Network algorithm.

Data models are built using classical ML Algorithms like Logistic Regression, Multi-Layer Perceptron, Random Forest, and Graph Convolutional Networks (GCN) on the dataset, and results are compared across a wide range of feature combinations. Random Forest eventually outperformed all these algorithms.

As of today, there are approximately 450 articles on Google Scholar about predicting illicit transactions in the Elliptic Dataset. Interested readers are directed at a survey by Jensen et al. [11] and the references therein.

2.1 Bitcoin Exchange

Like a fiat currency exchange, a bitcoin exchange is an online Bitcoin-to-currency exchange where anyone can buy and sell Bitcoins. Different fiat currencies or altcoins are used for trading Bitcoins. This exchange acts as an intermediary between buyers and sellers of the cryptocurrency. Some exchanges operate similar to banks, offering interest on customer savings. Like banks, users can own one or more than one wallet in the exchange. The wallets are used to sell or acquire Bitcoins. However, there

is still a significant risk of Bitcoin exchanges being hacked. Conversely, Marella [12] recommends that exchanges explicitly disclose all cyber-attack data to their customers. As a result, their operations will be more transparent. Mt. Gox attack resulted in a loss of 450 million dollars, Bitfinex attack resulted in a 23% drop in the value of Bitcoin, and DDoS attack on big exchanges has resulted in the loss of thousands of Bitcoins.

2.2 *Illegal Activities on Bitcoin*

Some of the illegal activities associated with Bitcoin transactions:

- hire murderers, terror funding
- drugs, weapons, and human organ trafficking
- Ponzi scams, signature forgeries, and illegal gambling
- money laundering, illegal mining
- information system hacking, ransomware attacks, and plain theft.

There are more than a billion Bitcoin addresses (wallets) controlled by approximately 24 million users, most of which use Bitcoin for criminal activities. According to the study of Bohr and Bashir [13], at least 6 million (25%) of Bitcoin users and roughly 300 million (44%) of Bitcoin transactions are involved in criminal activities. According to another study by Foley et al. [14], exactly 50% of all Bitcoin transactions are illegal. Illicit users prefer to transact in small quantities frequently with the same party to avoid being detected.

Bitcoin seizure incidents have forced illegal users to sell off their Bitcoins quickly. Users who spend Bitcoin on illicit purchases have between 25% and 45% more Bitcoin than users who do not spend Bitcoins on illegal goods. As a result, criminal acts must be investigated as soon as possible. The following is a study of some of the most serious illicit actions involving Bitcoin.

Ransomware is a computer virus similar to trojan horses, worms, and spyware. Because of its rapid rise on a global scale, ransomware has become a big problem. It exploits the easy way to collect ransom through Bitcoin systems. Hackers break into a system and encrypt the files on the host, threatening to divulge any sensitive or secret data the users may have or never decrypt the data unless they pay a Bitcoin ransom within a particular amount of time. The hacker then forces demands on the victim, who could be a company or an individual.

Theft: One-third of the loss in the Bitcoin network is due to software vulnerability and network-based attacks. Cyber-attacks against Bitcoin wallets might result from system security weaknesses, Bitcoin users' faults such as neglect or ignorance, or a DoS assault. DOS attack has two patterns: a small amount of Bitcoin transactions to the repeated addresses and transaction rate attacks producing parasitic worm structures. The transactions cannot be reversed. It is advantageous to criminals because it is impossible to fix theft-related faults in a public ledger. As a result, funds can be

stolen or taken without the consent of Bitcoin owners. Some of the many examples include the hacking of Binance (2019), KuCoin (2020), Mt. Gox (2011), Bitfinex (2016), Coincheck (2018), etc. [15].

Scam: The recent flurry of interest in cryptocurrencies has piqued the interest of a wide range of investors, but it has also piqued the interest of scammers. The most common goal of crypto scams is to obtain private information such as security codes or to dupe an unsuspecting user into sending cryptocurrency to a hacked digital wallet. According to Murphy et al. [16], some of the types of scams are social engineering scams, lottery scams, romance scams, imposter and giveaway scams, phishing scams, blackmail and extortion scams, investment or business opportunity scams, and new crypto-based opportunities: ICOs and NFTs scams, De-Fi rug pulls, cloud mining scams, etc.

Darknet: The term “ark-net” refers to an encrypted Internet network that can only be accessed through special browsers like the Tor browser. According to Weimann [17], approximately 57% of darknet content is illegal, and approximately 47% of all Bitcoin transactions take place on the darknet [14]. This indicates a deep association of illegal content with Bitcoin transactions on the darknet. The Darknet consists of 90% of the Internet and is not accessible by normal browsers. This provides anonymity to the users leading to exploitation. Bitcoin has become the go-to payment mode for illegal services.

Money Laundering exploits the loopholes in a legal process to convert money obtained from illegal sources like terror funding, drug traffic, etc. to funding that appears to have come from legitimate sources [18]. Online banking and cryptocurrencies have made it easier for criminals to commit transactions without detection. There is an internationally coordinated effort to prevent money laundering, especially targeting terrorist funding.

2.3 *Tainted Bitcoin*

Tainted Bitcoin is a term that implies that Bitcoin or cryptocurrency linked to illegal activities is “ilthy” and will remain thus permanently. As a result, there is concern that individuals may obtain tainted Bitcoin through no fault of their own, and as a result, their funds may be seized, accounts closed, or, at the very least, their accounts will be scrutinized more closely. In actuality, however, it is claimed that all bitcoins have been tainted at some point, with the exception of newly mined bitcoins.

3 Anti-money Laundering Analytics on a Bitcoin Dataset

In this section, we use Elliptic Bitcoin Dataset [3, 4] released by Elliptic (a company) that provides Anti-Money Laundering solutions and various crime-detecting solutions in cryptocurrencies. Then, we analyze the Bitcoin Transactions by applying

the Machine Learning (ML) models to the dataset. In the end, we provide a way of reasoning “how to compare and select a better ML model”.

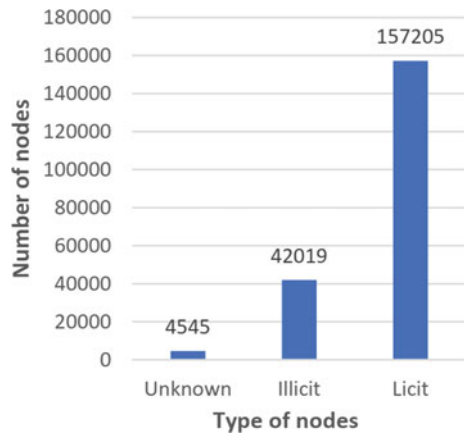
3.1 *Elliptic Bitcoin Dataset*

The Elliptic Dataset contains Bitcoin transactions which have entities representing Licit categories versus Illicit categories. The Licit categories involve miners, licit services, exchanges, wallet providers, etc. The Illicit categories involve ransomware, Ponzi schemes, scams, malware, terrorist organizations, etc. The transactions in Bitcoin are collected from anonymized data, which represents a graph with nodes (as transactions) and edges (as the flows of Bitcoins between two transactions). Each node contains 166 characteristics and has been classified as “licit,” “illicit,” or “unknown” as shown in Fig. 1. In the Bitcoin transaction graph, 203,769 nodes are transactions and 234,355 edges are the flow of Bitcoins. Class 1 is labeled with 2% (4,545) of the transactions (nodes) classified as illicit transactions. Class 2 has 21% (42,019) of the nodes as licit transactions. The remaining transactions are without labeling (i.e. unlabeled) because they are “unknown,” whether they are legal or illegal.

There are 166 features associated with each node, which has a time step to represent the amount of time taken to do a transaction broadcast in the Bitcoin network. The range of the time step is from 1 to 49, which are uniform for a two-week interval. Figure 2 shows the number of transactions in each time step.

Out of 166 features, the initial 94 features represent the information about the specific transaction, which includes the time steps, the transaction fee, the output volume, the number of inputs/outputs, an average number BTC received (spent), and an average number of incoming/outgoing transactions. The remaining 72 features are aggregated features generated by one-hop backward/forward processing of trans-

Fig. 1 Distribution of different classes in the dataset



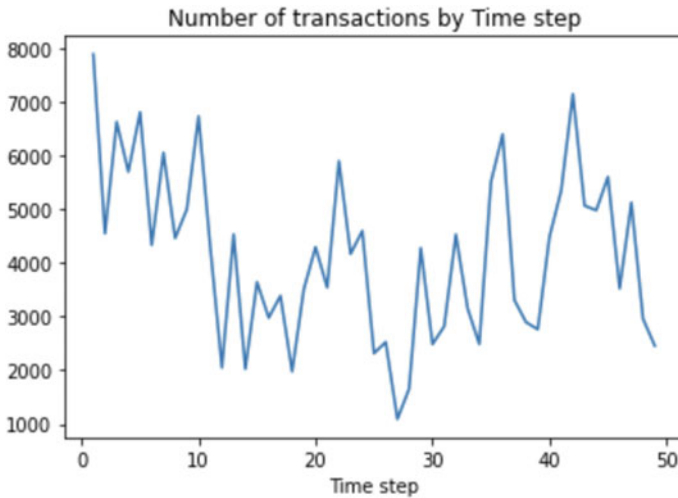


Fig. 2 Number of transactions per time step

action information from the center node, yielding the neighbor’s standard deviation, minimum, maximum, and correlation coefficients.

Deanonimization of the dataset: The Elliptic Dataset provides transactions that are anonymous in nature [19] deanonymized approximately 99.5% of these transactions, thereby disclosing the transaction hashes which can then be analyzed on Blockchain Explorer [20] or any other websites that provide transaction information from transaction hashes like [OXT.me](https://oxt.me).

3.2 Analysis of the Bitcoin Transactions

After reviewing prior reports of various illicit transactions such as hacker assaults, ransomware attacks, money laundering, and other illicit transactions, it was discovered that the illicit transactions follow several fundamental characteristics, making it difficult to trace the origins of the tainted bitcoin.

OXT—OXT.me is a blockchain explorer, visualization tool, and analysis platform owned and operated by Samurai Wallet. For years, OXT has been an undervalued asset for the ecosystem of Bitcoin. They have made it public in a free and open manner, giving expert analysis with the help of visualization and exploration tools. The OXT helps many of the researchers, organizations, and law enforcement agencies. OXT is proven helpful to make open-source actionable tools and detect: spam attacks and transaction denial on the blockchain network.

Mixers/Tumblr Services—A cryptocurrency tumbler is a Cryptocurrency Mixing Service, which combines potentially traceable (or “tainted”) cryptocurrency

assets with others in order to disguise the fund’s original source. This is typically accomplished by combining source funds from numerous sources, which are then spit into destination addresses. It’s tough to track specific coins because all the cash is pooled together and then given at random intervals. Tumblers were created for better bitcoin cryptocurrencies’ anonymity (thus also named as Bitcoin mixer), by storing all transactions on a public ledger. Tumblers have been used to launder cryptocurrencies due to their purpose of anonymity.

Fundamental transaction patterns (or methods) seen in illicit transactions are listed below.

A. Direct Exchange method: The direct method, as shown in Fig. 3, is separating tainted bitcoin into several new wallets and then using an exchange service to convert the tainted bitcoin into fiat currency before the wallet is marked. This method has become outdated because modern security measures would not approve this transaction.

B. Tree Structure method As shown in Fig. 4, the initial tainted coin is split and transferred into several new wallets with an approximately equal ratio. And the contents of the subsequent wallets are further split into newly created wallets. Repeating this for numerous iterations makes it difficult to keep track of all the wallets. Then, the bitcoins are converted into fiat currency using an exchange service before the wallets are flagged. This strategy works because a huge sum of bitcoins is split down into small amounts, making the transaction non-suspicious.

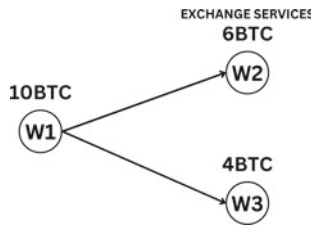


Fig. 3 Direct exchange method

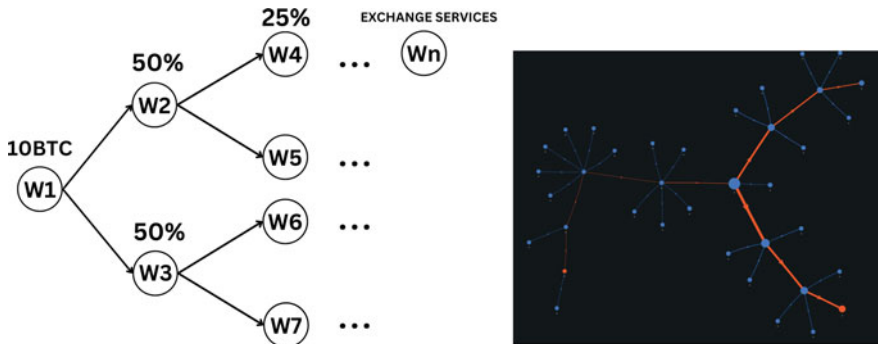


Fig. 4 Tree Structure method observed on an Elliptic transaction using OXT.me

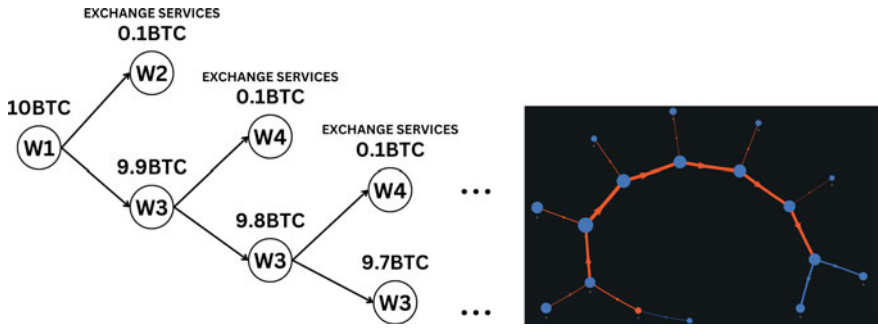


Fig. 5 Peeling method observed on a transaction of Elliptic Dataset on OXT.me

C. Peeling method: The initial tainted bitcoin is broken down into a large amount and a very small amount is transferred into new wallets, as shown in Fig. 5. The subsequent large amount is further broken down into a large part and a very small part, and this is repeated repeatedly until the large number of bitcoins is completely broken down into a very small amount by shaving off a very small amount one step at a time, hence naming it as the Peeling method. The small quantity of bitcoin that is peeled off is moved to a new wallet, and then the bitcoins are attempted to be converted for fiat currency through an exchange service following a series of transactions. This strategy is incredibly successful and extensively utilized because keeping track of Bitcoins becomes increasingly difficult with each successive step. As a result, even if bitcoin can be traced back to its source, it will take a long time to do so, one step at a time.

D. Coin convert method: Due to the countless improvements and innovations in blockchain and related services in recent years, cryptocurrency exchanges such as [Binance](#) have begun to offer coin conversion services, allowing users to convert one cryptocurrency into another and vice versa. Though these services provide customers with flexibility and convenience, they have also provided the opportunity for criminals to exploit these services through their ingenuity. Criminals take advantage of these services and split their tainted bitcoins into separate wallets, from which they can exchange them for Ethereum, Binance coin, Ripple coin, and other coins, which they can further change into other coins and then exchange for fiat currency at any moment or sent to a destination wallet as shown in Fig. 7.

To summarize: Criminals may employ a combination of these fundamental patterns (or methods) to make it incredibly hard for anybody to track the origins of bitcoins, successfully masquerade them as a non-tainted bitcoin, and escape prosecution. Criminals have started using these strategies (or methods) to disguise their transactions from the public eye because each time they convert, they jump onto a different blockchain. These findings show that illegal transactions have some distinguishing characteristics, such as the following. First, the wallets used are frequently anonymous and designed to be used only once. To substantiate it, it is observed that Bitcoin has 95% of empty wallets. Second, these transactions take place in a

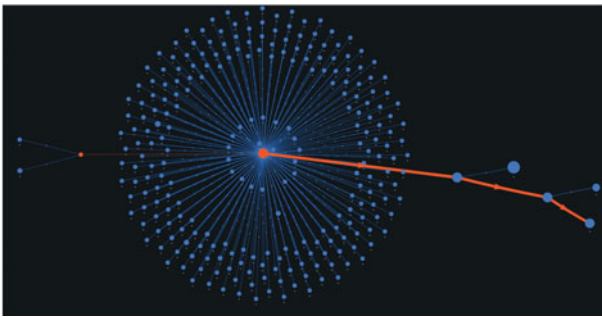
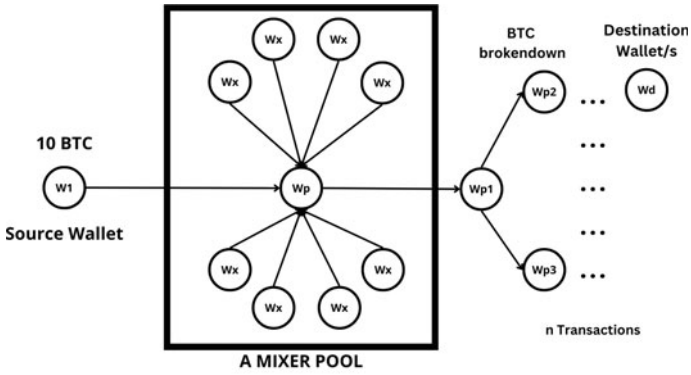


Fig. 6 Mixer method observed in Elliptic Dataset using OXT.me

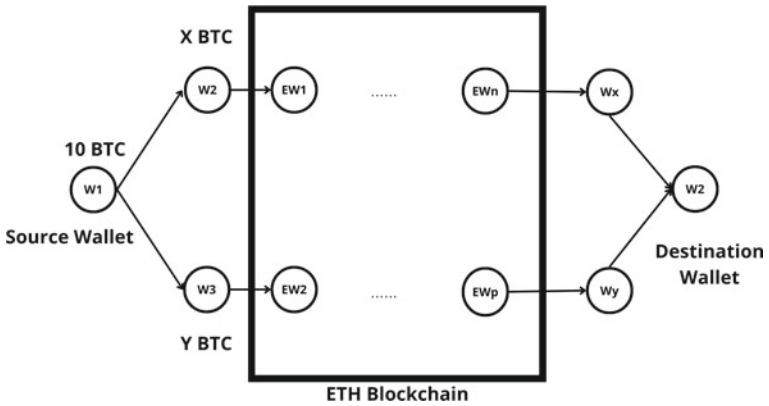


Fig. 7 Coin convert method

relatively short period of time. Third, the tainted coins are usually in transit and are traded as quickly as possible. Fourth, nowadays, tainted coins are more commonly moved through various mixer services and then converted to other cryptocurrencies before being withdrawn or spent, making it incredibly difficult to track the tainted coin.

3.3 Machine Learning of the Bitcoin Transactions

We discuss various machine learning algorithms and show results obtained on the dataset. Given that the Elliptic Dataset is semi-labeled, we used *supervised learning algorithms* to classify the data into illicit and licit transactions. We split the data into training (the first 34 time steps) and test datasets (the last 15 time steps).

We built models using standard classification models—*Logistic Regression, Decision Trees, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), XGBoost, and LightGBM*—to analyze our dataset and make accurate licit and illicit predictions. We used all 166 features (as **All Features**) as well as the first 94 features (as **Local Features**) separately to evaluate these models. Scikit-learn library for Python was used to calculate precision, recall, F1-score, micro-avg F1-score, cross-validation score, and execution time taken by each model.

Logistic Regression: We used binary classification where the dependent variable has two possible types, either licit or illicit transactions. The default parameters from the Scikit-learn package [21] were used, and results were recorded.

Decision Trees is a tree-structured classifier and handles decision-making automatically. Internal nodes are dataset attributes, branches represent decision rules, and each leaf node represents the result. It bisects the space into smaller regions. The decision tree is prone to overfitting. We set the default parameters of the Scikit-learn package [21] to record the results (Tables 1 and 2).

Random Forest: The Random Forest classifier collects predictions from all decision trees on different subsets of the dataset. Majority votes decide the final prediction. The default parameters from the Scikit-learn package [21] were used, and results were recorded.

Support Vector Machine (SVM): SVM uses each data item as a point in the n -dimensional space of features. The classes are identified by locating hyper planes. The default parameters from the Scikit-learn package [21] were used, and results were recorded.

K-Nearest Neighbor (KNN) is a supervised classification algorithm where a new data point is categorized based on its similarity to the data points in its immediate vicinity. There is no explicit training stage in the KNN algorithm, and all the work is done during prediction. We set the Scikit-learn package [21] with default parameters to record results.

XGBoost (Extreme Gradient Boosting) is scalable, which provides a parallel tree boosting. The basic technique can execute on clusters of GPUs or even over a network of computers because the library is parallelizable. This enables high-

Table 1 Illicit classification results for all ML models

ML models	Features	Precision	Recall	F1-Score	Micro Avg F1
Logistic regression	All features	0.454	0.633	0.529	0.928
	Local features	0.433	0.653	0.521	0.924
Decision tree	All features	0.502	0.666	0.572	0.937
	Local features	0.619	0.635	0.627	0.952
Random forest	All features	0.983	0.644	0.779	0.977
	Local features	0.911	0.655	0.762	0.974
SVM	All features	0.657	0.638	0.647	0.956
	Local features	0.786	0.623	0.695	0.965
KNN	All features	0.634	0.643	0.618	0.953
	Local features	0.690	0.651	0.670	0.960
XGBoost	All features	0.969	0.645	0.775	0.976
	Local features	0.897	0.637	0.745	0.972
LightGBM	All features	0.985	0.594	0.741	0.974
	Local features	0.904	0.598	0.720	0.971

performance ML tasks to be solved by training on hundreds of millions of training instances. The XGBoost Scikit-learn package for Python [21] is used to record the results.

LightGBM is a gradient-boosting framework that uses tree-based learning algorithms. It has the following advantages—fast training speed, high efficiency, low memory usage, better accuracy, and can handle large datasets. The default parameters of LightGBM classifier [21] are used to record the results.

3.4 Comparison Using the Cross Validation and Computing Time

To evaluate the performance of each model, the Cross-Validation (CV) score has been obtained and analyzed. LightGBM and Random Forest algorithms resulted in the best CV scores. The computing time of any algorithm is important when detecting illegal transactions in real time. Computing Time by each algorithm to complete execution is given below. Logistic Regression takes the least amount of time while SVM takes the most overall. LightGBM takes the minimum time as compared to the other Ensemble Learning techniques used.

Table 2 Illicit transaction classification results—cross-validation (CV) score and computing time (in sec) using all features and local features

ML models	CV score		Computing time (in sec)	
	All features	Local features	All features	Local features
Logistic regression	0.957	0.937	3.39	2.13
Decision tree	0.891	0.975	5.4	2.0
Random forest	0.987	0.985	23.1	16.3
SVM	0.967	0.965	186.3	99.4
KNN	0.964	0.974	13.6	17.9
XGBoost	0.982	0.984	20.3	13.7
LightGBM	0.985	0.988	9.9	8.0

4 Discussions and Conclusions

We presented *Anti-Money Laundering Analytics* on the Bitcoin Transactions. This paper provided a comparative analysis of various supervised learning algorithms to predict illicit transactions in the Bitcoin blockchain network to aid Anti-Money Laundering processes. We found that Boosting algorithms like XGBoost and LightGBM outperformed the traditional algorithms using the local and aggregated features provided in the Elliptic Dataset by evaluating on parameters such as Precision, Recall, F1-score, Cross-Validation score, and Computing Time. We included the Computing Time by each algorithm as an additional evaluation parameter because when it comes to detecting an illegal transaction in real time, the computing time of any algorithm becomes crucial. By analyzing the Computing Time by each algorithm, we concluded that Logistic Regression, Decision Tree, and LightGBM have the quickest computing time. Since the **LightGBM** offers high classification scores with quick computing time, it would be better to use LightGBM in a real-time setting.

In the future, we can use graph-based supervised learning techniques with appropriate pre-processing based on graph structure. Since the Elliptic Dataset has many unknown nodes, the data becomes imbalanced. We can also look into better data processing to balance the dataset. Deep learning models can also be taken into account in future studies.

References

1. Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system. Decent Bus Rev:21260
2. Chainanalysis-Team (2022) Crypto crime trends for 2022: Illicit transaction activity reaches all-time high in value, all-time low in share of all cryptocurrency activity. <https://blog.chainalysis.com/reports/2022-crypto-crime-report-introduction/>
3. Elliptic (2019) Elliptic data set. <https://www.kaggle.com/datasets/ellipticco/elliptic-data-set>

4. Weber M, Domeniconi G, Chen J, Weidele DKI, Bellei C, Robinson T, Leiserson CE (2019) Anti-money laundering in bitcoin: experimenting with graph convolutional networks for financial forensics. Preprint at [arXiv:1908.02591](https://arxiv.org/abs/1908.02591)
5. Alarab I, Prakoonwit S (2022) Effect of data resampling on feature importance in imbalanced blockchain data: comparison studies of resampling techniques. *Data Sci Manag*
6. Oliveira C et al (2021) Guiltywalker: distance to illicit nodes in the bitcoin network. Preprint at [arXiv:2102.05373](https://arxiv.org/abs/2102.05373)
7. Alarab I, Prakoonwit S, Nacer MI (2020) Comparative analysis using supervised learning methods for anti-money laundering in bitcoin, pp 11–17
8. Alarab I, Prakoonwit S, Nacer MI (2020) Competence of graph convolutional networks for anti-money laundering in bitcoin blockchain. In: 5th International conference on machine learning technologies, pp 23–27
9. Alarab I, Prakoonwit S (2022) Graph-based lstm for anti-money laundering: experimenting temporal graph convolutional network with bitcoin data. *Neural Process Lett*:1–19
10. Samsudeen F, Perera H (2021) Behavioral analysis of bitcoin users on illegal transactions. PhD thesis
11. Jensen R, Iosifidis A (2022) Fighting money-laundering with statistics and machine learning: an introduction and review. Preprint at [arXiv:2201.04207](https://arxiv.org/abs/2201.04207)
12. Marella V (2017) Bitcoin: a social movement under attack
13. Bohr J, Bashir M (2014) Who uses bitcoin? an exploration of the bitcoin community. In: 2014 Twelfth annual international conference on privacy, security and trust. IEEE, pp 94–101
14. Foley S, Karlsen JR, Putnins TJ (2019) Sex, drugs, and bitcoin: how much illegal activity is financed through cryptocurrencies? *Rev Financ Stud* 32(5):1798–1853
15. Crystal-Analytics-Team (2021) The 10 biggest crypto exchange hacks in history. <https://crystalblockchain.com/articles/the-10-biggest-crypto-exchange-hacks-in-history/>
16. Murphy C, Vera E, Kvilhaug S (2022) Beware of cryptocurrency scams. <https://www.investopedia.com/articles/forex/042315/beware-these-five-bitcoin-scams.asp>
17. Weimann G (2016) Going dark: terrorism on the dark web. *Stud Conflict Terrorism* 39(3):195–206
18. Chen J, Anderson S, Eichler R (2022) Money laundering. <https://www.investopedia.com/terms/m/moneylaundering.asp>
19. BenZz (2019) Deanonymization of elliptic dataset transactions. <https://habr.com/ru/post/479178/>
20. Bitcoin-Monitoring-Website (2022) Blockchain explorer. <https://www.blockchain.com/explorer>
21. Pedregosa F et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830