

Studies in Computational Intelligence 1065

Hiren Kumar Thakkar
Mayank Swarnkar
Robin Singh Bhadoria *Editors*

Predictive Data Security using AI

Insights and Issues of Blockchain, IoT,
and DevOps

 Springer

Studies in Computational Intelligence

Volume 1065

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

Indexed by SCOPUS, DBLP, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

Hiren Kumar Thakkar · Mayank Swarnkar ·
Robin Singh Bhadoria
Editors

Predictive Data Security using AI

Insights and Issues of Blockchain, IoT,
and DevOps

Editors

Hiren Kumar Thakkar
Department of Computer Science
and Engineering
Pandit Deendayal Energy University
Gandhinagar, Gujarat, India

Mayank Swarnkar
Department of Computer Science
and Engineering
Indian Institute of Technology BHU
Varanasi, Uttar Pradesh, India

Robin Singh Bhadoria
Department of Computer Engineering
and Applications
GLA University
Mathura, Uttar Pradesh, India

ISSN 1860-949X

ISSN 1860-9503 (electronic)

Studies in Computational Intelligence

ISBN 978-981-19-6289-9

ISBN 978-981-19-6290-5 (eBook)

<https://doi.org/10.1007/978-981-19-6290-5>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

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

Preface

In the recent past, due to the affordable and easy access to the Internet, the number of connected devices has grown exponentially. Contrary to a decade back, millions of new devices are connecting each day forming a giant network of Internet of Things (IoT) leading to enormous amounts of data communication across the devices. However, with the convenience of IoT enabled communications, substantial challenges are also emerging in the form of data snooping, data hacking, data phishing, man-in-the-middle attack, denial-of-service attack, etc. Moreover, few data attacks greatly affect the stability of applications by attacking the underlying Machine Learning (ML) and Deep Learning (DL) codes. The rate of the novel data attacks designed to compromise the securities is much faster than the rate of providing the remedial security solutions. Therefore, there is an increasing demand to refer to the existing security solutions and state of the arts in the data security domain. This book attempts to provide a comprehensive review of the existing types of data security attacks and corresponding solutions with discussion on the trends, technology. Moreover, the book focuses on Artificial Intelligence (AI) enabled data security solutions over and above the over-the-shelf conventional data security mechanism to provide the futuristic views. The book covers the domains such as Image forgery detection, security issues in ML and DL, reversible data hiding in healthcare, Blockchain-based authentication, security concerns in cloud, fog, and edge computing, and social media security analysis.

Gandhinagar, India
Varanasi, India
Mathura, India

Hiren Kumar Thakkar
Mayank Swarnkar
Robin Singh Bhadoria

Acknowledgements

We would like to first thank the almighty for providing us the strength to pursue the idea and carry it forward to make a complete book. Our sincere thanks to all the contributors who have provided their valuable time, support, and timely contributions to make a comprehensive book on data security. We would also like to thank all the reviewers for their informative and constructive suggestions to the contributors to improve the chapters. Finally, we would like to thank our Institutions Pandit Deendayal Energy University (PDEU), India, Indian Institute of Technology—Banaras Hindu University (IIT-BHU), India, and GLA University, India for providing all the required resources for drafting, proofreading, and editing the book.

Contents

A Comprehensive Study of Security Aspects in Blockchain	1
Pranav Singh and Sushruta Mishra	
An Exploration Analysis of Social Media Security	25
Shreeja Verma and Sushruta Mishra	
A Pragmatic Analysis of Security Concerns in Cloud, Fog, and Edge Environment	45
Manish Jena, Udayan Das, and Madhabananda Das	
Secure Information and Data Centres: An Exploratory Study	61
Pranav Pant, Kunal Anand, and Djeane Debora Onthoni	
Blockchain-Based Secure E-voting System Using Aadhaar Authentication	89
Ankit Kumar Jain, Sahil Kalra, Karan Kapoor, and Vishal Jangra	
DevOps Tools: Silver Bullet for Software Industry	105
Divya Srivastava, Madhushi Verma, Shashank Sheshar, and Madhuri Gupta	
Robust and Secured Reversible Data Hiding Approach for Medical Image Transmission over Smart Healthcare Environment	119
K. Jyothsna Devi, Priyanka Singh, José Santamaría, and Shrina Patel	
Advancements in Reversible Data Hiding Techniques and Its Applications in Healthcare Sector	133
Buggaveeti Padmaja, Maharana Suraj, and V. M. Manikandan	
Security Issues in Deep Learning	151
Shrina Patel, Parul V. Bakaraniya, Sushruta Mishra, and Priyanka Singh	
CNN-Based Models for Image Forgery Detection	185
Shyam Singh Rajput, Deepak Rai, Deeti Hothrik, Sudhanshu Kumar, and Shubhangi Singh	

Malicious URL Detection Using Machine Learning 199
Mayank Swarnkar, Neha Sharma, and Hiren Kumar Thakkar

About the Editors

Dr. Hiren Kumar Thakkar received his M.Tech in Computer Science and Engineering from IIIT Bhubaneswar, India, in 2012 and a Ph.D. degree from Chang Gung University, Taiwan, in 2018. Later, he worked as a postdoctoral researcher in the Department of Occupation Therapy, Motor Behavioral Research Lab (MBRL), Chang Gung University, Taiwan. Currently, he is an Assistant Professor in the Department of Computer Science and Engineering, Pandit Deendayal Energy University, Gujarat, India. Dr. Thakkar has published several journal research papers in the areas such as optimization, machine learning, and reinforcement learning. He is a member of IEEE.

Dr. Mayank Swarnkar is currently an Assistant Professor in the Department of Computer Science and Engineering at the Indian Institute of Technology (BHU) Varanasi. He completed his Ph.D. from the Indian Institute of Technology Indore in 2019. He completed his M.Tech in Wireless Communication and Computing from the Indian Institute of Information Technology Allahabad Prayagraj in 2013 and B.E. in Information Technology from Jabalpur Engineering College in 2011. He joined IIT(BHU) in 2020. He also worked as Software Engineer in NEC Technologies India during 2013–2014. His primary areas of interest are network and system security. He works mainly in network traffic classification, zero-day attacks, intrusion detection systems, IoT security analysis, network protocol vulnerability analysis, and VoIP spam detection. He has several publications in international journals and conferences. He is a member of IEEE and ACM.

Robin Singh Bhadoria completed his Ph.D. degree from the Indian Institute of Technology Indore in January 2018. He also finished his M.Tech. and B.E. in CSE from different institutions affiliated with RGPV Bhopal in 2011 and 2008, respectively. He has been awarded the University Gold Medal for his M.Tech. Degree at

Vidhan Sabha of Madhya Pradesh in 2011. He has published over 07 edited books and over 29 journal papers, and 21 peer-reviewed international conference papers. He is a member of IEEE (USA), IAENG (Hong-Kong), Internet Society, Virginia (USA), IACSIT (Singapore), and IEI (India).

A Comprehensive Study of Security Aspects in Blockchain



Pranav Singh and Sushruta Mishra

Abstract Knowledge is power, and in this digital age, knowledge is represented by data, making it one of the most valuable assets. With rapidly evolving technology, there are challenges that directly or indirectly threaten the integrity of data, such as cybercrime, privacy concerns, theft, malware, and viruses. The development of Blockchain Technology has helped in the mitigation of some of these problems by safeguarding online data resources. In this chapter, we introduce the concept of blockchain, discuss its structure and features, and understand its operation. The main focus of this chapter is to observe the vulnerabilities of this technology and scrutinize several attacks exploiting them to understand their outcomes. We go over a few security improvements in an attempt to protect from attacks and alleviate the existing threats. In addition, we explore its application and implementation in various fields. We conclude by discussing the major challenges this technology is facing at present and may encounter in the future.

Keywords Blockchain technology · Vulnerabilities of blockchain · Attacks on blockchain · Applications · Future challenges

1 Introduction

Blockchain was established in 2008 by an unknown entity Satoshi Nakamoto as an underlying technology for Bitcoin, for the maintenance of records. Blockchain is defined as a distributed ledger that consists of an ordered series of blocks, generated by cryptography, and linked together sequentially. Every single block carries a hash of the preceding block, a timestamp, and transaction data (Merkle tree). It is an electronic ledger that is duplicated and synchronized across numerous nodes in the network. The underlying principle combines cryptography, peer-to-peer networking, and mathematical analysis of interactions (game theory).

P. Singh · S. Mishra (✉)

Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India

e-mail: sushruta.mishrafcs@kiit.ac.in

Basically, data and information are stored, managed, and shared in this manner among parties and organizations. When one party creates and dispatches a set or block of information, it is verified by a myriad of nodes distributed across the network. After the block's verification, it is added to an immutable chain. Once any data is added, it becomes unalterable and can only be appended [1]. Modifying any single record would break the link with the adjacent block which would then require changing the entire chain composed of millions of blocks. That is somewhat impossible. Blockchain was initially used for Bitcoin and is currently the most admired representation of this technology. Blockchain has gained vast applicability in almost every industry including finance, health, supply chain management, the Internet of things, etc. With the commercialization of this technology, numerous Blockchain applications and platforms came into existence, like Bitcoin and Ethereum. Blockchain can be applied to a variety of fields far beyond bitcoin, which presently overshadows other blockchain categories. Data security and privacy, an emerging field of blockchain, is one such category that is receiving a lot of attention lately [2].

There are three types of blockchains [3]:

1. *Public blockchain*: These are permission-less and everybody can read, send, or receive transactions. Any participant can join in transactions and validate them through a consensus decision-making procedure before being added to the blockchain. They are considered to be fully decentralized blockchains. Bitcoin and Ethereum are public blockchains.
2. *Private blockchain*: These blockchains are restricted where write permissions are strictly confined to a single authority, who is responsible for granting read/write access to only a selected section of participants in the network. It's akin to a centralized system, but it's cryptographically protected as well as cheap. Everybody is not authorized to read, write, audit, or make transactions. In other words, they are permissioned blockchains. Ripple is an example of a private blockchain.
3. *Consortium blockchain*: This blockchain exists between the two extremes of private and public chains. Rather than a single governing authority, multiple designated authorities have write permissions that can administer and check the consensus procedure to approve a block. The read is not open to mass but only to a set of participants in the network, making it partially decentralized. It facilitates quicker transactions and preserves data by providing multiple points of failure [4].

2 Characteristics of Blockchain Technology

The features of blockchain, as described in Fig. 1, are discussed as follows [5]:

- (1) ***Decentralization***: In contrast to conventional database systems, blockchain technology does not rely on a centralized system to authenticate transactions since every single node in the blockchain has its own replica of data. Thus, the control does not reside solely on a single server or system.

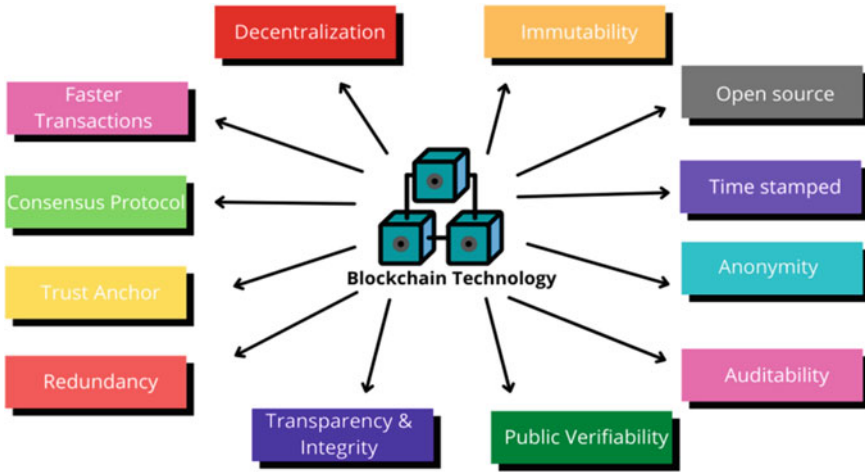


Fig. 1 Core attributes of blockchain technology

- (2) **Immutability:** Every block carries the block header, Merkle root, the hash of the preceding block, and transaction details which are altogether hashed using a famous hashing algorithm like SHA-256 (Secure Hash Algorithm 256). This unique hash value is difficult to reverse engineer, and even a small change in input completely changes the output. The cryptographic hash is difficult to generate but is easily verifiable by miners. As hashing incorporates meta-data from the previous block, it sets up a connection between the current and previous nodes and the cycle continues as new blocks are chronologically added to the chain. Thus, if an attacker attempts to alter one block, its hash value will change too thereby breaking the chain. In order to restore all the subsequent blocks, each hash value will need to be recalculated which requires tremendous computational power.
- (3) **Anonymity:** Users interacting with blockchains are assigned with public and private keys. A public key is an identifier that is used to manage and verify the identity of a user and can be shared freely. On the other hand, the private key is like a password that must never be shared. A person’s real identity is concealed, making transactions anonymous. In addition, a user’s identity can be either anonymous or pseudonymous.
- (4) **Time-stamped:** Each block stores the date and time at which it was mined and successfully added by the blockchain network.
- (5) **Auditability:** Prior to a transaction getting incorporated into the blockchain, it should be verified by the node. The cryptocurrency should actually be owned by the spender and he/she should possess sufficient balance to carry out any given transaction. Thus, transactions rely on previous unspent transactions. This makes the verification and tracking process easy.

- (6) **Public Verifiability:** Each node in the network validates its own copy of transaction data through a general consensus protocol. Therefore, it can be verified by anyone.
- (7) **Transparency & Integrity:** Blockchain data is updated and synced across nodes for public validation. Every node has a separate copy of the data. The digital record is public and transparent to each node. This system enables public verification so anybody can certify it while maintaining its integrity.
- (8) **Redundancy:** Blockchain technology is based on a decentralized architecture which means each node holds data which is consistent across all nodes in the network. In contrast, centralized systems rely on backups and physical servers to get the stored information giving rise to redundant data.
- (9) **Trust Anchor:** Providing read and write access to a system is the responsibility of the trust anchor. They control, access, and grant permissions.
- (10) **Consensus Protocol:** Blockchain eliminates the need for intermediaries. As there is no central authority to validate the transactions, it is important to reach a consensus among the untrustworthy nodes. Nodes cannot trust each other in order to identify illegal and invalid transactions. Hence to maintain the consistency through all the nodes, they settle on an agreed consensus protocol thus ensuring accountability. Consensus mechanisms resolve this problem of trust between corrupt nodes. New blocks are appended to the chain when the majority of the nodes verify the transactions unanimously.
- (11) **Faster transactions:** With blockchain technology, transactions are completed within a matter of few minutes thus conserving time. Traditional payment methods involve various paperwork which takes ample time for approval.
- (12) **Open Source:** Most blockchain systems are open source which means everyone can use blockchain technologies to build their own applications.

3 Working of Blockchain

To fully comprehend the vulnerabilities of blockchain, it is essential to get acquainted with its working [5]. Before we dive deeper into this subject, we examine a few fundamental key components of a blockchain.

Node: It is simply a computer system that preserves replicated copies of the database as well as stores details associated with payment and ownership. There are several nodes depending on the level of participation and type of blockchain network. Full nodes work separately and inspect all the rules of the system. Lightweight nodes only download the headers of blocks and confirm the transactions through SPV (Simple Payment Verification).

Block: Similar to a record in a public ledger, a block stores the information of all the transactions happening over a time period. These blocks are connected to each other via a hash pointer, which points to its predecessor's data. The very first block in the starting of each chain has no parent block and is called the genesis block. A

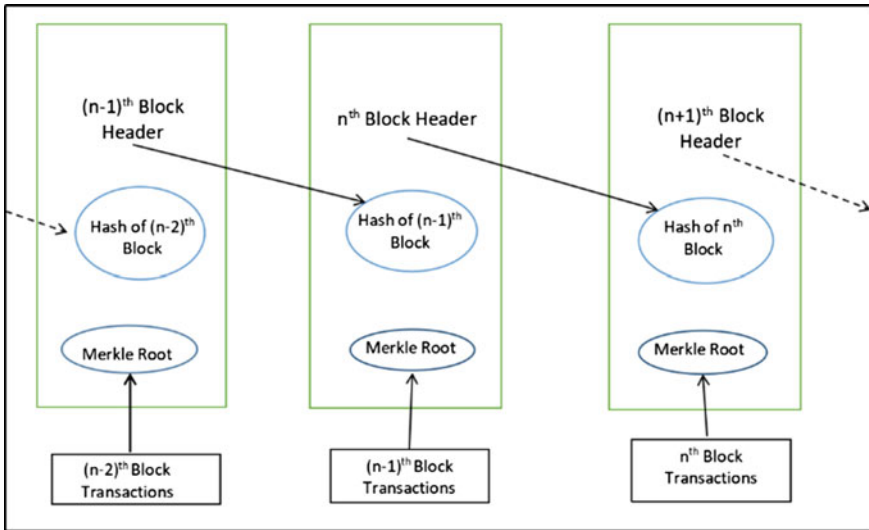


Fig. 2 Structure of a block

block is divided into two components: block header and the block body. The block header contains metadata such as block version, Merkle tree root hash, nBits, parent block hash, timestamp, and nonce. Transactions make up the block body along with a transaction counter. The typical structure of a block is shown in Fig. 2.

Transaction: Transactions are data structures that represent the exchange of digital currency between a sender and receiver over a blockchain network. Every transaction is kept in an invalidated transaction pool and distributed in the network by applying a flooding protocol called the Gossip protocol. Miners typically prefer transactions with a higher transaction fee [6].

Miner: Mining is the most important feature of blockchain through which new cryptocurrency is added to circulation. In blockchain, mining is the process of adding new blocks of transactions, which are then authenticated. The peer who uses its computing power to mine a block is called a miner [7].

Hash and Hash function: Hash function is a mathematical algorithm that takes an input and converts it into an output of a specific length. It is collision resistant which means that it is quite challenging to create the input data again from the hash value alone. It is associated with the immutability feature of blockchain, discussed earlier.

Consensus mechanism: To maintain data consistency, it is necessary to reach consensus among the untrustworthy nodes through a set of predefined rules known as consensus algorithms. Figure 3 shows some of the most common consensus procedures like PoW (Proof of Work), PoS (Proof of Stake), DPoS (Delegated Proof of

CRYPTOCURRENCY	CONSENSUS MECHANISM
Bitcoin	PoW
Litecoin	PoW
Zcash	PoW
Monero	PoW
Ethereum	PoS
Cardano	PoS
EOS.IO	DPoS

Fig. 3 Consensus mechanisms used by well-known cryptocurrencies

Stake), PBFT (Practical Byzantine Fault Tolerance), etc. used by popular cryptocurrencies to solve the Byzantine Generals Problem. Each consensus mechanism has its own merits and demerits.

Digital signature: By using a cryptographic algorithm, a digital signature certifies if data is legitimate or not. Each participant has two types of keys. One is public, which represents transactions and is openly visible to everyone, so anyone in the network can decrypt the transaction. As for the private key, it is used to digitally sign the transactions and prove ownership. The signature is composed of 256 bits which means any attacker needs to apply 2256 permutations to fake it, which is simply a waste of resources [8].

How Does a Transaction Work in Blockchain?

Figure 4 summarizes the entire transaction process in blockchain. A sender issues a transaction and this transaction is added to the unconfirmed transaction pool. The node clubs all the transactions in a given period of time into a block. Before a miner adds the block, he checks its validity by verifying the digital signature and blockchain's history to see if the user has sufficient currency to trigger the said transaction [9]. The miners compete to solve an extremely complicated mathematical puzzle to generate a hash value, whose value should be less than the current target value. Its difficulty is determined by the network and is dynamically set by the system after every 10 min. When a winner emerges, he gets to add the block to the distributed ledger. The successful miner is rewarded with some cryptocurrency for using his resources to create this block. Any transaction fees acquired by the miner are also sent in this transaction. The verified block is then relayed to its peers, who may or may not choose to mine the transaction. All the peers in the network verify the new block using a consensus mechanism. If two blocks have the same parent block, a fork is created. Blockchain protocol deems the longest chain of two branches to be valid. Each miner will have this same blockchain making it consistent across the network [10].

After verification, the block containing the transaction is added to the existing blockchain and is now considered legitimate. The current block links itself with the another newly created block by using a cryptographic hash pointer. In other words,

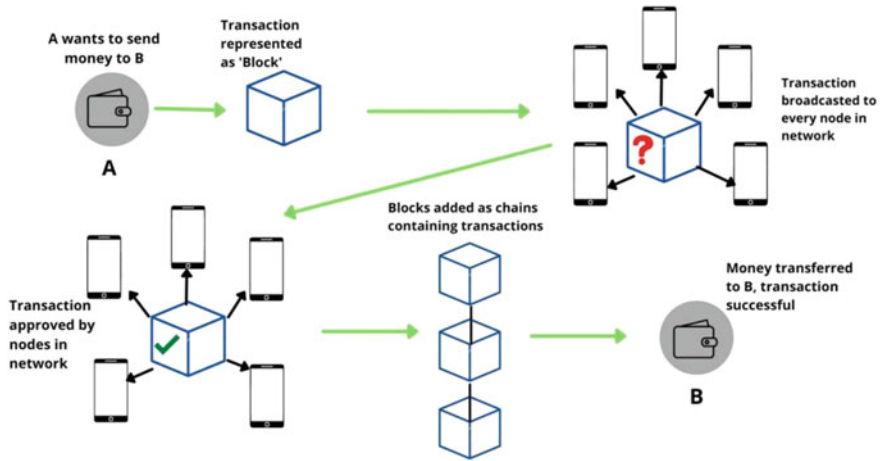


Fig. 4 Roadmap of a transaction in blockchain

this hash pointer points to the data in the parent block. Now the block receives its first confirmation while the transaction obtains the second confirmation. Additionally, with each new block added to the chain, the transaction is reconfirmed.

4 Analysis of Security in Blockchain

Given the risks involved with blockchain, there arises the urgent need to safeguard data and preserve online data resources. Over time security assaults have become more effective and powerful, resulting in old techniques becoming weak or redundant. The protection of important data in the present times can be ascertained with the help of blockchain technology.

The global outbreak of the virus WannaCry in 2017, which demanded ransom payments from the targeted users, made blockchain a hot topic in the world [11]. It caused damage worth millions and billions of dollars by infecting computers worldwide. A snapshot of the ransom message is displayed in Fig. 5.

4.1 Risks to Blockchain

Blockchain security is affected by many factors like the types of attacks, network state, scientific advancement, etc. A malicious actor can abuse this technology’s vulnerabilities and gain unauthorized access for nefarious purposes [12]. Below we discuss some common risks to Blockchain technology:



Fig. 5 Snapshot of WannaCry ransomware attack (Source: wikipedia)

- (1) **51% Attack:** Blockchains are distributed ledgers that use consensus protocols to authenticate transactions and maintain data consistency. In PoW (Proof of Work)-based blockchains, if an individual or a group of miners' computational hashing power is $>50\%$ of the total hashing power of the complete blockchain network, then this attack can be launched. Giving them an edge over honest miners, they can solve the puzzle faster, allowing them to earn undue rewards for completing new blocks. Higher the hashing power faster the attack.

It is possible for an attacker to exploit this vulnerability by posing the following threats:-

- a. The attacker might be able to avoid new transactions from receiving confirmations, stopping them between merchants and clients.
 - b. They can manipulate transactions and reverse them allowing them to spend the same coins many times. (Double Spending)
 - c. Adversely affect the mining power of honest miners.
 - d. Exclude or alter the correct sequence of transactions.
- (2) **Double Spending:** Double spending is the situation in which the consumer utilizes the same cryptocurrency numerous times for carrying out transactions [13]. Before blockchain, it was difficult to ascertain the order in which the transactions arrived in the pool. Transaction A might happen prior to Transaction

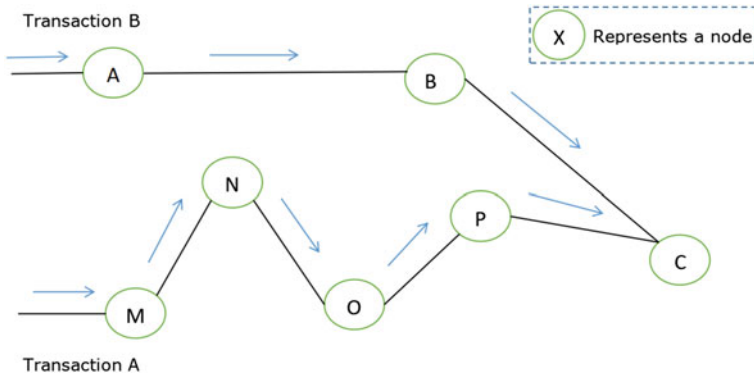


Fig. 6 (Double spending) Transaction B gets recorded due to fewer nodes than Transaction A

B, but the number of nodes Transaction A has to go through could be greater than that of Transaction B. As a consequence, Transaction B gets recorded in a node first. Figure 6 demonstrates this problem in a comprehensible manner. This leads to the problem of double spending and was resolved to some extent with the introduction of blockchain technology. Double spending can be carried out in yet another way.

PoW-based blockchain is more vulnerable to double spending problem since the attacker can utilize the transitional time between a transaction’s initiation and verification. We illustrate this problem using the diagram shown in Fig. 7. We assume that an attacker performs two transactions. First transaction X is sent to the merchant’s address and second transaction Y to a colluding address, which is owned by the attacker himself.

Cryptocurrency can be doubly spent provided that the following requirements are fulfilled:

- (a) X is added to the merchant’s account.
- (b) Y is mined valid by the blockchain network.
- (c) Before any anomaly can be detected, the merchant transacts the output to the attacker.

Transaction X is recognized as an invalid transaction eventually, and the attack becomes successful. The attacker has received the merchant’s output while still owning the cryptocurrency resulting in double spending. The attacker revels in the service despite not spending any currency.

- (3) **Private key security:** Unique private and public keys are assigned to each client of the blockchain. The public key acts as the address which allows transactions to be received or sent. The private key is like a password used to access the transaction output (cryptocurrency). It also acts like a proof of ownership. As opposed to public keys, a person can only have one private key. This key is

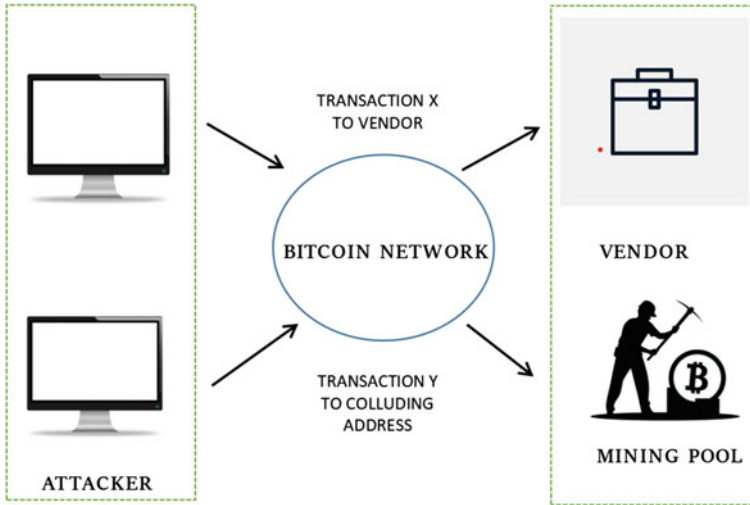


Fig. 7 Double spending

used to encrypt and sign transactions ensuring their security. If this private key is lost, it is quite challenging to retrieve. The user's blockchain account will be compromised if this key falls into the wrong hands. Since there is no central governing authority in blockchain, it is tough to track the malicious activities of a criminal and undo any tampered data. Hence, private keys must not be shared.

(4) **Criminal Activity:** Criminals rely on funds and cannot operate usual bank accounts for storing them as they are highly regulated institutions [14]. Alternatively, Bitcoin can be conveniently used in illegal activities since it offers anonymity. Users can engage in illegal trade and buy or sell anything without being tracked. Bitcoin acts as the financial enabler of money laundering, fraud, human trafficking, cybercrime, terror funding, and other unlawful acts. According to a blog from Chainalysis, cryptocurrency crime reached an all time high in the year 2021. Some frequent activities involve the following:

- a. **Ransomware:** Criminals use Bitcoin as the trading currency when it comes to deploying ransomware for ulterior motives as it offers anonymity. These programs encrypt the user's files and demand money in exchange for the decryption key which is essential to restore the affected files. In July 2014, a ransomware CTB-Locker (Curve-Tor-Bitcoin Locker) was transmitted around the globe by concealing itself as downloadable mail attachments. When a user clicks on this attachment, the virus infiltrates the system and begins encrypting files using elliptic curve cryptography. Unless the victim pays the attacker within 96 h through Bitcoin, the files remain scrambled leading to permanent loss of data.
- b. **Underground market:** Bitcoin is frequently used as the currency in the underground market. Silk Road is an incognito and international online black

market which runs on Tor hidden service. It allows the trade of illegal goods and services like drugs, illicit content, proprietary information, military-grade arms, etc. with Bitcoin, leaving no trail behind. This ultimately poses a threat to the social security of the countries.

- c. **Money laundering:** Bitcoin's features of anonymity and virtual payment allow consumers to hide their assets and launder money across seas. Dark Wallet uses advanced cryptography and zero-knowledge proofs to enable users to hold their own money and control their finances. It provides advanced security features that improve upon the existing Bitcoin protocol. It facilitates money laundering by mixing the user's authentic cryptocurrency with chaff coins.
- (5) **Transaction privacy leakage:** Blockchain systems take actions to safeguard the transaction privacy of users. In Bitcoin and Zcash, one-time accounts are used for stocking the received cryptocurrency and a private key is allocated to each transaction. This is done so that the attacker cannot deduce if the transactions are collected by the same user. In Monero, users may comprise some chaff coins (known as mixins) when a transaction is initiated. As a result, the attacker is oblivious of the actual coins expended by the transaction. Unfortunately, blockchain does not provide adequate privacy protections. Blockchain transactions are public and the addresses of the sender and receiver are easily traceable. Using statistical analysis, an attacker can make out the total amount of cryptocurrency being transferred. In other words, the transactions from a targeted user can be linked making the user's behavior traceable. Analysis shows that 66.08% of all Monero transactions don't include any mixins and 62.32% of transaction inputs having mixins are deducible.
 - (6) **Criminal Smart Contracts:** Criminals can use smart contracts for a variety of malicious activities, which may pose a threat to our daily life. CSCs (Criminal Smart Contracts) can promote the exposure of confidential information, theft of cryptographic keys, and multiple real-world delinquencies like terrorism, murder, arson, etc.
 - (7) **Vulnerabilities in smart contracts:** Smart contracts may have security vulnerabilities caused by program defects that can be taken advantage of in the following ways:
 - a. **Transaction-ordering dependence:** TOD (Transaction-Ordering Dependent) contracts rely on the miners heavily for smooth execution. Blocks may contain multiple transactions making the order in which they are mined and added to the blockchain crucial for TOD contracts.
 - b. **Timestamp dependence:** Each block in the blockchain consists of a timestamp. Some smart contracts' trigger conditions rely on this timestamp that the miner sets according to its local standard time. This endangers the contract since an attacker may try to alter it.
 - c. **Mishandled exceptions:** Sometimes contracts are linked to each other and depend on the other's completion. When contract X calls contract Y, if Y

runs unusually, Y will halt and give back false. Suppose contract X calls contract Y, which unusually halts running and conveys back a false value. This makes it necessary for contract X to specifically check if the call has been executed successfully. Not checking the exception information may make contract X vulnerable.

- d. **Reentrancy vulnerability:** When a smart contract is invoked, the actual state of the contract account is altered after the call is finished. The intermediate state can be used by the malicious actor to invoke a call that sets a chain of calls that repeatedly call to the smart contract.
- (8) ***Under-priced operations in Smart Contract:*** Gas is simply a transaction fee in Ethereum based on the computational resources and parameters like bandwidth, execution time, memory occupancy, etc. It is payable by the sender and used to reward miners. Gas is the product of gas price (set by the sender) and gas cost. Every transaction has a gas limit which acts as a protective measure to prevent the abuse of resources by throwing an out-of-gas exception, if the execution is more gas costly. For example, if some demanding input–output operations’ gas values are fixed at a very low level, these operations can be repeatedly implemented in quantity in a single transaction. If the gas cost set for EVM operations is not just then an attacker can commence DoS (Denial of Service) attack, whose purpose is to waste valuable computational resources at a low cost. This slows down the network leading to low transaction processing speeds which consequently pulls down the market value of Ethereum. EXTCODESIZE and SUICIDE are two DoS (Denial of Service) attacks that abuse this vulnerability of under-priced executions in smart contracts. Ethereum has applied new settings to mitigate such attacks but it still does not solve the problem completely [15].

4.2 Attacks on Blockchain

Various types of attacks have emerged since the advent of blockchain technology which take advantage of different weaknesses of the system. The BCSEC reports that around 2 billion dollars of economic losses were incurred because of blockchain security incidents in 2018. This rising trend in losses from attacks is clearly shown in Fig. 8. As the blockchain’s value grows, the likelihood of attacks will also rise [16].

Securing data is always the focus of people’s attention, and it is also the main reason why blockchain has not been widely used all over the world. Below we discuss some attacks that threaten the basic nature of this technology.

- (1) ***Selfish mining attack:*** The main strategy behind this attack is keeping a new set of blocks secret and releasing them at an advantageous time earning them undue rewards while invalidating the work of honest miners. The attacker or selfish miner keeps undiscovered blocks private and tries to fork a private chain to create another branch/chain of blocks. A miner strategically publishes blocks

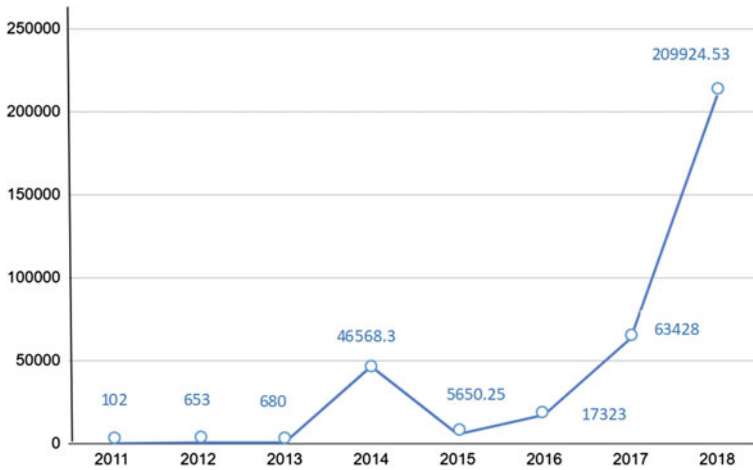


Fig. 8 Losses due to blockchain security incidents in tens of thousands of dollars (Year 2018)

once the public chain reaches the length of the private chain and consequently the blocks in the private chain become stale or orphaned [17]. The private chain now becomes the longest chain in the network so the honest miners are motivated to switch and start mining this chain. Hence, the honest miners end up wasting their time and resources in solving a computationally expensive puzzle fruitlessly. This attack is designed to waste the computational power of others. Selfish-Mine is a proposed attack plan, in which initially the length of the public as well as private chains are the same.

The following three situations can arise afterward:

- a. Both the selfish miners and honest miners discover the first new block simultaneously which gives rise to two concurrent forks initially of the same length. Honest miners may mine any branch of the two whereas the selfish miners carry on mining the private chain. If selfish miners discover the second new block too, they publish it to the chain and are rewarded for mining the pair of blocks. The private chain will now become the valid branch due to the fact that it is the longest chain. On the condition that honest miners initially discover the second new block and then add it to the private chain, selfish miners collect the first new block’s rewards, whereas honest miners will acquire rewards associated with the second new block. On the other hand, if the said block is written to the public chain, honest miners gain rewards for two new blocks while selfish miners receive none.
- b. Considering the hashing power of selfish miners is relatively less and assuming the public chain is longer than the private one, they may update the private chain based on the public chain. This scenario does not reward them.

- c. Selfish miners discover both the first and second blocks. In this scenario, they keep on mining blocks on the private chain while secretly withholding the two blocks. On the discovery of the first block by the sincere honest miners, selfish miners will broadcast their own first new block and the same process follows for the second new block. The selfish miners will carry on responding in this manner until the public chain exceeds the private chain by one block. Selfish miners will publish their last new block ahead of its discovery by honest miners, whereupon, the private chain will be deemed valid. As a result, selfish miners will be credited unwarranted rewards for all new blocks.
- (2) **BGP hijacking attack:** BGP (Border Gateway Protocol) is an external routing protocol governing how IP packets are forwarded to their destination. The main purpose is to intercept and divert the traffic by an ISP. In order to intercept the blockchain's network traffic, attackers either use BGP routing or manipulate it. Generally, hijacking BGP involves gaining control over the network operator, which can be exploited. An attacker can intercept the blockchain network by manipulating the BGP, and then data can be routed and the traffic can be modified to the attacker's favor. Taking a look at the node-level as well as network-level attacks, on Bitcoin, the number of the successfully to-be-hijacked Internet prefixes has a direct correlation with the distribution of mining power. Because of the high centralization of some Bitcoin mining pools, if they are encroached by BGP hijacking, it will have a striking effect. The hacker can slow down the pace of block propagation and divide the blockchain network.
- (3) **Liveness attack:** This type of attack can hinder the time it takes for a targeted transaction's confirmation or acknowledgment and cause unnecessary delay. Liveness attack consists of three phases which are (I) attack preparation phase, (II) transaction denial phase, and (III) blockchain delay phase.
- a. Attack preparation phase: Similar to the selfish mining attack, an attacker tries to attain a profit over the honest miners and builds his private chain, before a specific transaction is broadcasted to the public chain. In this scenario, the private chain is of a greater length than the public chain.
 - b. Transaction denial phase: The attacker tries to delay the block comprising the targeted transaction by withholding it from the public chain. When he can no longer hold the block, he moves on to the next phase.
 - c. Blockchain delay phase: As the public chain grows, it is no longer possible to hold the targeted block and the block is broadcasted. In some blockchain systems when the depth of a block is greater than a constant, it will be declared valid and so the attacker keeps on mining the private chain for the sake of building an advantage over the public one. He will then publish privately held blocks in the public chain and try to slow down the growth rate of the chain. When the targeted transaction's validity is confirmed by miners in the public chain, the liveness attack ends.

- (4) **Balance attack:** In balance attack, the attacker can compensate for his low mining power by momentarily disrupting the connections between subgroups that have equivalent mining power. By abstracting the blockchain, the attacker creates a DAG (directed acyclic graph) tree. He then inserts a delay between valid subgroups and performs a transaction in one subgroup termed as transaction subgroup and mines the other sub-tree (block subgroup). This is done so that the block sub-tree dominates the transaction sub-tree and the attacker can disregard that transaction and the attacker can alter the block containing the concerned transaction with high probability despite the transaction being committed. The attacker recognizes the subgroup with the targeted merchant and issues transactions to purchase goods from them. Afterward, the attacker gives out transactions to this subgroup and mines the blocks in the group. The attacker delays the communication until the merchant ships the goods. The attacker could re-create another transaction using exactly the same cryptocurrency given that with a high probability, the DAG tree seen by the merchant is outweighed by another tree. The attacker makes the nodes disregard the valid transaction allowing him to double spend successfully. The balance attack clearly proves that block obliviousness is a limitation of PoW-based blockchain systems. This means that a malicious actor can issue a transaction to a merchant and later remove it from the main valid branch.
- (5) **DDoS attack:** A denial-of-service (DoS) attack is a type of cyber attack that disrupts the normal operations of the host's services by overwhelming the network resource or machine making it unavailable to its users. It overloads the target system or network by flooding it with unwanted Internet traffic so as to cripple the normal services of the host. A DDoS attack refers to a distributed DoS attack in which unexpected traffic emerges from different distributed sources spanning all over the Internet network. DDoS attacks utilize the compromised systems to send large amounts of data to the host ultimately clogging the communications. Counteracting this by isolating each individual source and jamming them one by one is barely effective in this case. By knocking out a network partially or completely, this attack can render the blockchain inaccessible. On one side is the recovering rate of the compromised nodes and on the other is the success rate of the attack on the network.
- (6) **Sybil attack:** In Sybil attack, a malicious node forges multiple identities and operates them simultaneously in real time, which obliterates the reputation of the blockchain network. To an outside viewer, these identities appear as separate real entities and give unfair majority influence over the network to a particular node. Sybil attacks are capable of being executed directly or indirectly through a middle node. Private blockchains automatically prevent this attack as they authenticate a node's identity before it joins that network. Consensus algorithms like PoW, PoS, and DPoS make Sybil attacks impractical [18]. These algorithms make the mining process so resource intensive that it discourages a miner from attempting a Sybil attack and it is in the attacker's favor to continue mining honestly.

5 Security Enhancements

- (1) **Hawk:** As we have already discussed that the privacy protection measures in blockchain are not very robust. Research suggests that deanonymization is attainable with careful analysis of transactions. In blockchain, not only the transactions but also the smart contracts are publicly visible. Hawk is a framework that separates smart contracts into portions that are public and private, thereby granting some degree of privacy to smart contracts. Private data like personal information of the user and financial transactions can be confined to the private part and the rest of the data can be written to the public portion. It allows the users to write easy encryption-free code as Hawk provides its own cryptographic protocol to conceal confidential information. Hawk ensures transactional privacy through contractual security and on-chain privacy [19].
- (2) **SegWit:** SegWit or Segregated witness is an additional protocol upgrade that runs alongside blockchain. It separates the signature information or witness information and stores it outside in a side chain, allowing more transactions to be included in each block. This increases the throughput of the blockchain system, reduces transaction fees, and also improves scalability. In addition, the segregation of data makes the network more secure. Note that it does not increase the block size. It was originally deployed to combat transaction malleability problems.
- (3) **Oyente:** Oyente is an open-source tool used for finding bugs in smart contracts deployed on Ethereum. It uses the technique of symbolic execution with constraint solver Z3 to analyze the bytecode of the smart contracts without having to execute them. Ethereum blockchain stores this EVM (Ethereum virtual machine) bytecode. As illustrated in Fig. 9, it is made of four key components: CFG Builder, Explorer, Core Analysis, and Validator. The CFG builder accepts the byte code and global state as input to construct a CFG (control flow graph) in which nodes represent the execution blocks and edges represent the jump between them. Explorer runs the contract virtually and executes the states until there are none remaining or the time runs out. Z3 eliminates the proven impractical traces. The core analysis identifies the four major security flaws: TOD, timestamp dependence detection, mishandled exceptions, and reentrancy detection. The last part, the validator eliminates false positives. Oyente flags the contracts which are potentially vulnerable. This feature can be used by users for writing problem-free and better contracts.
- (4) **Lightning network:** Bitcoin users have to wait for a fixed period of time before transactions are confirmed and assured that they won't be reversed. In bitcoin, users have to wait for six block confirmations or about an hour to make the transactions full and final. With payments involving small amounts, the transaction fee is minimal which makes the transaction uneconomical. The lightning framework, a two-layer transaction mechanism, was introduced to resolve these problems. These transactions do not rely on block confirmations but on double

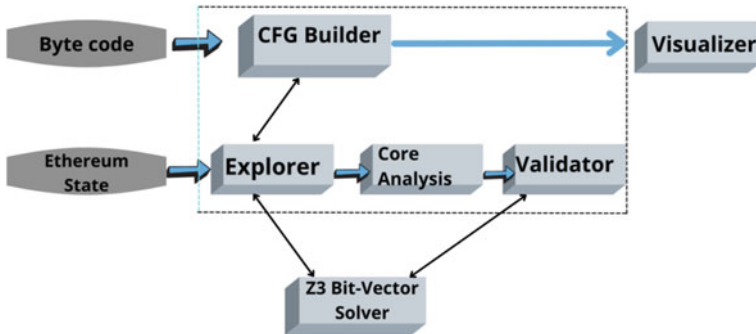


Fig. 9 Architecture of Oyente

signing from both the involved parties of the channel. The channel is bidirectional and can be opened by committing a simple fund. It allows users to conduct transactions between themselves and off the blockchain without making them public. Each user can have multiple channels across the network allowing for mass transfer of funds. These are free from the interference of a third-party miner. They can be closed unilaterally when one party decides to end the channel or when the transactions are completed. Only when the channel is closed, the final settlement is added to the blockchain [20].

6 Applications of Blockchain

Blockchain technology has a wide range of applications. A vast variety of fields like healthcare, voting, real estate, energy, governance, education, and many more can benefit from the advancement of blockchain technology. Blockchain is also expected to have an impact on the digital environment just like the Internet [21]. Initially, when the Internet was introduced, nobody had any idea about the great changes that it would bring to the world. Now in the present world, nearly every small and big activity is dependent on it. The time will come when Blockchain will have a similar impact too as it is gradually progressing every day. Outlined in Fig. 10, we highlight a few of the applications of blockchain that researchers across the globe have suggested.

Healthcare

- Introduction of blockchain in healthcare systems eliminates the need for intermediate entities making the traditional system more efficient and cheaper. It allows patients to timely receive required aid without any unwanted delay.
- Every patient has a distinct physical profile and the treatment strategy may differ from doctor to doctor. Blockchain technology improves the interoperability of health reports by integrating the medical records, allowing doctors/specialists from different institutions to work in cohesion and deliver better healthcare to the

Fig. 10 Applications of blockchain technology



patients. It makes every patient's absolute medical history easily accessible by granting a secure sharing platform.

- The BCT will provide transparency in the healthcare system when all the medical practitioners have access to accurate and unaltered data and information which prevents malpractices.
- Blockchain can be a solution to the fabrication of drugs in the pharmaceutical industry as all the transactions added to the distributed ledger are unalterable and digitally timestamped, which makes it easier to track any product and discourage such malignant practices. These drugs endanger a patient's life rather than curing the diseases; they may have negative side effects that can be fatal.
- The current medical record-keeping system lacks privacy. Blockchain offers significantly more security to keep each patient's sensitive information private. This decentralized information is also difficult to hack or alter as multiple copies are maintained in the network.
- Smart contracts can be utilized to implement health insurance that can release funds when the patient gets discharged. It streamlines the claim verification process and can be done automatically when some conditions are met, preventing false claims.

Insurance

- Smart contracts can be used to automate the insurance policies. This reduces the cost by eliminating the administration, processing, and other overhead costs. The terms are structured using digital protocols that precisely follow the predefined terms which prevent frauds by misinterpretation and disagreement of conditions.
- It reduces the level of paperwork, making underwriting easier and storage of data related.

Identity Management

- In real life, verifying identity using physical documents and IDs is easy but it is hardly effective in online systems. Blockchain may permit people to make an encrypted identity, which requires neither username nor password making it more secure and giving the user more control of his private data.
- Blockchain technology might allow the consumer to access and verify online payments by the mere use of an app for authentication rather than the usual username, password, and biometric security system enhancing transparency.
- Cryptography segregates data from the individuals' identities for better security. With the help of separate data, management companies can acquire only that data which is of their use thereby preventing the misuse of personal information like frauds in banking, trading, etc. It only protects from the leakage of confidential data by companies.
- By using blockchain-based identity management systems, we can eliminate the involvement of third parties for digital or manual identity management, identity theft, and identity sprawl.

Supply Chain

- Blockchain can drastically improve the transparency of products. It can verify the genuineness of products making the supply chains more efficient and competent.
- Blockchain improves the management and storage of inventory and reduces the related paperwork.
- It enhances traceability in a supply chain by offering real-time tracking of goods. It helps in locating the goods and hastening the operations. Tracking allows for locating the goods in cases of human error and fraud and makes operations swift and secure.
- Smart contracts in blockchain can connect the logistics related to delivery and its payments using digital contracts, improving the efficacy of the supply chain.
- Blockchain can transform the automobile industry. The whole vehicle history can be stored on a distributed ledger with an immutable feature which allows the buying of used vehicles trust-less and reserves the resale price of the currently used new vehicle. Information on the public ledger will help the future buyers know the exact value of the vehicles and also help the owners to receive the correct value for their vehicles. This technology will also help to eliminate the counterfeiting in the automotive industry.
- Blockchain can transform the automobile industry. A vehicle's complete history can be stored on the distributed ledger which makes buying and selling second-hand automobiles trustworthy and easy. The exact market value of a vehicle can be evaluated during resale, eliminating frauds and counterfeiting in the industry.
- In food-based supply chains, blockchain can be helpful in tracing the food products in case of an outbreak of a food-borne disease and identifying the contaminated sources.

Financial Services

- With blockchain technology, the transactions become faster which are predominantly slow in the existing banking systems [22]. Sometimes the transfer of funds across borders takes days to be successful. Blockchain provides for seamless transactions which are processed and verified within a matter of seconds across time zones.
- Blockchain facilitates inter-banking borrowing and lending of funds that are speedy, robust, and transparent.
- Auditing can be done in significantly less time as data stored in blocks is immutable and verifiable. Audits generally take from weeks to months to finish. The distributed ledgers can make verifying the integrity of transactional data using digital fingerprints possible.
- Investors can invest in decentralized hedge funds which eliminate the use of a hedge fund manager and minimize security risks.
- Credits score reports can be stored in blockchains. Due to its immutable nature, this information cannot be tampered with or sold or leaked. It enables small businesses to easily get approval for credits.

Music Industry

- Blockchain fixes the existing problems in the music industry of ownership in-clarity, royalty distribution, and monetization of music. Smart contracts in blockchains can deliver the required reliability and transparency, with music owners getting their rightful share of royalty.
- Blockchain can provide a vast decentralized platform for the music industry where artists and musicians can collaborate and directly share their content, eliminating the need for any intermediaries. Blockchain can also be used to form separate music streaming platforms.
- Copyright claims and issues are a big problem in the music industry and blockchain technology can provide the means to authenticate copyrighted music and prevent the sharing of pirated versions. It gives the music artists their well-deserved rights to the music.

Other Applications

- Blockchain can provide an E-voting system where people can simply vote from their mobile phones or PCs at a cheaper rate. It will guarantee a secure, anonymous, transparent, truly democratic, and completely safe system with data encryption and no breach of security.
- Blockchain in real estate assures direct means of connection between the buyers and sellers thus reducing the costs of intermediaries' fees and commissions. It also allows the estate to be tokenized and traded like cryptocurrencies thereby offering a chance for fractional ownership too.
- The distribution and encryption of data in blockchain provide transparency and records of the transactions. It also ensures reduced costs and increased speed of transactions hence making the entire IoT(Internet-of-Things) ecosystem proactive.

7 Trade-Offs and Challenges of Blockchain Technology

Blockchain is a fascinating and revolutionary technology and is being adopted by various industries belonging to diverse sectors. As per May 2019, 44% of the total global institutions have implemented blockchain [23]. However, like every other emerging technology, it has its disadvantages and restrictions. In this section, we talk about some fundamental difficulties encountered by blockchain technology.

- (1) **Performance & Scalability:** Latency and throughput are two variables in performance. Throughput refers to the number of transactions completed in a unit time, whereas latency refers to the time taken to add a block to the chain. Protocols like Proof of Work deliver low throughput and high latency because a lot of computational power and resources are required to work out the puzzles before a block of transactions can be added to a chain. For example, Bitcoin provides a low throughput of 6–7 transactions per second.

With a huge number of transactions taking place each minute, miners prefer those which have a high transaction fee since the size of each block is limited. The size of blocks in Bitcoin is 1 MB. This delays the small transactions making the system slow. A decentralized blockchain cannot have all the characteristics of decentralization, consistency, and scalability as stated in the DCS triangle given below in Fig. 11. Blockchain can fulfill only two conditions of DCS at a time. One provides high latency and low throughput while the other provides reduced transaction speed and low volume. Using alternative consensus algorithms can be a remedy to this problem.

- (2) **Energy consumption:** The blockchain network deals with a tremendous amount of message exchange and processing which includes computing complex mathematical problems to satisfy the consensus protocols. Hence, the consensus algorithms like Proof of Work are energy inefficient making them unsustainable. Whenever a miner successfully adds a block to a chain, he is rewarded Bitcoins for this painstaking work and this serves as an incentive to attract more miners

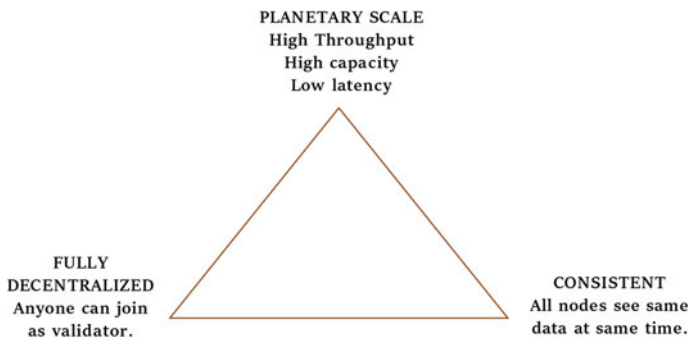


Fig. 11 DCS triangle

to run high-energy demanding devices. This process in turn adds more cryptocurrency to circulation. Subsequently, the total energy consumption of Bitcoin reached a new high. According to International Energy Agency, the total energy consumption of Bitcoin is higher than in a few countries. It is also predicted that the mining of Bitcoin could add to significant global warming and temperature may rise as much as 2 °C within a span of three decades [24]. According to estimates, each transaction alone has a carbon footprint of 274.29 kg of CO₂ per transaction. Although the application of blockchain technology in several cases cuts down intermediary costs, it comes at a greater price to pay.

- (3) **Privacy leakage:** Although blockchain provides security and privacy, it does not assure transactional privacy. A peer is anonymous in the blockchain network but research shows that IP addresses can be traced back to the peers' pseudonyms exposing the user's identity, even while hiding behind a firewall or NAT (network address translation). Peers may also be identified through its set of connected nodes. The main reason behind this is that all the transactions and the associated public keys are openly visible.
- (4) **Initial cost:** The initial cost of building the infrastructure of blockchain is quite high. It includes the cost of software development and a team of experts to launch the application.
- (5) **Public trust & perception:** People lack a technical understanding of blockchain technology usually. Blockchain involves several complicated terms and jargons, which makes it important for people to be fully aware of it before entering this ecosystem. Common man finds blockchain technology synonymous with Bitcoin. Furthermore, people also don't have much faith and trust in this new form of money (cryptocurrency) hence it requires meticulous marketing strategies to gain people's confidence and make it a worthy investment.
- (6) **Regulatory problems:** Cryptocurrency weakens the control of central banks on the economy of a state. There are no international laws to regulate cryptocurrency or bitcoin and its legal status varies from country to country. Canada has a bitcoin-friendly status and treats it like a commodity while Australia considers it a currency. There are no uniform regulations in the European Union. In contrast, Bitcoin is banned in countries like Iran, Ecuador, Pakistan, Morocco, etc.
- (7) **Immutability:** The immutable nature of blockchains may not be a fit model for scientific research where new findings and literature are published every day. This requires the scientific literature to be updated, changed, or may altogether contradict the previous findings.
- (8) **Cybercrime:** We have already discussed the role of blockchain-based cryptocurrency and how it is being exploited by criminals and terrorists to further aid their ill motives. This puts in question the legitimacy of blockchain as means of storing and managing data in information centers and libraries [25].

8 Conclusion

We have discussed the core concepts of blockchain technology and some of the most important characteristics. Blockchain technology is widely recognized and highly reputed due to its decentralized and immutable nature. In this chapter, we theoretically discussed various vulnerabilities as well as attacks on blockchain that obstruct the increased adoption of this technology. While we enjoy the perks of this disruptive technology, it is important to stay cautious of the existing security risks. With the expansion of its applications, new security threats are emerging as well. The way to strengthen the security is to divert the emphasis from applications to the analysis of blockchain security and advance research in this area.

Blockchain technology has shown great advancements in various fields since its establishment in 2009 extending from the traditional cryptocurrency to the present smart contract. Although it is still in the infant stage, we should not underestimate the optimistic socio-economic advantages of this remarkable technology. Some of its issues have been solved but it has still got a long way to go. The governments of various nations have to make regulatory laws for this technology before it can be embraced by even more companies and organizations. To conclude this chapter, in the final section, we have covered the uses, benefits, and future applications.

References

1. Mishra, S., Dash, A., Ranjan, P., Jena, A.K.: Enhancing heart disorders prediction with attribute optimization. In: *Advances in Electronics, Communication and Computing*, pp. 139–145. Springer, Singapore
2. Ekblaw, A., Halamka, J.D., Lippman, A.: A case study for blockchain in healthcare: MedRec prototype for electronic health records and medical research data (2016)
3. Azaria, A., Ekblaw, A., Vieira, T., Lippman, A.: MedRec: using blockchain for medical data access and permission management. In: *International Conference on Open and Big Data, OBD*, pp. 25–30
4. Yue, X., Wang, H., Jin, D., Li, M., Jiang, W.: Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. *J. Med. Syst.* 218 (2016)
5. Mishra, S., Mishra, B.K., Tripathy, H.K.: A neuro-genetic model to predict hepatitis disease risk. In: *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*, pp. 1–3. IEEE (2015)
6. Mishra, S., Panda, A., Tripathy, K.H.: Implementation of re-sampling technique to handle skewed data in tumor prediction. *J. Adv. Res. Dyn. Control Syst.* **10**, 526–530 (2018)
7. Gervais, A., Karame, G.O., Wüst, K., Glykantzis, V., Ritzdorf, H., Capkun, S.: On the security and performance of proof of work blockchains. In: *The 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 3–16
8. Mishra, S., Thakkar, H.K., Mallick, P.K., Tiwari, P., Alamri, A.: A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection. *Sustain. Cities Soc.* **72**, 103079 (2021)
9. Christin, N.: Traveling the silk road: a measurement analysis of a large anonymous online marketplace. In: *The 22nd International Conference on World Wide Web*, pp. 213–224 (2013)
10. Juels, A., Kosba, E.S.: The ring of gyges: investigating the future of criminal smart contracts. In: *The ACM SIGSAC Conference on Computer and Communications Security*, pp. 283–295

11. Zhang, F., Cecchetti, E., Croman, K., Juels, A., Shi, E.: Town crier: an authenticated data feed for smart contracts. In: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, pp. 270–282
12. Chen, T., Li, X., Luo, X., Zhang, X.: Under-optimized smart contracts devour your money. In: IEEE 24th International Conference on Software Analysis, Evolution and Reengineering, SANER, pp. 442–446
13. Mondal, S., Tripathy, H.K., Mishra, S., Mallick, P.K.: Perspective analysis of anti-aging products using voting-based ensemble technique. In: Advances in Systems, Control and Automations, pp. 237–246. Springer, Singapore (2021)
14. Tripathy, H.K., Mishra, S., Thakkar, H.K., Rai, D.: Care: a collision-aware mobile robot navigation in grid environment using improved breadth first search. *Comput. Electr. Eng.* **94**, 107327 (2021)
15. Eyal, I., Sirer, E.G.: Majority is not enough: bitcoin mining is vulnerable. In: Financial Cryptography and Data Security—18th International Conference. Lecture Notes in Computer Science, vol. 8437, pp. 436–454 (2014)
16. Apostolaki, M., Zohar, A., Vanbever, L.: Hijacking bitcoin: routing attacks on cryptocurrencies. In: IEEE Symposium on Security and Privacy, pp. 375–392 (2017)
17. Yan, H., Oliveira, R., Burnett, K., Matthews, D., Zhang, L., Massey, D.: BGPmon: a real-time, scalable, extensible monitoring system. In: Cybersecurity Applications Technology Conference for Homeland Security, pp. 212–223 (2009)
18. Singh, A., Ngan, T., Druschel, P., Wallach, D.S.: Eclipse attacks on overlay networks: threats and defenses. In: 25th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies (2006)
19. Tripathy, H.K., Mallick, P.K., Mishra, S.: Application and evaluation of classification model to detect autistic spectrum disorders in children. *Int. J. Comput. Appl. Technol.* **65**(4), 368–377 (2021)
20. Greveler, U., Justus, B., et al.: A privacy preserving system for cloud computing. In: 11th IEEE International Conference on Computer and Information Technology, pp. 648–653 (2011)
21. Mishra, S., Tripathy, H.K., Thakkar, H.K., Garg, D., Kotecha, K., Pandya, S.: An explainable intelligence driven query prioritization using balanced decision tree approach for multi-level psychological disorders assessment. *Front. Pub. Health* **9** (2021)
22. Wang, Q., Wang, C., et al.: Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing. IEEE (2010)
23. Mohapatra, S.K., Mishra, S., Tripathy, H.K., Bhoi, A.K., Barsocchi, P.: A pragmatic investigation of energy consumption and utilization models in the urban sector using predictive intelligence approaches. *Energies* **14**(13), 3900 (2021)
24. Gervais, A., Karame, G.O., Wüst, K., Glykantzis, V., Ritzdorf, H., Capkun, S.: On the security and performance of proof of work blockchains. In: The ACM SIGSAC Conference on Computer and Communications Security, pp. 3–16 (2016)
25. Stolfo, S.J., Salem, M.B., Keromytis, A.D.: Fog computing: mitigating insider data theft attacks in cloud. In: IEEE CS Security and Privacy Workshop (2012)

An Exploration Analysis of Social Media Security



Shreeja Verma and Sushruta Mishra

Abstract Social media security is a rising concern among today's generation, as when the pandemic started, a lot more people than before have begun using social media due to lack of entertainment or other reasons. There is a rise of 62% in Ransomware since 2019 (pre-pandemic), as mentioned by the Cyber Threat Report by SonicWall. As cybersecurity attacks are becoming more severe, this number of attacks is still set to rise. So in order to investigate the possible security issues, this paper digs deep into the concepts of social media security, potential threats and feasible solutions. The importance of having users' data secure and protected from various threats such as malware attacks, identity theft, cyberbullying and so on, is addressed so that neither the user nor the developers suffer from any loss. Organizations may do more effective patch management to prioritize security-related patching and update their software in accordance with the solutions discussed in this paper.

Keywords Exploits · Vulnerabilities · Malware · Steganography · Metadata · Third-party

1 Introduction to Social Media Security and Its Evolution

As we have stepped into the second decade of the twenty-first century, we have seen technology grow remarkably in comparison to the times when smartphones were non-existent. With the advent of social media platforms like WhatsApp, Facebook, Instagram and Twitter, we have not only witnessed the creation of a new standard of communication but also an increase in security issues despite the various efforts made by the developers of these platforms [1].

In the early 2000s, with the advent of social media, cybercrime began to take its toll. The influx of personal information into profile folders has led to the emergence of ID fraud. People with malicious intent use information to set up bank accounts, suspend credit cards or engage in other forms of financial fraud in various ways.

S. Verma · S. Mishra (✉)

Kalinga Institute of Industrial Technology, Deemed to Be University, Bhubaneswar, Odisha, India
e-mail: sushruta.mishrafcs@kiit.ac.in

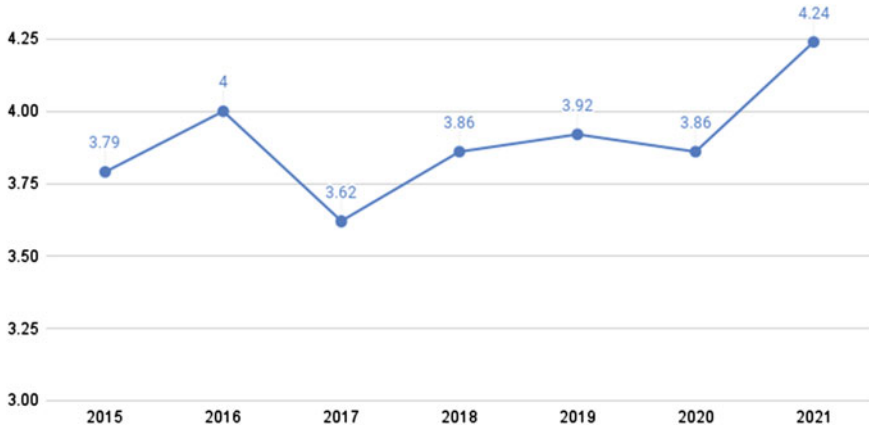


Fig. 1 Average total cost of a data breach (in US\$ millions)

Currently, ransomware threats are increasing because criminals are pursuing threats to disclose information they have gained through the social media profile of an individual or organization. Ransomware is a type of malware program that limits users and system administrators' access to files or all networks. If a malware program invades systems, attackers will send a ransom note usually authorizing payment via Bitcoin. Ransomware made history in 2020 contributing to the first reported deaths related to cyberattacks. In this case, the German hospital was closed off from its programs and could not treat patients. A woman who needed urgent care was taken to a nearby hospital 20 miles away but did not survive. According to the report rendered by IBM, In 2021, inflation in the average total cost of a data breach was the highest amount in 7 years [2]. Data breach costs have increased considerably between the years 2020 and 2021, growing from \$3.86 million in 2020 to \$4.24 million in 2021 (Fig. 1).

With such extreme threats and vulnerabilities, social media cybersecurity comes into action. For the users to be able to rely on these social media platforms, it is necessary to constantly monitor the security threats and vulnerabilities and fix them as soon as possible. It is also important to be able to detect such threats in advance before it comes to the notice of a pernicious client who later exploits these vulnerabilities [3].

2 Important Issues Involving Security for Social Media

While millions of users share their content on social media without any worry for their data, there exist issues that concern the security of these vulnerable platforms that are prone to be exploited by hackers. Some extremely important ones associated with it are listed below.

2.1 Privacy of Data

When users share their personal information on social media, there are many ways in which this data can be misused by corrupt clients around the world [4].

2.1.1 Metadata

Social media content has a lot of metadata, which can be used by cyberstalkers to gather information about their targets. For instance, if they have access to a person’s location, they can easily find out their device’s details. In our day-to-day lives, we come across metadata more frequently than we know. Each time you open an e-mail, read a book or order something off Amazon, or while communicating over the telephone, you’ve come upon metadata (Fig. 2).

Law enforcement organizations around the globe are infamous for using metadata from e-mails, digital messages and other forms of telecommunications to conduct investigations and achieve their goals. In 2015, when the Australian Parliament made it obligatory for communication carriers to keep a 2-year database of all telephone metadata, a media storm erupted as critics considered the site a hacker’s honeypot, claiming that data theft could lead to serious damage to citizens’ privacy [5].

2.1.2 Shared Ownership

In the age of sharing a humongous amount of data with each other, where just a single client can manipulate the protection settings of the particular multimedia, often results in proprietorship loss on the content. For example, you may have access

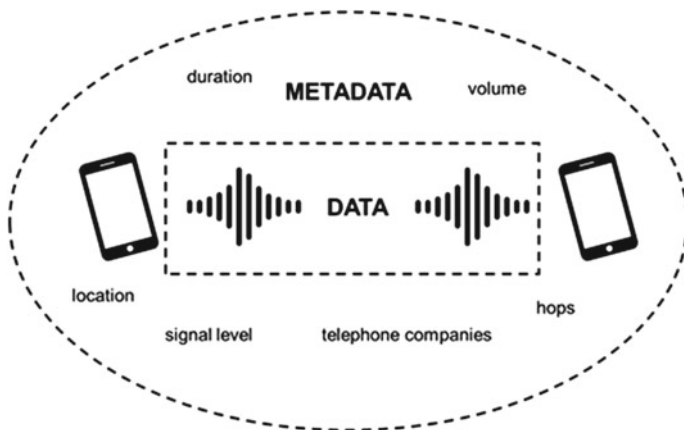


Fig. 2 Example of metadata [36]

to edit a google document but someday due to personal grievances, the creator might change the access settings to private, resulting in loss of ownership of the document.

2.1.3 Tagging

While posting multimedia content, a user not only puts his/her personal information at risk but also puts the ones, who are not active users or may not be a user of the particular platform at all, in danger as well by revealing personal information about more than one user, by tagging others in a single post [6].

2.1.4 Social Engineering

Undermining the security of sensitive data of an organization to acquire it for malicious activities is called Social Engineering which happens due to the exploitation of personal trust. Social networking sites, such as [7] Facebook, allow you to create your own page and interact with a network of people who may contain malicious users. By exploiting the trust the user has in his or her network, cybercriminals can embed malware or computer viruses on the content of their Facebook friends which can further lead to more users being duped.

2.2 Data Mining

We all leave our digital footprint anytime we interact with the internet [8]. At any point in time, if we create a social media account, we leave behind a set of traceable digital activities, actions, contributions and information. This data trail can be stored in various ways and can be exploited by attackers.

2.2.1 Third-Party Access

Users may delete or deactivate their social media accounts, but they may not be able to erase the data that the platform has shared with the third-party vendors, who are usually in a high-risk area for privacy breaches. Targeted advertising, which is another byproduct of all the data provided to the third-party platforms, may collect our real-time information without our knowledge.

According to a blog post by Yaffa Klugerman (Director at Panorays), listed below are the Five Most Noteworthy Third-Party Data Breaches in the year 2021:

1. **Accellion:** Late last year, the vulnerabilities in Accellion's File Transfer Appliance were used by cybercriminals [9]. The Accellion's File Transfer Appliance is used for the mobility of large and crucial files inside a network. Their victims

included the State Bank of New Zealand, Washington State, Kroger Grocery Store, the University of Colorado, cybersecurity company Qualys and many others.

2. **Audi and Volkswagen:** Earlier in the year, the Volkswagen Group of America, Inc. came to know about the insecure data that was dumped on the internet by one of its sellers, which was acquired by some illegitimate group. The violation hit 3.3 million consumers, which is more than 97% of Audi customers and interested consumers [10–12].
3. **Click Studios:** In April, Click Studios informed its customers that a business password manager, Passwordstate, a business password manager, was hacked by cybercriminals who exploited Click Studios' update system and delivered malware to customers. Click Studios notified its users about this in April as it had affected more than 370,000 security and IT professionals in 29,000 companies globally.
4. **Cancer Centers of Southwest Oklahoma:** An unauthorized access to protected data which included names, Social Security numbers, addresses, treatments and medical diagnoses, for around 8000 oncology patients was made in 2021. The cloud-based storage provider for the Cancer Centers of Southwest Oklahoma, Elekta, received an unusual outburst on their network which lead to this breach.
5. **Kaseya:** There exist many groups of malevolent attackers infamous for their malicious intent, one of them, known as REvil ransomware team, attacked Kaseya VSA, which is a remote monitoring and management software platform. Kaseya had to close down both the on-prem and the SaaS servers as a precautionary measure after that, as about 1500 companies were hit worldwide [13].

2.2.2 Profiling and Profile Cloning

Clients with malicious intent may often keep a close watch on their target's social media account to extract every bit of information from the content they post. Thorough data analysis can further help the attackers in profiling and profile cloning, which can lead to blackmailing, cyberbullying and cyberstalking.

2.2.3 Corporate Espionage

Big organizations using social media accounts for marketing purposes may help attackers to assemble internal data through the content that is being posted. Competing companies might use this information to destroy the reputation of the concerned organization and hence encourage such spies to keep a close eye on every bit of information [14–16].

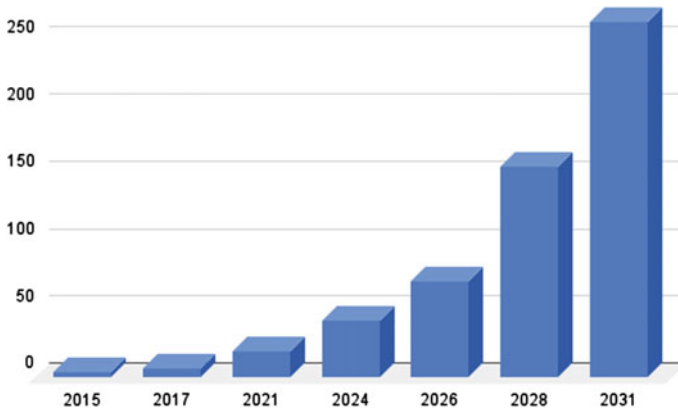


Fig. 3 Global ransomware damage cost prediction (in billion USD)

2.3 Virus and Malware Attacks

2.3.1 Malware Attacks

Attackers often use malevolent URLs to direct users to a site that installs malware in their systems. These further help the attackers in extracting confidential data from the user's PC such as passwords and account details. Cybersecurity Ventures revealed in an article that by 2031, ransomware will probably take a toll on its targets by costing them more than \$ 265 billion (USD) annually, with new entries every 2 s as ransomware hackers gradually improve their loading of malicious software once and for all, and related fraudulent activities [17]. The dollar figure is based on a 30% annual growth rate in damages over the next 10 years (Fig. 3).

2.3.2 Phishing

Phishing, a word derived from the word fishing which means 'baiting' people [18], is generally done via emails or texting. It guides the user to a fraudulent website and asks for private information such as usernames and passwords of bank accounts or other official accounts. 2021 Tessian research has revealed that employees get an average of 14 malevolent emails per year. A few industries were affected worse than the others, with retail workers receiving an average of 49. Mishra et al. [19] CISCO's 2021 report on cybersecurity threats shows that at least one person has clicked on a link to a phishing website at about 86% of organizations. Company data suggests that phishing accounts for about 90% of data breaches. CISCO also found that phishing is often high during the holidays, based on the fact that the number of phishing attacks increased by 52% in December.

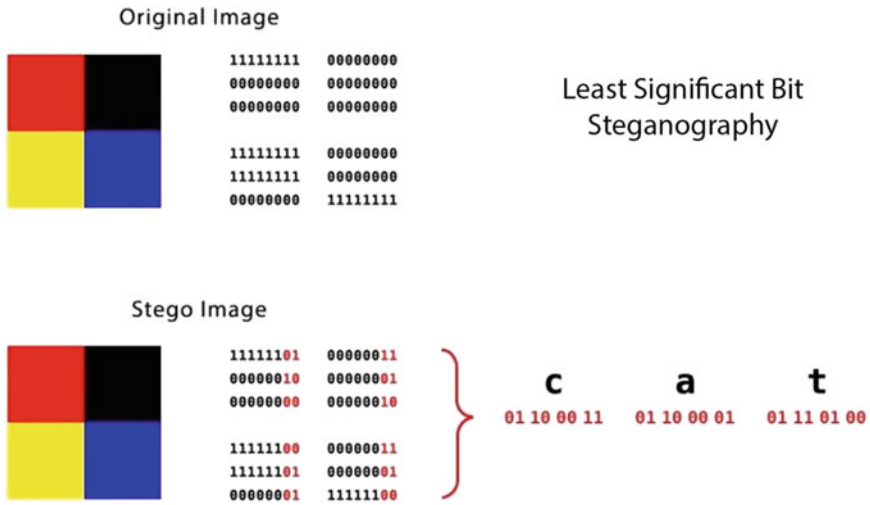


Fig. 4 Example of steganography—least bit steganography [37]

2.3.3 Steganography

The practice of concealing information in different objects has been observed for ages. Before the computer was built, sensitive messages were hidden in furniture, quilts, paintings and so on. But in today’s time of technological advancements, attackers are using digital media to conceal malicious data in plain sight. For instance, an image of a dog may contain some malevolent code that gets triggered once clicked and it deletes all the system files from your computer. Using images to conceal data is the most common form of steganography in today’s time and attackers may change just the least significant bit that does not affect the appearance of the image and is not detectable by the naked eye (Fig. 4).

It is important to note that Cryptography and Steganography are not alike and they have complementary purposes. Information might be encrypted and then concealed using Steganography.

2.3.4 Click-Jacking

It is an attack that lets a user click on a web page element that is hidden or corrupted with a malicious element that can lead to unwillingly installation of malware, transfer money or submit important usernames and passwords [20].



Fig. 5 Session Hijacking attack [38]

2.3.5 Session Hijacking

A type of man-in-the-middle attack used to exploit social media accounts. It occurs when an attacker steals a session cookie, which can provide identity, access and tracking information when active. It can also occur when the attacker injects malicious code into frequently used websites by its target (Fig. 5).

According to Alabrah et al. [21], there are three types of Session Hijacking:

1. Active Session Hijacking—Attacker attacks on an already operational session between user and server. The attacker puts himself in the position of a legitimate user with a DOS attack.
2. Passive Session Hijacking—Here, the attacker positions himself in the middle of the legitimate user and the server. He receives the packet in between the transition of the packet from the user to the server. The attacker can also alter the information carried by the packet to achieve his mischievous goals. The only disadvantage for the attacker here is that he can have access to the packets only till the session is active [22].
3. Hybrid Session Hijacking—It is a combination of both active session and passive session hijacking. It can further be divided into categories: Blind Spoofing Attack and Non-Blinding Spoofing Attack [23].

2.4 Legal Issues

2.4.1 Grooming

When an adult (18 years or above) makes an attempt to trap minors through the internet (games or social media platforms) with the intent of sexual assault, it is called grooming [24]. Social media opens the door for such paedophiles, as they can easily communicate with their targets. They may impersonate someone else or try to befriend them through their jovial act.

Table 1 Data breaches in well-known digital platforms

Organization	Year	Impact
Aadhaar	2018	1.1 billion people
LinkedIn	2021	700 million users
Facebook	2021	533 million users
Twitch	2021	700 million users
Twitter	2018	330 million users
Dubsmash	2018	162 million users
MeetMindful	2021	2.28 million users
Raychat	2021	150 million users
Tinder	2020	70 thousand female users
TikTok	2020	235 million accounts
Instagram	2019	49 million records

2.4.2 Cyber Bullying

With more and more young teenagers joining social media platforms, there has been a drop in social media security. There have been many instances where media content of young girls has been manipulated and exploited on social media by pernicious clients [25].

3 Risks and Challenges of Social Media Security

3.1 Information Revelation

The term ‘data spill’, according to the National Security Agency, refers to the transfer of isolated or sensitive information to unauthorized systems, people, applications or media. A multimedia substance with its metadata when uploaded on social media platforms, or Data breaches at third-party apps lead to the revelation of personal and confidential information to the attackers. Malware that gets installed on a user’s PC through a malevolent URL might extract all the personal data stored on the system, hence leading to Data Disclosure [26] (Table 1).

3.2 Location Spillage

Third-party apps that have access to the user’s real-time data, data breaches, metadata and multimedia content exposure on social networking sites often result in the leak of the location of the user [27].

3.3 Cyberbullying and Cyberstalking

With the increasing ease of communicating with people around the globe, there has also been an upsurge in cyberbullying. A user can interact with his/her friends or even with celebrities through public comments on their posts. Many users use this as a medium to pass lewd comments, spread rumours and hatred around them. Hackers easily get access to a user's account and their data, after which they may misuse it or even blackmail the user [28].

3.4 Cyber Terrorism

The nature of terrorism threats has considerably changed over the past 20 years. Social media has now also become a medium for terrorists to spread hatred and rumours around society and hence convince a larger group of people to support their inhumane cause. In 2017, organizations in more than 150 countries were struck by the two highly vicious attacks of WannaCry and NotPetya, that caused losses estimated at more than USD 300 million along with the damage to reputation because of the loss of customer data. And as the cyberattackers get more and more adapted to the older security barriers of the social media accounts, we can now also observe the shift in the nature of cyberattacks; from individual consumers to global political and economical systems.

3.5 Reputation Misfortune

Often corporate employees, school students or social specialists who have the agenda of degrading a superior's reputation, make use of social media platforms. These lead to the downfall of a person's or an organization's prestige.

3.6 Identity Theft

Social media platforms like Facebook or Instagram have billions of users who have stored their personal information or constantly post tremendous amounts of content that increases the risk of leaking personal information has become a hub for attackers to utilize this data to apply for loans or commit other financial fraud without being caught.

4 Social Media Networks Security Solutions

Social media threats are more ostensible now than ever and thus, many researchers have developed various security measures to curb vulnerabilities and cybercrime [29].

4.1 Watermarking

Watermarking represents the commercial application of steganography and its potential aim is to reduce cybersecurity risks. It is the process of embedding a piece of code, sound or any tag to identify the proprietorship of the content in which it is implanted [30]. It can be robust, i.e. information can be restored after a dangerous or weak attack, where the information cannot be duplicated or verified after basic flag correction. There also exists semi-delicate watermarking which is half powerful as delicate watermarking.

Currently, watermarking is used for

- Copyright protection—preventing third parties from infringing on digital media ownership.
- Fingerprinting—to obtain information about a digital media receiver (owner) to track distributed media copies.
- Copy protection—to prohibit data copying devices from copying the digital media if it is copy-protected.
- Image verification—originality of the digital media is verified.

The basic idea of watermarking involves the addition of a watermark signal to the host data that will be marked in such a way that the mark is invisible and protected in the signal. Watermark can be partially or fully obtained only if the cryptographically secure key is available (Fig. 6).

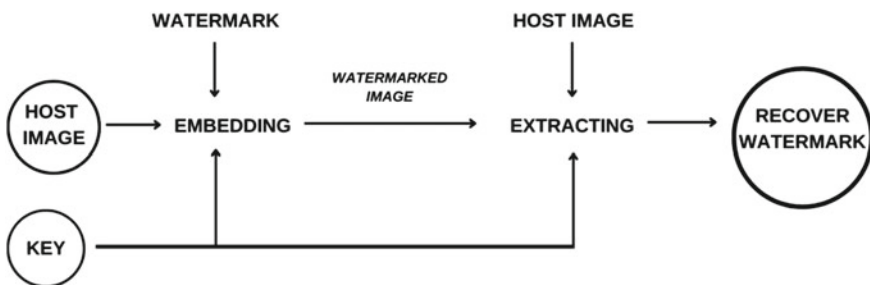


Fig. 6 Embedding and extracting process of digital marketing

4.2 Steganalysis

It is an approach to detect steganography by inquiring at differences between bit patterns and remarkably big files. The aim is to recognize suspected hidden data, find out whether or not they have got obscured messages encrypted in them and recover those messages if feasible. Steganalysis typically starts with numerous suspected information streams because of the uncertainty of whether these information streams incorporate a secret message. The steganalyst then begins by reducing the large set to a subset of most likely altered data streams. Analysis of encrypted information may also take several forms: obtaining, removing, disqualifying or demolishing encrypted information. The technique used in steganalysis relied on the information that can be gathered by the steganalyst, which includes only the steganography medium available for analysis, the hidden message known and so on [31]. In the procedure of steganalysis, the stego image is blocked and the steganalyzer also sometimes ties to extract the hidden message or information. Figure 7 shows the block diagram of the generic steganalysis process.

4.2.1 Steganalysis Techniques

Unusual Patterns

Suspicious for steganography arise from the detection of unusual patterns. There is a decline in the quality of digital media if it is being altered to conceal an object. For example in the case of Network Steganography, Packet headers are stuffed with strange patterns due to the fact that these packet headers are hardly read by anyone in general, thus making them a good hiding place [32].

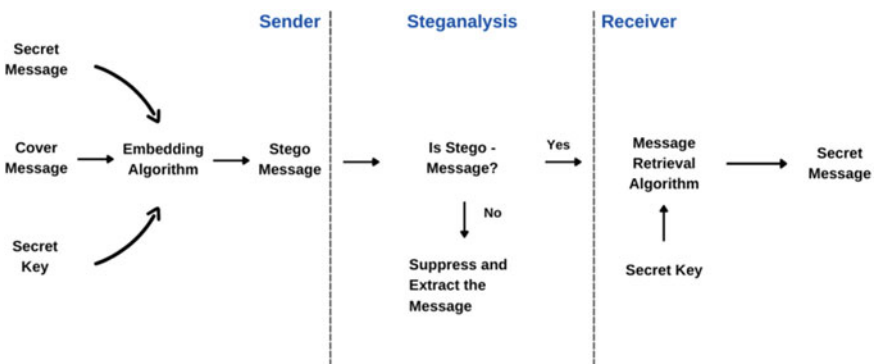


Fig. 7 Block diagram of steganalysis

Visual Detection

The identity of hidden data or a steganography tool can be detected with the help of analyzing repetitive patterns. Comparing the original image with a stego and thereafter observing noticeable differences is one of the methods to test these patterns. Such attacks where the carrier is known and discovered easily are called known-carrier attack. In case of unavailability of cover images, even the signatures that are received are enough to indicate the existence of a hidden message or to identify the tool that was used to embed these messages [33]. These hidden message signatures can be obtained by comparing multiple images. Other visible indicators of a hidden message include cropping or padding of images because a certain stego tool cuts blank spaces to fit a stego image into a fixed size. Other methods that can come under Visual Detection of Steganalysis Technique are noticing the variation in file sizes of the cover image and the stego image, or the differences in the colours of both the images.

Tools to Detect Steganography

Several tools to detect steganography are available, such as EnCase by Guidance Software Inc., ILook Investigator by Electronic Crimes Program, Washington DC, various MD5 hashing utilities, etc. An automated tool for detection, known as Stegdetect, provided by Niels Provos, is also quite popular. Stegdetect uses specific steganography-based applications, is used to find hidden data in JPEG images.

4.3 Digital Oblivion

Digital Oblivion is commonly known as the right to forget [34], meaning that data must be identifiable and usable for a limited time. This approach is very helpful in protecting the privacy of astronomical amounts of data since Online Social Networks are thriving with billions of users. Generally, there are two ways to implement digital oblivion using expiration dates, the first one relies on cryptography (such as employing expiration dates) and another in which the data is stored on an external, highly secured server [35]. Tools like X-Pire software creates encrypted copies of images and asks users to give each one expiration date.

But there are certain challenges that hinder the implementation of digital oblivion as pointed out in a recent EU report:

- (i) allowing a person to locate where their data is stored,
- (ii) tracking all the copies and the information deduced from the item,
- (iii) to ascertain whether a person has the right to request for the removal of a data item and
- (iv) affecting the erasure or removal of all exact or derived copies of the item in the case where an authorized person exercises the right.

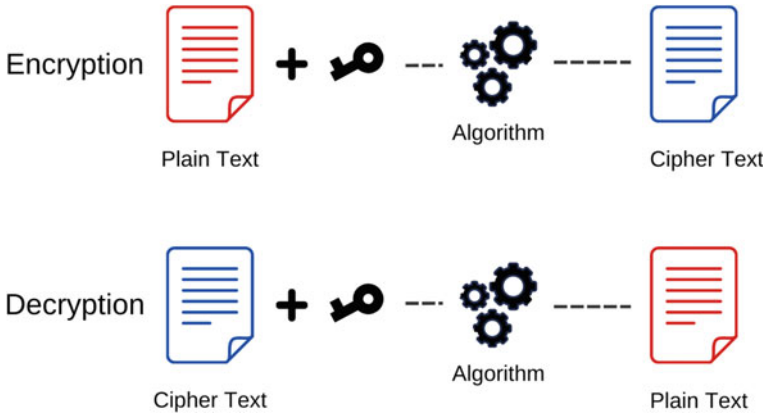


Fig. 8 Encryption and decryption of data

4.4 Storage Encryption

As many new social media platforms are emerging day by day, not each one of them owns private data centres. The majority of them store the user data on a third-party server. Personal user information may be shared with unauthorized groups by these third-party data centres for their own advantages and neglect the user's consent. So using encrypted storage and site encryption goes a long way in reducing the risk of data exploitation. In today's time, we use electronic tools to develop different data encryption algorithms to shuffle your data. Plain text or any type of data can be encoded using this methodology. The encrypted data can only be decoded by someone who has the decryption key to keep the data safeguarded. The encrypted data is generally referred to as ciphertext, while unencrypted data is called plain text (Fig. 8).

4.5 Detection of Malware and Phishing

Detection of Malware can be performed in multiple ways. Malware is a grave threat that leads to other serious problems such as ransomware and one may even remain unaware that their system has been infected by malware until it's too late.

4.5.1 Techniques for Malware Detection

Anomaly-Based Detection

There are two phases through which anomaly-based diagnosis can occur—the training (learning) phase and the adoption (monitoring) phase. During the progression of the learning phase, the detector tries to learn normal behaviour. The main advantage of anomaly-based discovery is its ability to detect zero-day attacks. The Zero-day attack as described is a previously unknown attack on the malware detector program. This process has two basic limitations: a high level of false alarms and complexity.

Static Anomaly-Based Detection

In statistical anomaly-based detection, specifications about the file system under review, are used to capture malicious code. The main advantage of statistical anomaly-based detection is that its use can make it possible to detect malicious software without allowing malware to run on the host system.

Hybrid Anomaly-Based Detection

- (i) **Ghostbuster:** Wang et al. [7] suggest how to identify the type of malware that they call ‘ghostware’ by offering a ‘cross-view diff-based’ approach. Ghostware is a malicious computer program that can hide its presence in the operating system query programs. For example, all the resources of the ghostware will be invisible to a user if he/she executes a command to list files in the current directory, ‘dir.’
- (ii) **Self-nonsel:** The method proposed by Forrest et al. [8] is generally not accomplishable in real detection aids. The goal of the suggested method is to detect changes in protected data. The limit of this method is that it cannot detect the extraction of objects in a protected database.

4.5.2 Phishing

The majority of the strategies to detect phishing sites revolve around machine learning procedures in which highlights of sites are used in recognizing phishing websites. The phishAri process of ongoing ID theft crime that takes place on Twitter is a program that combines tweets posted with URLs into two categories: identity theft or authenticity using tweet’s objects. Another program, WarningBird, identifies malicious links posted via Twitter. They may be able to withstand the onslaught of identity theft by hiding malicious URL-based redirects.

4.6 Prediction of Cyberattacks Through Monitoring Social Media

Cybercriminals use socially explicit information to create exploit code as part of a series of cyber killing chains that follow a series of online attacks from the initial stages to data extraction. Social Media forums such as Twitter provide predictable details of potential zero-day attacks. This examination of potential pre-existing threats creates a time lag in the exploitation process; from the time, criminals receive information about the vulnerability, to the ultimate exploitation. It is found that using a monitored machine learning system such as Random Forest can help detect potential exploitation through the information available on Twitter (Fig. 9).

Sapienza et al. [13] developed a model that warns of the latest cyber dangers by constantly observing social media posts of highly qualified security researchers, analysts and dangerous criminals of white hats, posts related to attacks, threats and vulnerabilities. Sauerwein et al. [23] noted that there were discussions on about one-fourth of the perceived vulnerabilities on Twitter before they were disclosed in public and that Twitter could provide an opportunity to respond to newly identified risks, thus being a good source for preventing exploitation, crippling cyber crimes and benefit organizations in performing efficient patch management and prioritize security-related patching as well.

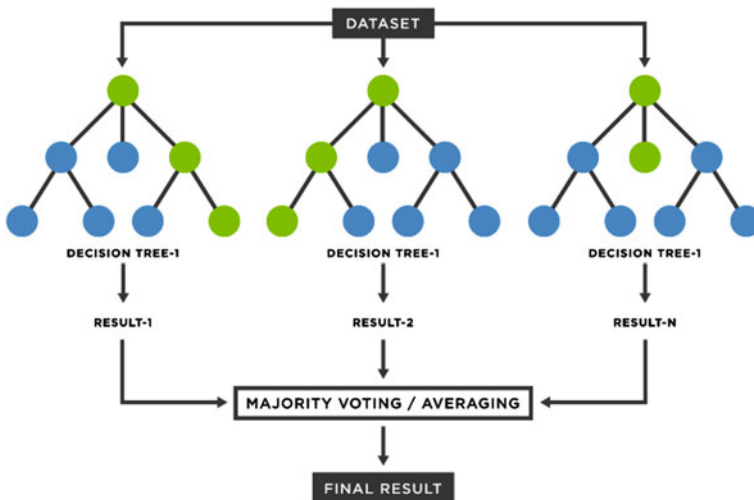


Fig. 9 Random Forest [39]

4.7 Time Lag-Based Modelling for Software Vulnerability Exploitation Process

Only after detection, potential risk can be ill-used. From the moment, the attackers get the information about a new vulnerability that has been discovered, to the final act of exploitation taking place, there exists a finite time lag in between the whole process of exploitation. Using this time lag approach as a benefit, Choudhury et al. [17] developed a model for risk exploitation that takes place in many phases. The time between the detection and exploitation of the vulnerability is limited by the memory kernel function over a finite period of time.

4.8 Session Hijacking Counter Measures

It is said that the attacker takes advantage of the lack of awareness of the user regarding the security of their information or sometimes fools the user to steal their information. The corrective measures for the session hijacking are divided into two layers from the OSI (Open System Interconnection) model:

- Network Layer
- Application Layer

4.8.1 Network Layer

A very crucial step to safeguard our information from getting stolen is always using the Secure Socket Layer that provides end-to-end data encryption. These SSL channels use a 28-bit public key and a 256-bit symmetric key to make encryption tougher and more secure. Even if the attacker gets his hand on the data it will be very difficult to get real data in the pockets.

The use of Secure Socket Shell (SSH) provides a robust authentication and encryption mechanism between two systems on an unprotected network, which helps keep users away from any type of session attack. It is also very important to note that, whenever we visit a website, we must use an HTTPS connection.

4.8.2 Application Layer

According to Sabottke et al. [12] Session ID provides a unique identity for each session and is useful to track user progress and authentication status in the web application. Having a strong and sophisticated Session ID will make it difficult for the attacker to access it. Using a long Session ID by generating one from a random Session ID generator also makes it difficult for the attacker to guess the ID.

4.9 Privacy Set-Up on Social Networking Sites

Long-range social media platforms like Facebook provide many types of privacy settings such as not displaying individual data, for example, mobile number, email ID and so on. These must be updated from time to time from the user's end.

5 Conclusion

In this exploratory analysis, we walked through different possible threats to the security of social media. As more and more users are joining these platforms, it has become a primary cause of lookout for malicious actors. With the increasing amount of data, comes the grave responsibility to keep it secure and this gives a lot of benefits to the attackers while planning to extract valuable and personal information from their target's account. To curb these extensively outspread cybercrimes such as data theft and identity theft, several existing solutions were described in detail. To conclude, social media can become more secure if the users are educated about cybersecurity along with the ongoing efforts made by researchers and analysts to create a secure online environment.

References

1. Das, S., Das, S., Bandyopadhyay, B., Sanyal, S.: Steganography and steganalysis: different approaches (2011)
2. Stokes, K., Carlsson, N.: A peer-to-peer agent community for digital oblivion in online social networks (2013)
3. Mishra, S., Thakkar, H.K., Mallick, P.K., Tiwari, P., Alamri, A.: A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection. *Sustain. Cities Soc.* **72**, 103079 (2021)
4. Mishra, S., Panda, A., Tripathy, K.H.: Implementation of re-sampling technique to handle skewed data in tumor prediction. *J. Adv. Res. Dyn. Control Syst.* **10**, 526–530 (2018)
5. Druschel, P., Backes, M., Tirta, R.: The right to be forgotten—between expectations and practice, Deliverable, ENISA, November 2012 (2012)
6. Weaver, N., Paxon, V., Staniford, S., Cunningham, R.: A taxonomy of computer worms. In: *Proceedings of the 2003 ACM Workshop on Rapid Malcode* (2003)
7. Wang, Y.M., Beck, D., Vo, B., Roussev, R., Verbowski, C.: Detecting stealth software with strider ghostbuster. In: *Proceedings of the 2005 International Conference on Dependable Systems and Networks*, pp. 368–377 (2005)
8. Forrest, S., Perelson, A.S., Allen, L., Cherukuri, R.: Self-nonsel self discrimination. In: *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, May 1994 (1994)
9. Tripathy, H.K., Mishra, S., Suman, S., Nayyar, A., Sahoo, K.S.: Smart COVID-shield: an IoT driven reliable and automated prototype model for COVID-19 symptoms tracking. *Computing* 1–22 (2022)

10. Mishra, S., Mishra, B.K., Tripathy, H.K.: A neuro-genetic model to predict hepatitis disease risk. In: 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), pp. 1–3. IEEE (2015)
11. Aggarwal, A., et al.: Phishari: automatic realtime phishing detection on Twitter (2013)
12. Sabotke, C., Suci, O., Dumitras, T.: Vulnerability disclosure in the age of social media: exploiting twitter for predicting real-world exploits. In: Proceedings of the 24th USENIX Security Symposium (USENIX Security 15), pp. 1041–10567 (2015)
13. Sapienza, A., Ernala, S.K., Bessi, A., Lerman, K., Ferrara, E.: DISCOVER: mining online chatter for emerging cyber threats. In: Proceedings of WWW Companion for the Third International Workshop on Computational Methods for CyberSafety, pp. 983–990 (2018)
14. Hartung, F., Kutter, M.: Multimedia watermarking techniques. Proc. IEEE (1999)
15. Tripathy, H.K., Mallick, P.K., Mishra, S.: Application and evaluation of classification model to detect autistic spectrum disorders in children. Int. J. Comput. Appl. Technol. **65**(4), 368–377 (2021)
16. Mishra, S., Dash, A., Ranjan, P., Jena, A.K.: Enhancing heart disorders prediction with attribute optimization. In: Advances in Electronics, Communication and Computing, pp. 139–145. Springer, Singapore (2021)
17. Choudhury, B., Das, R., Baruah, A.: A Novel steganalysis method based on histogram analysis. In: Lecture Notes in Electrical Engineering, pp. 3–4, November 2015
18. Baitha, A.K., Vinod, S.: Session hijacking and prevention technique. Int. J. Eng. Technol. **7**(2.6), 193–198
19. Mishra, S., Tripathy, H.K., Thakkar, H.K., Garg, D., Kotecha, K., Pandya, S.: An explainable intelligence driven query prioritization using balanced decision tree approach for multi-level psychological disorders assessment. Front. Pub. Health **9** (2021)
20. Mallick, P.K., Mishra, S., Mohanty, B.P., Satapathy, S.K.: A Deep neural network model for effective diagnosis of melanoma disorder. In: Cognitive Informatics and Soft Computing, pp. 43–51. Springer, Singapore (2021)
21. Alabrah, A., Bassiouni, M.: Preventing session hijacking in collaborative applications with hybrid cache supported one-way hash chains. In: 2014 International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom). IEEE (2014)
22. Jain, V., Sahu, D.R., Tomar, D.S.: Session hijacking: threat analysis and countermeasures. In: International Conference on Futuristic Trends in Computational Analysis and Knowledge Management (2015)
23. Sauerwein, C., Sillaber, C., Huber, M.M., Mussmann, A., Breu, R.: The tweet advantage: an empirical analysis of 0-day vulnerability information shared on Twitter. In: Janczewski, L., Kutylowski, M. (eds.) ICT Systems Security and Privacy Protection, pp. 201–215. Springer (2018)
24. Mondal, S., Tripathy, H.K., Mishra, S., Mallick, P.K.: Perspective analysis of anti-aging products using voting-based ensemble technique. In: Advances in Systems, Control and Automations, pp. 237–246. Springer, Singapore (2021)
25. Mohapatra, S.K., Mishra, S., Tripathy, H.K., Bhoi, A.K., Barsocchi, P.: A pragmatic investigation of energy consumption and utilization models in the urban sector using predictive intelligence approaches. Energies **14**(13), 3900 (2021)
26. Anand, A., Bhatt, N., Kaur, J., Tamura, Y.: Time lag-based modelling for software vulnerability exploitation process (2021)
27. Huber, M., Kowalskiy, S., Nohlbergz, M., Tjoa, S.: Towards automating social engineering using social networking sites, Proceedings of International Conference on Computational Science and Engineering (2009)
28. Kumar, S., Saravanakumar, N., Deepa, K.: On privacy and security in social media—a comprehensive study. In: International Conference on Information Security & Privacy (ICISP2015), 11–12 December 2015, Nagpur, India
29. Tripathy, H.K., Mishra, S., Thakkar, H.K., Rai, D.: Care: a collision-aware mobile robot navigation in grid environment using improved breadth first search. Comput. Electr. Eng. **94**, 107327 (2021)

30. Wang, Z., Huang, D., Zhu, Y., Li, B., Chung, C.-J.: Efficient attribute-based comparable data access control. *IEEE Trans. Comput.* **64**(12), 3430–3443 (2015)
31. Wu, M.-Y.: On runtime parallel scheduling for processor load balancing. *IEEE Trans. Parallel Distrib. Syst.* **8**(2), 173–186 (1997)
32. Tian, W., Zhao, Y., Xu, M., Zhong, Y., Sun, X.: A toolkit for modeling and simulation of real-time virtual machine allocation in a cloud data center. *IEEE Trans. Autom. Sci. Eng.* **12**(1):153–161 (2015)
33. Raja, A.S., Vasanthi, A.: Secured multi-keyword ranked search over encrypted cloud data. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2**(10) (2012)
34. Mishra, S., Tripathy, H.K., Mallick, P.K., Bhoi, A.K., Barsocchi, P.: EAGA-MLP—an enhanced and adaptive hybrid classification model for diabetes diagnosis. *Sensors* **20**(14), 4036 (2020)
35. Chattopadhyay, A., Mishra, S., González-Briones, A.: Integration of machine learning and IoT in healthcare domain. In: *Hybrid Artificial Intelligence and IoT in Healthcare*, pp. 223–244. Springer, Singapore (2021)
36. Ereth, J.: If data is the new oil, metadata is the new gold, 12 April 2017. <https://www.eckerson.com/articles/if-data-is-the-new-oil-metadata-is-the-new-gold>
37. Black Slash: How to hide secret data inside an image....., 9 July 2018. <https://null-byte.wonderhowto.com/how-to/steganography-hide-secret-data-inside-image-audio-file-seconds-0180936/>
38. The Ultimate Guide to Session Hijacking aka Cookie Hijacking, 16 November 2020. <https://www.thesslstore.com/blog/the-ultimate-guide-to-session-hijacking-aka-cookie-hijacking/>
39. What is a Random Forest? <https://www.tibco.com/reference-center/what-is-a-random-forest>

A Pragmatic Analysis of Security Concerns in Cloud, Fog, and Edge Environment



Manish Jena, Udayan Das, and Madhabananda Das

Abstract With the emergence of Fog and Edge architecture, optimization has become a significant aspect of Cloud computing. Not only do these changing architecture necessitate re-evaluating cloud-native optimizations and uncovering Fog and Edge-based outcomes, but the goals also necessitate a significant shift from focusing just on latency to focusing on energy, security, dependability, and cost. As a result, it appears that optimization targets have become broader, with the Internet of Things (IoT)-specific objectives emerging recently. Furthermore, in certain applications that need low latency, the delay generated by transferring data to the cloud and subsequently back to the application can have a significant impact on their performance. Existing IoT designs are becoming increasingly centralized, relying heavily on cloud solutions for data processing, analytics, and decision-making. This survey highlights the main security and privacy challenges that fog and edge computing confront, as well as in what way security concerns may influence the development and usage of edge and fog computing.

Keywords Cloud computing · Edge computing · Fog computing · Security issues · Edge nodes

1 Introduction to Cloud Computing

Cloud computing has risen to prominence as the most popular data storage and processing platform in recent years. It has extended to a variety of industries, including healthcare, real estate, banking, manufacturing, and so on. Instead of maintaining data on their local systems, businesses preserve it on the cloud. The Internet of

M. Jena · U. Das · M. Das (✉)
Kalinga Institute of Industrial Technology, Bhubaneswar, India
e-mail: mndas_prof@kiit.ac.in

M. Jena
e-mail: 2005240@kiit.ac.in

U. Das
e-mail: 2005280@kiit.ac.in

Things (IoT) is among the most disruptive technologies of the previous decade, and it is at the heart of a number of emerging trends, including smart cities [1]. Latency and network bandwidth are the only problems, which arise in IoT environments. Despite its various capabilities, the cloud's security remains one of its most crucial aspects. Every cloud computing server is concerned with privacy and security in a distinct way. Cloud analysts search for loopholes in the architecture and possible attacks to exploit them for cloud safety and security [2].

The emergence of mobile cloud computing technology is due to the convergence of mobile computing with cloud computing. Using a mobile device's thin native client or web browser, these centralized programs are then retrieved across wireless networks. Processes are enabled by a centralized server that follows a collection of rules. Computer software and middleware are used to allow smooth communication between devices linked via the cloud. Data is frequently copied by cloud computing service providers to defend against security threats, data loss, and data breaches, among other things. We separate cloud systems into two groups to better understand how they function. The front end is one thing, and the back end is another. Both are connected to one another via a network, most often (Fig. 1)

the Internet. The front end is the user's or client's side of the application. All cloud computing platforms do not have to have the same UI/UX. On the back end, the cloud is made up of several computers, web servers, and storage devices. A cloud computing model may run any program, from data processing to computer games [3].

Cloud computing allows people to access critical data from afar and eliminates the need to invest in expensive computer and storage infrastructure. Cloud Service Provider provides basic infrastructure like a computer-generated interface for users to keep their data in Infrastructure-as-a-Service. Users utilize Operating systems and apps to process, network, and store data on their installed applications. This design includes hardware resources such as CPU, memory, the disc, and bandwidth. Users may run and manage their apps with Platform as a Service. They're given one base operating system, some development packages, and technology for developing applications [4]. Some resources are supervised, such as software frameworks and storage. Under Software as a Service, the cloud merchants provide all of the infrastructure and software, which is also known as 'on-demand software.

Cloud computing reduces the need for expensive computer and storage infrastructure and enables users to access data from any location. According to the demands of the users, resources are allotted dynamically. BNA (Broad Network Access) lets a person manage data from remote places using standard platforms such as smartphones, laptops, etc. Elasticity refers to the ability to scale resources up or down in response to demand [5]. For all sorts of emergency scenarios, including natural catastrophes and power outages, cloud-based systems provide speedy data recovery. Only 9% of non-cloud users claim that in less than four hours disaster recovery can be done. A global study found that around 39% of IT executives desire to invest in cloud-based emergency preparedness methods [6]. Cloud infrastructures enable virtual services to be powered rather than the actual software and hardware, lowering energy costs and cutting computer-related emissions. A cloud owner's full-time task

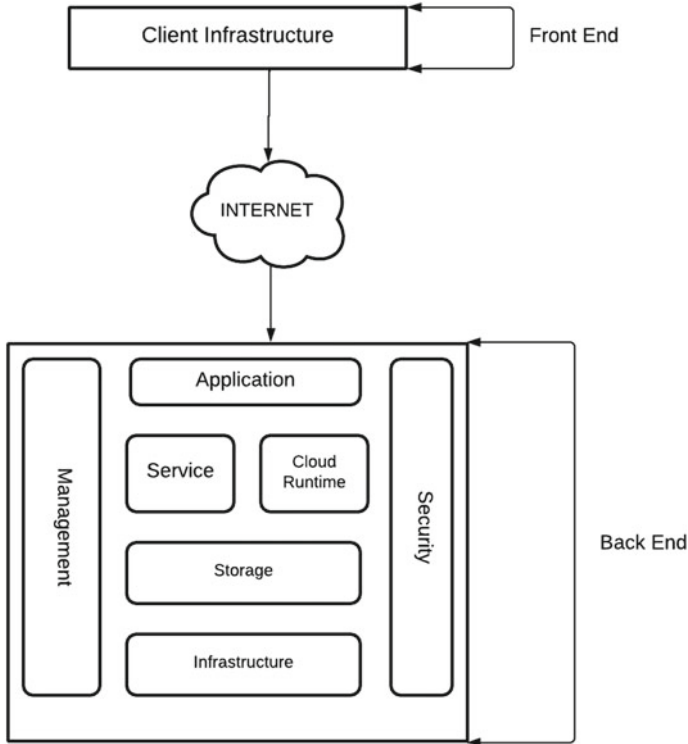


Fig. 1 Cloud computing architecture

is to thoroughly examine protection that is significantly more efficient than a standard internal system. According to Rapid Scale, 94 percent of producers observed a substantial improvement in security after transitioning to the cloud.

2 Introduction to Fog Computing

Fog computing adds a layer between the data’s origin and the cloud. It is performed at fog nodes, which gather data from a variety of edge devices. It is a platform with a lot of virtualization which gives network and storage operations between end devices and standard cloud-based servers but not at the network’s edge. Fog computing utilizes specialized networking devices known as Fog nodes to execute numerous computational activities at the network edge. Its primary characteristics are edge awareness, low latency, mobility support, real-time interaction, and heterogeneity.

All nodes aren’t kept active at all times in fog computing [7]. When the data load is diminished, the computational unit of the Fog nodes may be switched off and activated as needed. As a result, the Fog environment is extremely scalable and of

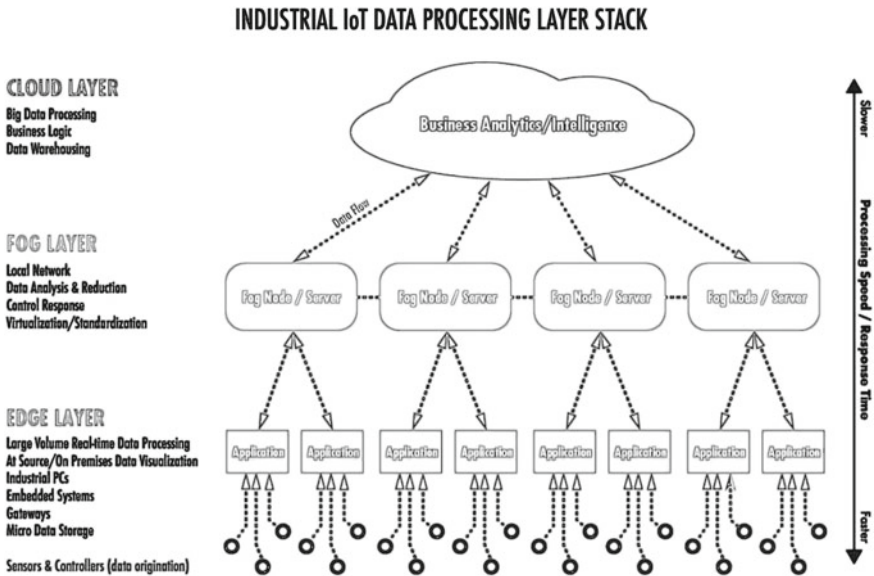


Fig. 2 Fog computing layers [6]

low-energy. In addition, security protections for data privacy and intrusion prevention can be applied to each communication channel between the nodes. As a result, secure data transmission is possible. In rare instances, the fog network may be regarded as a point-to-point network. A fog-based system design is similar to the cloudlet idea, but with a stronger focus on the overall system resilience.

Fog computing is a three-tier system, with the lowest tier consisting of edge devices such as sensors, vehicles, or apps that generate data, the second level consisting of fog nodes that help in collecting data from multiple edge devices, and the topmost layer consisting of cloud data centres that gather data from fog nodes, as shown in Fig. 2.

As a result, fog systems must work together with cloud servers in order to perform tasks like orchestration, big data analysis, and distinct service delivery, implying that fog couldn't completely substitute cloud computing, and the two must work together to offer users on-time value-added services.

3 Introduction to Edge Computing

Edge computing was initially introduced around 2002, and it was primarily used to deliver applications across Content Delivery Networks. Taking advantage of the close proximity and resources of CDN end nodes in order to achieve a lot of scalabilities is the main purpose of this. Routers, base stations, and switches make up an edge node,

which directs network traffic. In order to manage the packet data from several subnetworks they integrate, these devices perform complex processes. Edge computing, in a larger sense, may be defined as an agreed-upon strategy that additionally considers node ownership. Edge nodes, resembling cloudlets, are located near end nodes [8]. Edge Computing for Mobile was launched by the European Telecommunications Standard Institute (ETSI) in 2014 as a unique platform that provides IT and cloud computational power within the Radio Access Network close to users.

In an edge computing paradigm, data generated by devices is stored on the device or near the device rather than being transmitted to the cloud. Before delivering the information to the cloud, these gateways pre-process it. They provide a layer of protection against illegal access from other devices. The primary feature of Mobile Edge Computing is the presence of network control and storage resources at the mobile Radio Access Network's edge, with the goal of lowering latency. As a result, MEC is not a replacement for cloud computing, but rather a complement: The delay-complicated component of software can operate on Data centres, whereas the delay-tolerant compute-heavy component of an application can run on a cloud server [9]. Recently, fog computing was proposed as an extension of MEC, and edge devices are now defined as anything from smartphones to set-top boxes. These two regularly conflict & terminology is frequently interchanged (Fig. 3).

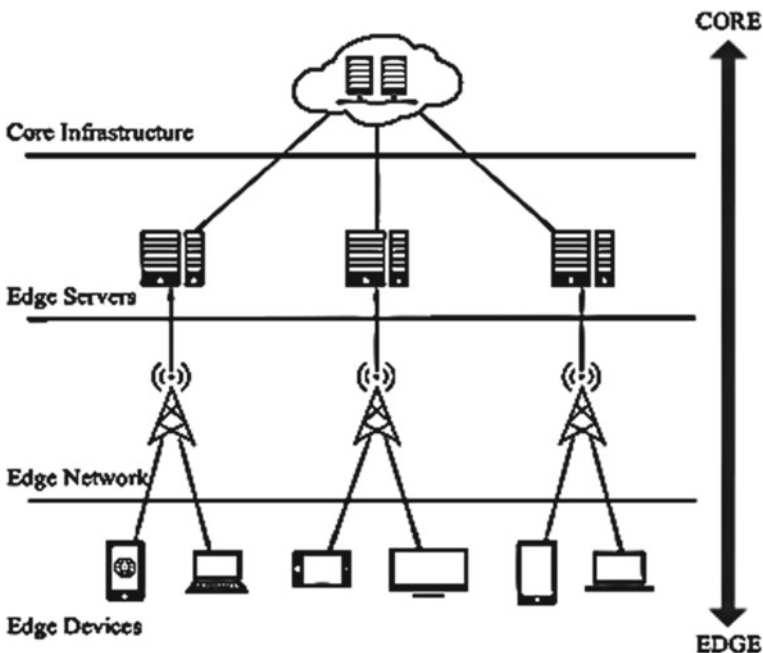


Fig. 3 Edge computing architecture [17]

In manufacturing, adaptive diagnostics in an industrial setting can advance the uptime of systems and equipment, deducting service expenditures. Edge-compute-generated fault codes get combined with historical repair information which can provide the framework for technicians, speeding up troubleshooting and repairs [10]. Computation on the edge permits public infrastructures and facilities to be checked for greater productivity in lighting, heating, etc. In traffic management systems, cameras and signals can mend safety and traffic flow. In the coming years, self-driving vehicles, where near-zero latency is critical, will be the most noticeable and dramatic examples of real-time edge computing. Smart wearable devices can store data on heart rate, temperature, and other metrics, then offer reminders for medication. Also, edge computing enables creators to ensure sensitive data, such as therapeutic imagery, which does leave the device to boost security and confidentiality. Cloud gaming establishments are looking to create edge servers in close proximity to gamers in order to decrease latency and deliver a fully receptive and immersive gaming experience.

It is preferable to cloud computing since it allows for real-time data processing with little latency. There will be less network traffic and faster data processing as a result. By filtering out sophisticated data and delivering it to the data centre at the edge device, the data's safety is increased. Low latency and bandwidth limits are important aspects of edge computing. The effect of low bandwidth at a certain site is reduced when the burden is moved closer to the user. Data analytics, Network Function Virtualization (NFV), and web monitoring are some of the examples of edge computing applications. Performing NFV on the edge layer increases efficiency while cutting expenses. On the gadget, data may be examined, and a single data summary should be stored in a centralized cloud.

4 Security Threats of Cloud Fog and Edge Computing

A plethora of new security concerns and risks was introduced in the cloud. Data is stored in the cloud by an unauthorized provider which is accessible over the internet, restricting both discernibility and handling of data. It is vital for all people to understand their role and the safety risks that cloud computing entails. The risks and difficulties related to cloud security are primarily the responsibility of cloud providers. Consumers are responsible for the security of their data stored on the cloud, whereas the cloud service provider is in charge of the cloud's security. Every cloud computing user is constantly responsible for safeguarding and controlling access to their data [11].

There are three services, namely:

- I. **Software as a Service (SaaS):** Cloud security concerns in this service are unquestionably data and access-related. 'On-demand software' is another term for this service. Every company should be worried about the sort of data it sends to the cloud, who has access to it, and what fortifications it has in place [12]. The

SaaS provider's role as a possible access point to the administration's data and procedures should also be evaluated. All of these occurrences demonstrate that attackers regard software and cloud providers as a tool to target more resources [6]. As a result, attackers are concentrating their efforts on this potential flaw.

Below is a list of security issues in this service:

- Authentication and authorization are ineffective.
- Data breaches and data losses.
- There is a lack of staff with the requisite skills to manage cloud application security.
- Data to and from cloud apps cannot be monitored.
- Identity supervision on the cloud is immature.

II. **Infrastructure as a Service (IaaS):** In this, protecting information is risky. When consumer concern spreads to apps, OS, and network traffic, new risks develop. Administrations must analyse the current evolution of risks that extend beyond data as the focal point of IaaS risk. Clients obtain the hardware configuration as a virtual medium to host their information from the Cloud Service Provider. Users may create their own OS and apps, as well as process, network, and store data on them. Analysing the ability to prevent theft and govern access when developing cloud infrastructure is risky. Choosing who can send data to the server, tracking resources varies in order to detect anomalous behaviour, protecting and toughening instrumentation gears, and Increasing network traffic evaluation as a possible indicator of infiltration are all swiftly emerging common processes in defending large-scale cloud installations. Bad actors grab computing resources to mine bitcoins, then reuse such assets to target different aspects of the firm.

Some of the security issues in this service are:

- An assault that spreads from one cloud capacity to another.
 - Unable to supervise the cloud activities and applications.
 - Observing a virtual computer from the perspective of another virtual machine.
 - Using the host computer to inspect virtual machines.
 - Malicious actors are stealing data hosted in cloud infrastructure.
- III. **Platform as a Service (PaaS):** This solution enables organizations to create, manage, and accomplish Web applications without having to invest in costly setups. Consumers are given a base operating system, development tools, and the infrastructure they need to create apps. Providers of PaaS might specialize in a variety of fields. Database-specific PaaS providers exist, as well as a newer form known as the highly efficient application PaaS, which generates applications utilizing a graphical, low-code approach. PaaS includes infrastructure such as servers, storage, and networking, the same as IaaS does. Database management systems, middleware, programming tools, and business intelligence elements are all addressed.

- Deficiency of Secured Software Development Process with CSPs.
- Legacy software's given by the merchants.
- PaaS Platform's security controls and self-service rights aren't properly organized.
- Hackers can gain unauthorized access and modify configurations.
- Insufficient necessities in Service Level Agreement (SLA).

Computing in fog has recognized unique archives in the contemporary communication era by overcoming fundamental technological difficulties and restrictions in cloud computing. However, this technology is expected to pose a number of security and privacy risks in relation to facts and services. Several academics have planned literary works alleging that security vulnerabilities are prevalent in fog computing. As illustrated in Fig. 4, the changing characteristics for fog computing, like spatial dispersion, motility, as well as variability, prevent current cloud computing security and privacy solutions from working in a fog computing network. In the future, new cutting-edge security technologies will be necessary to solve the security and privacy concerns highlighted by fog computing [13]. Although fog computing offers advantages over cloud computing, there are a number of security problems that may prevent future fog-based systems from being implemented (Fig. 5).

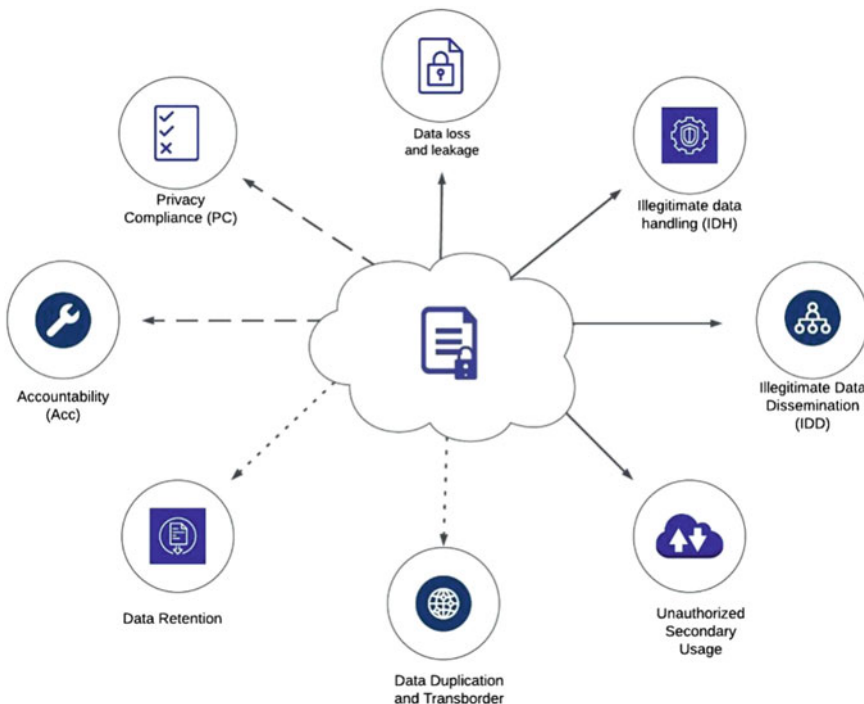


Fig. 4 Privacy in the cloud computing environment

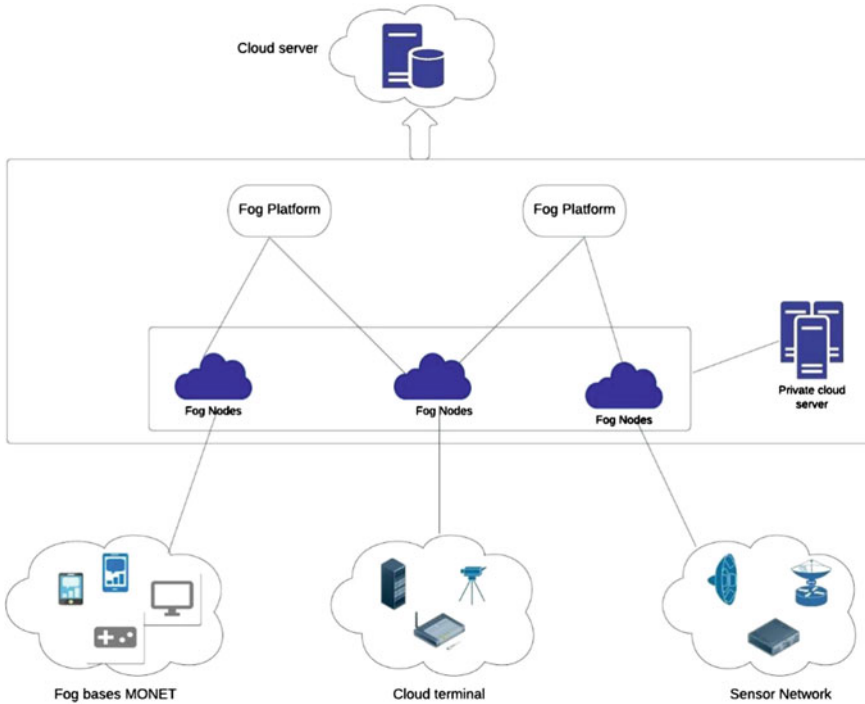


Fig. 5 The architecture of fog computing

Some privacy and security concerns are:

- Malicious fog node problems.
- Ways of attack recognition are ineffective.
- Multitenancy issues.
- Data recovery and backup are difficult when a system outage happens.
- Risky communication meetings between devices.
- Virtualization issue.

Besides the above privacy and security threats, there are other kinds of attacks in Fog Computing namely:

- (a) **Tampering:** In this type of attack, attackers playfully modify the information to be communicated. It's difficult to detect this intrusion because data transfer may get delayed due to the flexibility and the transmission route's wireless nature in end-user.
- (b) **Spam:** It is related with the unwanted data created by hackers, such as erroneous data received from people's computers and supplementary data. Spam causes enormous resource consumption in the network and deluding.
- (c) **Virtual Machine Based Attack:** This is an exploit in which a hacker steals control of the hypervisor and uses it to create a virtual atmosphere within a

virtual system. On a virtual machine, there are four basic forms of attack: guest to host, virtual machine to virtual machine manager, virtual machine to virtual machine, and virtual machine to virtual machine manager external attacks.

- (d) **Denial of Service (DoS):** It is a well-known exploit that can have an impact on a range of scenarios, including but not limited to fog computing, in which attackers send bogus information to fog nodes, which are then bombarded with limitless fake requests, rendering them unavailable to real clients.
- (e) **Sybil:** Network attackers occasionally utilize bogus identities to manipulate fog computing's efficacy and performance, as well as the consistency of the nodes. This type of attack is known as a sybil attack. The attackers generate entirely untrustworthy crowd-sensing reports. They are also capable of hiding a legitimate end user's personal information.
- (f) **Collusion:** This type of security breach includes more than two organizations conspiring to deceive and cheat lawful end users. They trick and attack a group of fog nodes, fog nodes combined with IoT nodes, or IoT nodes combined with cloud nodes in order to maximize the assault's effectiveness.

Edge computing design, like Fog Computing, is plagued by major security issues, putting customers' private or confidential data at risk, current edge computing security and privacy challenges, as well as encryption techniques and solutions [8]. Edge computing faces two major challenges: First, majority of the edge nodes are linked to a huge number of IoT devices with limited resources and different core components, resulting in a variety of routing processes to broadcast messages. This type of modification in the components might result in certain security issues. As a result, there may be a number of challenges with access control in the IoT context [14]. The key management of communications is the second problem that has some people concerned about edge computing security. While edge computing can enable end-to-end connection for IoT devices via multiple routing protocols, data concealment and integrity remain a concern. To address these problems, a superior key management and circulation method should be created.

Before we go into the security risks of Edge Computing, let's have a look at some of the factors that lead to these vulnerabilities in edge computing networks, which expose end users' personal information:

- I. Edge nodes are closer to consumers in edge computing, which might result in the receiving of a large volume of sensitive data. If any of this information is taken, the repercussions will be disastrous.
- II. In comparison to cloud computing, edge computing only has a portion of the network resources, hence advanced encryption methods aren't available.
- III. The dynamic environment of the edge computing network is always changing. As a result, assailants may easily blend in with the group. Furthermore, developing security mechanisms for such a dynamic network is quite difficult.

Some of the common privacy and security risks are as follows:

- (a) **Data storage, backup, and protection risks:** As previously stated, data stored on the outside lacks the physical safety barriers seen in data centres. In fact,

it is possible to steal an entire database by removing the disc from an edge computing resource or inserting a memory stick to transfer data.

- (b) **Perimeter defense risks:** Edge computing obstructs the entire perimeter protection as it widens the IT perimeter. It's possible that edge systems may be needed to validate their apps with data centre partner apps, and the authorizations for this are commonly maintained at the edge. This means that a compromise of edge security might disclose access credentials to data centre assets, posing a severe security risk. Threats might be more difficult to deal with because security capabilities may be constrained at the edge due to changes in hosting architecture when dealing with perimeters.
- (c) **Distributed Denial of Service (DDoS) Attack:** It's a type of attack in which the attacker uses distributed resources, such as a set of compromised edge devices, to compromise the routine services offered by numerous servers. When an attacker sends an unlimited number of packets to the victim's device from a hacked distributed device, it drains the target's hardware resources, making it impossible to handle any other packet. As a result, any valid request is not met on time.
- (d) **False Data Injection:** False data injection is a network attack in which an intruder injects false code that collects all stored data from the database and transfers it to the attacker.
- (e) **Physical Attack:** It arises when the physical protection provided by edge technology is insufficient or shoddy. Physical threats will impair services in particular geographical zones as the deployment of edge servers spreads.
- (f) **Cloud adoption risks:** Because cloud computing is still a new topic in IT, the hazards connected with edge computing in conjunction with cloud computing are very important. The risks are determined by the precise interaction between the edge and the cloud, which is easy to ignore because different cloud computing platforms and services address edge aspects differently. If the edge devices are simple supervisors, granting them secure access to cloud resources and apps may be problematic, needing a detailed examination of the cloud-to-edge connection, access control, and overall security protocols.

5 Potential Solution of Cloud Fog and Edge Computing

Security issues in cloud computing must be addressed effectively. If proper measures are not implemented, the cloud environment will grow increasingly exposed to hackers and invaders. To address the issues, the following are some of the solutions that should be considered while thinking about cloud computing [15].

- **Data Encryption:** When it comes to data security, encryption is said to be a more secure way. Before sending data to the cloud, it should be encrypted. The owner of the data can provide access to certain members.
- **Security check events:** We must guarantee that the cloud service provider provides specifics on promise fulfilment and problem reporting. These security

check events will reveal the cloud computing service provider's accountability and actions.

- **Monitor the Data Access:** Cloud service providers must assure who, when, and for what purpose data is being utilized. For example, several websites have a security issue with attackers listening to phone conversations, reading emails, and accessing personal data, among other things.
- **Taking the backup of the data:** The cloud service provider should back up the data that is kept in their cloud, but only with the user's consent. That backup data will be useful in the event of a data breach or loss.
- **Access control:** Multi-factor authentication, strong passwords, and automated key rotations should all be used by cloud service providers to safeguard and control access. If these measures are implemented, there will be many fewer instances of unwanted access, and the threat of data theft from cloud data centres will be better managed.
- **Use of firewall:** In spite of the fact that new techniques of safeguarding networks have grown popular in recent years, an efficient firewall is still the best approach for preventing unwanted access. It may be feasible to prevent malevolent hackers from gaining access to the server by taking extra precautions to ensure that the firewall only allows as much access as is required.
- **Password security:** When it comes to cloud server security, the password is the most important component. A common error made by many individuals when setting up a cloud is being careless with passwords. Because the cloud is based on trust, a single compromised password might destroy it.
- **Restricted Access:** Restricted user access entails securing some login forms using basic username/password protection and a challenge-response test. When a user or employee no longer needs access to the cloud data centre, their data centre access credentials must be withdrawn promptly.

Fog computing is now in its infancy, and there is still a long way to go. Because of its distinct characteristics, this computing faces a variety of obstacles. As we all know, it takes advantage of user devices' idle resources, which are not completely reviewed by any standard body, causing security and privacy risks in the fog network. Because multiple devices are involved in fog application processing, safe and quick authentication techniques are required.

Some of the methods proposed are:

- **Authentication plan:** The user's identity can be authenticated by comparing the user credentials to the information stored in the database through an authentication server. This aids in the protection against hostile invasions. If a user is completely authorized in the system, fog computing allows them to access fog nodes from the fog infrastructure.
- **Blockchain Security:** The blockchain idea was created to provide safe transactions in cryptocurrency applications. However, as time went on, everyone realized that the blockchain method was the way to go, with its exceptional security properties, may also be utilized to safeguard computing networks. As a result, the blockchain approach can improve the security of the fog environment.

- **Decoy Technique:** It's a method for verifying a user's data on a network. It substitutes the user's genuine information with a phony version, which is subsequently sent on to the attackers. When a hacker compromises the system's security, it discovers a bogus file in lieu of the real one. The decoy file is the name of this file, and the decoy methodology is the name of the approach. To provide greater security, fake files are produced from the start. The system hides the genuine data, which can only be viewed by authorised users, and replaces it with a decoy file by default for system intruders.
- **Modified Decoy Technique:** It's an improved and updated form of the original decoy strategy, in which attackers are provided fictitious data and system or user nodes to exploit, while the hidden file collects information about their identification, such as their Mac address or IP address.
Edge computing, as demonstrated in this study, has unique properties, therefore security solutions developed for cloud and fog computing services are ineffective with edge computing. As a result, some proposed solutions that may be implemented using edge computing's unique qualities include [16]:
- **Full-time Monitoring:** To protect computing from bad users or hackers, it's critical to keep a continual check on all edge networks and nodes and to give network awareness to all users through an interactive interface.
- **Cryptographic Techniques:** To cope with security breaches perpetrated by hackers or intruders, cryptographic measures are employed. In these methods, a secret key is employed that is only shared by the sender and the recipient. This secret key is used to decode the message that was received. However, if a thief could gain this secret key from the network's delivered packets, he could also steal the data included in the message.
- **Data Confidentiality:** Several data confidentiality strategies based on encryption techniques have been suggested to address the numerous privacy issues caused by network attackers' illegal data activities, data loss, data manipulation, data breach, and so on. The author of the article [17] suggests Query Guard, a latency-aware query optimization tool, as a privacy-preserving solution. This method fulfils two objectives: first, it handles the problem of privacy-aware distributed query processing, and second, it optimises requests for transmission without delay. When compared to typical query optimization methods, it yields better results in terms of computation time and memory use.
- **Edge Node Security:** Edge node security refers to ensuring that all nodes in the edge network have the same level of security and that suitable safety standards are followed. In the event of varying security levels, a hacker may be able to get past the node with the weakest protection, triggering a system issue [18]. Furthermore, numerous security levels might make it difficult for system administrators to determine which node has inadequate security, allowing a security compromise.
- **Proper Encryption:** New state-of-the-art encryption algorithms that are incredibly difficult to decrypt are being employed as modern technology advances. These algorithms use a highly secure secret key that is only shared by the sender and receiver. Only real users have access to this secret key, which allows them to decrypt the file and read the data [19].

- **User Behaviour Profiling:** It is the practice of watching and tracking a user's activity in order to detect any divergence from normal behaviour that might indicate the presence of a malicious user.

6 Conclusion and Future Scope

As a consequence of our research, we've determined that Cloud Fog and Edge Computing is an ever-growing industry, but that as data and devices rise, we'll need to improve our cloud system. Because the entire system of systems is only as safe as its weakest component, network and data security are major concerns that can only be addressed by employing secure hardware and conventional security techniques in the cloud. Machine learning is used in all of these processes to analyse, transform, and categorize data. To date, much of the research in this field has concentrated on the cloud as the execution environment. There is still research that can be done to lower the cloud's latency and bandwidth requirements even more without affecting the system's security. After all of the requirements have been met, the system should be able to operate on its own. For efficient resilience orchestration in modern systems, there is a critical requirement to develop a logical knowledge of the capabilities of nodes and roles, and at the same time, research on security and privacy in this context must be deeply done, since these are vital elements for cloud computing to gain the trust of their users. Finally, after initial configuration of the cloud to meet all of the requirements, it should operate without the need for human intervention. Cloud computing will probably be replaced by fog and edge computing in the future. More research may be done in this sector to improve latency without affecting the user's privacy.

References

1. Alwarafy, A., Al-Thelaya, K.A., Abdallah, M., Schneider, J., Hamdi, M.: A survey on security and privacy issues in edge-computing-assisted Internet of Things. *IEEE Internet Things J.* **8**, 4004–4022 (2020)
2. Badidi, E., Ragmani, A.: An architecture for QoS-aware fog service provisioning. *Procedia Comput. Sci.* **170**, 411–418 (2020)
3. Mebrek, A., Merghem-Boulahia, L., Esseghir, M.: Efficient green solution for a balanced energy consumption and delay in the IoT-fog-cloud computing. In: *Proceedings of the 2017 IEEE 16th International Symposium on Network Computing and Applications (NCA)*, Cambridge, MA, USA, 30 October–1 November 2017, pp. 1–4
4. Marbukh, V.: Towards Fog Network Utility Maximization (FoNUM) for managing fog computing resources. In: *Proceedings of the 2019 IEEE International Conference on Fog Computing (ICFC)*, Prague, Czech Republic, 24–26 June 2019, pp. 195–200
5. Naha, R.K., Garg, S., Georgakopoulos, D., Jayaraman, P.P., Gao, L., Xiang, Y., Ranjan, R.: Fog computing: survey of trends, architectures, requirements, and research directions. *IEEE Access* **6**, 47980–48009 (2018)
6. <https://www.winsystems.com/cloud-fog-and-edge-computing-whats-the-difference/>

7. Zeyu, H., Geming, X., Zhaohang, W., Sen, Y.: Survey on edge computing security. In: Proceedings of the 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Fuzhou, China, 12–14 June 2020; pp. 96–105
8. Suma, V., Bouhmala, N., Wang, H.: Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN 2020; Springer: Berlin/Heidelberg, Germany (2021)
9. Novak, M., Shirazi, S.N., Hudic, A., Hecht, T., Tauber, M., Hutchison, D., Maksuti, S., Bicaku, A.: Towards resilience metrics for future cloud applications. In: Proceedings of 6th International Conference Cloud Computing. Services Science, vol. 1, pp. 295301 (2016)
10. Mariani, L., Monni, C., Pezze, M., Riganelli, O., Xin, R.: Localizing faults in cloud systems. In: Proceedings of IEEE 11th International Conference Software Testing, Verification Validation (ICST), April 2018, pp. 262273 (2018)
11. Mishra, S., Thakkar, H.K., Mallick, P.K., Tiwari, P., Alamri, A.: A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection. *Sustain. Cities Soc.* **72**, 103079 (2021)
12. Tripathy, H.K., Mishra, S., Thakkar, H.K., Rai, D.: Care: a collision-aware mobile robot navigation in grid environment using improved breadth first search. *Comput. Electr. Eng.* **94**, 107327 (2021)
13. Aldossary, S., Allen, W.: Data security, privacy, availability, and integrity in cloud computing: issues and current solutions. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, vol. 7, no. 4, pp. 485–498
14. Tripathy, H.K., Mishra, S., Suman, S., Nayyar, A., Sahoo, K.S.: Smart COVID-shield: an IoT driven reliable and automated prototype model for COVID-19 symptoms tracking. *Computing* 1–22 (2022)
15. Tripathy, H.K., Mallick, P.K., Mishra, S.: Application and evaluation of classification model to detect autistic spectrum disorders in children. *Int. J. Comput. Appl. Technol.* **65**(4), 368–377 (2021)
16. Rajegore, P.B., Kadam, S.G.: Issues & solution of SAAS model in cloud computing. *IOSR J. Comput. Eng. (IOSR-JCE)* 40–44
17. Zhang, J., Chen, B., Zhao, Y., Cheng, X., Hu, F.: Data security and privacy-preserving in edge computing paradigm: survey and open issues. *IEEE Access* **6**, 18209–18237
18. Mishra, S., Tripathy, H.K., Thakkar, H.K., Garg, D., Kotecha, K., Pandya, S.: An explainable intelligence driven query prioritization using balanced decision tree approach for multi-level psychological disorders assessment. *Front. Pub. Health* **9** (2021)
19. Mishra, S., Tripathy, H.K., Mallick, P.K., Bhoi, A.K., Barsocchi, P.: EAGA-MLP—an enhanced and adaptive hybrid classification model for diabetes diagnosis. *Sensors* **20**(14), 4036 (2020)

Secure Information and Data Centres: An Exploratory Study



Pranav Pant, Kunal Anand, and Djeane Debora Onthoni

Abstract Getting delicate information is the objective of the overwhelming majority. Cyber-attack programs target data driven information because majority of strategic and touchy information are available there. Thus, associations should focus on data set security, and the initial step is information knowledge—knowing what touchy information one has, how their data set framework is designed, and who approaches it. It involves a common sense that the web isn't secure. Many occasions have shown that there are individuals in this enormous interconnection of organizations that need to, with different aims, take others' data, disturb the administration of an overall specialist co-op, and assault frameworks to get entrance or to cut them down. Network security has been a principal component of each association to guarantee secure web availability and insurance against information breaks. While numerous associations have turned towards data centre specialists to save their time and effort on obtaining, establishing and securing of equipments, servers, and gadgets, data centres themselves are not secure from hooligans on the web. It is time for the Data Centre to demonstrate its reliability to clients by getting their information and disconnection from different clients that share a similar framework and offering continuous assistance with a base measure of personal time. To get Data Centres organizations and forestall information breaks, various sellers and Data Centre experts have proposed different arrangements, of which some have been examined in this paper. Besides, as Data Centre innovation has been created to adjust to mechanization through programming reflection, virtualization has become an indivisible piece.

Keywords Network security · Data centre · Cyber-attack

P. Pant (✉) · K. Anand

Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India
e-mail: Pranavpant26@gmail.com

K. Anand

e-mail: Kunal.anandfcs@kiit.ac.in

D. D. Onthoni

Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes (NHRI), Miaoli 350012, Taiwan
e-mail: Djeane@nhri.edu.tw

1 Introduction

Current organizations use PCs in practically all parts of carrying on with work correspondence, data capacity, bookkeeping, and everyday business capacities. A data centre is a brought together virtual office where corporate PCs, organization, stockpiling, and other IT hardware help business tasks live. The PCs in a data centre contain or work with business-basic applications, administrations, and information. Data centres come in all sizes—they might fill a storeroom, a committed room, or a stockroom. A few organizations with IT hardware in their data centres might require more than one data centre office [1]. Likewise, organizations can lease server space and have another person keep up with their data centre. A data centre could reach out outside a virtual office by utilizing a private or public cloud to expand its activities or capacity. A virtualized data centre can involve servers in distant areas when expected to run more enormous responsibilities. The boom of data centres came during the website air pocket of 1997–2000. A quick Internet, along with the relentless activity, was the need of every organization to convey frameworks and to lay out a presence on the Internet. While it was not practical for some highly modest organizations to introduce not such gear, they began assembling huge offices, known as Internet data centres (IDCs), that gave improved capacities [2] (Fig. 1).

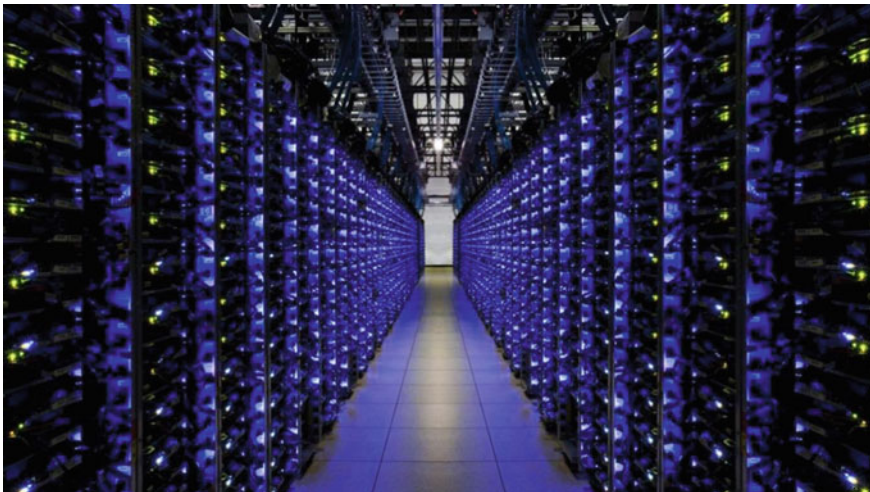


Fig. 1 Data centre [2]



Fig. 2 NASA mission control computer room c. 1962 [3]

1.1 History of Data Centre

Early PC frameworks were complex to work with and keep up with; an exceptional working climate was essential [3]. Many links were crucial to associate the parts, and techniques to oblige and put together were concocted. A single mainframe requires much force and must be cooled enough to avoid exposure to extremely high temperatures. The high cost of PC and their frequent utilization for military purposes emphasized the security aspect of the PCs. The accessibility of modest systems administration hardware, in association with new norms for the organization-organized cabling, made it conceivable to include a progressive plan to put the servers in a specific room inside the organization. The utilization of the expression “data centre” began to acquire wide acknowledgement regarding this time (Fig. 2).

1.2 Importance of Data Centres in a Business Environment

In the realm of big business IT, data centres help business applications and exercises that include the following:

1. Email and record sharing.
2. Productivity applications.
3. Customer relationship with the executives (CRM).

4. Enterprise asset arranging (ERP) and information bases.
5. Large information, computerized reasoning, and AI.
6. Virtual work areas, interchanges, and coordinated effort administrations.

2 Core Parts of a Data Centre

Data centre configuration incorporates switches, firewalls, capacity frameworks, servers, application conveyance, and regulators [4]. Since these parts store and oversee business-basic information and applications, data centre security is essential in the data centre plan. Together, they give security and privacy to data centres.

2.1 Network Infrastructure

The network infrastructure usually consists of hardware components like routers, switches, security appliances, and firewalls. These data centre resources are vital for the association and mix of the various data centre equipment frameworks. Famous brands incorporate Cisco, Brocade, Juniper, and F5 Networks, which are only the tip of the iceberg. These interfaces include servers (physical and virtualized), Data centre administrations, stockpiling, and outer availability to end-client areas [5].

2.2 Storage Infrastructure

Capacity framework alludes to IT stockpiling parts like organization appended capacity (NAS), directly joined stockpiling (DAS), substantial state drive (SSD) streak clusters, and tape capacity. Famous capacity gadget brands incorporate any semblance of HPE, Dell EMC, NetApp, and IBM. Information is the ammunition of the cutting-edge data centre. Capacity frameworks are utilized to hold this vital product.

2.3 Server Infrastructure

It refers to the rack, blade, and tower servers where data can reside and applications servers can likewise be virtualized conditions inside an actual machine, yet those are excluded from the data centre parts clarified inside this article since they are not the existing foundation.

2.4 *Computing Resources*

Applications play a crucial role and are the driving force of a data centre. These servers give the handling, memory, neighbourhood capacity, and organization network that drive applications.

2.5 *Categories of Data Centre Facilities*

The progression of the data centre framework became the reason for the growth and categorization of data centre facilities [6].

Undertaking Data Centre Facilities—Generally coordinated offices are straightforwardly possessed and worked by a solitary association. For the most part, these are situated nearby, and an in-house group administers upkeep, IT organizations, equipment redesigns, and Network observing.

Colocation Data Centres—These comprise a shared data centre arena where an association can lease headroom for servers and other equipment. The advantages of colocation versus in-house data centres are that the office gives the structure, power, HVAC, web data transfer capacity, and actual security, while the client should supply and keep up with the equipment.

Overseen Data Centre—An organization rents the actual foundation in a supervised administration data centre course of action, and an outsider-managed specialist co-op deals with the equipment and office.

Cloud Data Centre—This data centre facility has become more famous over the new years. A cloud data centre is an off-premises office that is web open; however, one has no liability regarding keeping up with the connected foundation.

3 **Requirements of a Modern Data Centre**

Because data centres contain so much expensive IT equipment, they have special requirements for security and power.

3.1 *Abundant, Reliable Power*

The gear in a data centre frequently requires much force, from an unsusceptible source to interferences through quickly accessible backup power [7]. Virtualized or programming characterized data centres are more effective and need significantly



Fig. 3 A bank of batteries in a large data centre used to provide power until diesel generators can start [7]

less energy than conventional ones. A depository of batteries in a massive Data centre provides power until the alternative power supply options can start (Fig. 3).

3.2 Cool Conditions

The entirety of the power and hardware in a data centre creates a great deal of hotness, so data centres frequently require some cooling gear to work ideally. Water can annihilate PCs, so sprinklers cannot be utilized to safeguard the hardware in a data centre from fire. Data centres can use synthetic fire-retardant frameworks covering flares without hurting electronic gear. Ordinary cold walkway arrangement is with server rack fronts confronting one another and cold air disseminated through the elevated floor [8] (Fig. 4).

3.3 Physical and Virtual Security Measures

Security is a significant part of any data centre due to the business-basic applications and data it contains. A break where touchy client or organization information becomes uncovered can cost a considerable number of dollars and obliterate an organization's image and business in most pessimistic scenarios. Physical and virtual



Fig. 4 Typical cold aisle configuration with server rack fronts facing each other and cold air distributed through the raised floor [8]

safety efforts are essential to guarantee that a data centre stays secure and that organizations are not helpless against an information break [9]. A data centre should be shielded from the robbery with actual safety efforts like locks, video observation, and limited admittance. Organization and application security programming can give fundamental virtual safety efforts.

4 Tiered Data Centres

The most generally embraced norm for data centre plan and foundation is ANSI/TIA-942. It incorporates principles for ANSI/TIA-942-prepared accreditation that guarantees consistency with data centre levels appraised for levels of overt repetitiveness and adaptation to internal failure [10].

Level 1: Basic site foundation. A Tier 1 data centre renders restricted insurance against actual occasions. It has single-limit parts along with solitary, nonredundant appropriation way.

Level 2: Redundant-limit part site foundation. This data centre offers further developed insurance against actual occasions. It has repetitive limit parts and a solitary, nonredundant conveyance way.

Level 3: Concurrently viable site foundation. This data centre safeguards against basically all occasions, giving spare parts and different autonomous dispersion ways. Every element can be taken out or supplanted without disturbing administrations to end clients.

Level 4: Fault-open-minded site foundation. This data centre gives the most significant levels of adaptation to non-critical failure and overt repetitiveness. Excessive parts and different autonomous conveyance ways empower simultaneous practicality and one issue in the establishment without causing personal time [11].

4.1 Uptime Institute

The Uptime Institute standard characterizes four Tiers:

Level I—Basic Capacity

A Tier 1 Data centre gives the fundamental limit level expected to help IT for an office. It requires an uninterruptible power supply (UPS) for blackouts, lists, and spikes; a region for IT frameworks; dedicated cooling hardware that runs outside available time; and a motor-generator for blackouts.

A Tier 1 office safeguards against human blunder yet offers restricted insurance against startling disappointments or blackouts and should close down totally for fixes and support. Accordingly, it gives 99.671% uptime, no overt repetitiveness, and will encounter 28.8 long periods of personal time each year [12].

Level II—Redundant Capacity

Level 2 Data centres offer superior security against actual occasions. They give upkeep and prosperity against aggravations through gear like cooling systems, energy generators and limits, fuel tanks, and siphons. Like a Tier I office, a startling closure influences the framework, bringing about 99.749% uptime and 22 h of personal time each year.

Level III—Concurrently Maintainable

Not at all like Tier 1 and 2 Data centres, a Tier 3 office should not be closed down when gear support or substitution is required featuring redundant parts and dissemination ways that promise it is at the same time suitable. A Tier 3 Data centre is more fit for more prominent organizations and will safeguard against most actual occasions. It offers 99.982% uptime, is N + 1 issue open-minded to give somewhere around 72 h of blackout insurance, and encounters under 1.6 long periods of personal time each year [13].

Level IV—Fault Tolerant

Level 4 Data centres highlight accessible, segregated frameworks that make spare parts and appropriation ways. This guarantees that arranged or random disturbances

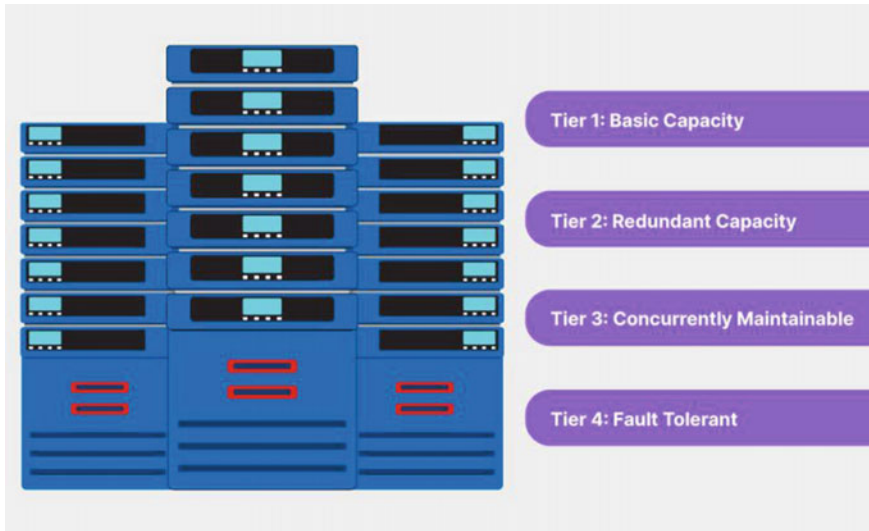


Fig. 5 Tiered data centres (uptime institute) [13]

will not influence the office and IT activities. All IT gear in a Tier 4 office should have an issue lenient power plan, and the structure requires constant cooling so the climate stays stable. A Tier 4 Data centre is ordinarily reasonable for great business partnerships. It furnishes 99.995% uptime with simply 26.3 min of yearly personal time. It additionally offers a $2N + 1$ completely excess framework and 96-h blackout assurance (Fig. 5).

5 Challenges in Data centre Networking

Data Centre organizing is the consolidation of registering administrations, including switches, load-adjusting, and examination programming to empower the assortment and appropriation of information [14].

Current Data Centre organizing difficulties present a costly effect that might reach across the associated assortment of information assets, including virtual machines, holders, and uncovered metal applications [15]. This may adversely influence the brought together observing and granular security controls.

A few difficulties in Data Centre organizing include the following.

5.1 Data Security

One reliable wellspring of data centre organizing difficulties is security. An information break could cost many dollars in lost protected innovation, private information spillage, and individual data. Focus, for instance, lost \$162 million in light of an information break. Therefore, all data centre chairmen should think about hazards to the board and safeguard both put away and appropriated information across the organization. As indicated by a study directed by the Information Management Society, 32% of CIOs positioned security as their top concern [16].

5.2 Power Management

While server solidification and virtualization decrease how much equipment is in the Data centre, they do not continuously bring down energy utilization. Notwithstanding being more proficient, edge servers consume four to multiple times the energy of past information stockpiling advancements. In addition, power and cooling prerequisites are becoming more significant as hardware necessities change [17].

5.3 Capacity Planning

Keeping up with ideal execution requires working the data centre at the most extreme limit. IT administrators frequently leave an edge for the blunder, a little security hole, to guarantee that exercises do not endure interferences. Over-provisioning is expensive and misuses space, PC handling power, and power. Data centre executives are becoming more stressed over running out of reach, so a developing number of data centres are carrying out DCIM projects to identify inactive handling, stockpiling, and cooling limits. DCIM empowers data centres to work at the most extreme limit while limiting gamble [18].

What is DCIM?

Data centre foundation infrastructure management (DCIM) is the climax of Data Centre Operations and IT that can rule the roost for an ideal Data centre execution. DCIM devices and best practices can be utilized to observe the executives of Data centre components like power dispersion components, servers, capacity equipment, and organization hardware (Fig. 6).

The meaning of foundation, notwithstanding, is advancing. By and large, it used to allude to on-premises equipment [19]. With the proceeded and expanding dependence on the cloud, the limits of customary IT foundation parts are extending. However, regardless of how one scopes it, the significant important point of the framework of the board is that it addresses the whole cluster of the executives' works, including the following:



Fig. 6 Sunbird DCIM Software [19]

- Knowing what one has.
- Deciding the qualities (What is the gauge? What is excellent? What is atypical? awful?).
- Guaranteeing uptime.

DCIM and Data centre executives can be utilized for the accompanying exercises:

- Frameworks Discovery—Network gadget disclosure can be utilized by Data centre administrators to take stock of all IT gadgets in and around the office. Associated estimation instruments for power conveyance and air quality should likewise be represented and thought of.
- Observing and Reporting—Once one has an exact office stock, Data centre equipment checking should be utilized to quantify execution pointers. However, tragically, thorough announcing of such frameworks can be... debilitating. Therefore, the organization applies the usefulness of DCIM and reports actual occurrences and episodes that permit an user to figure out the commotion and recognizes the requirements [20].
- Representation—Infrastructure and organization planning apparatuses can be utilized to hoist the presentation of Data centre supervisors by giving a natural comprehension of the office and stream of data.

5.4 The Internet of Things (IoT)

The ability to control sensors in pretty much every framework raises extra issues for data centres. As indicated by Gartner, the Internet of Things is a troublesome power

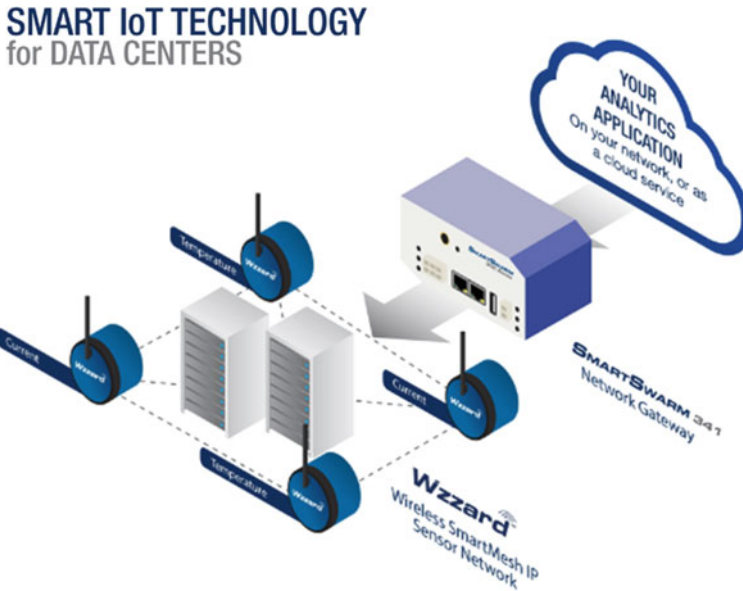


Fig. 7 IoT Technology for data centres [20]

that will change the data centre, attributable to the sheer volume of information it will deliver. Therefore, the IoT information should get handled, focused on, put away, and investigated.

Since IoT information is produced in mass, new data centre advances like edge figuring is essential to monitor the volume (Fig. 7).

5.5 Mobile Enterprise

Data centre organizing difficulties plague versatile registering specialist organizations and their “own gadget” methodologies, similarly as they are to the security of these gadgets. Workers have quick admittance to business-basic information through handheld gadgets, yet these gadgets should stay controlled and secured. To avoid the deficiency of classified data with arising data centre organizing difficulties, information access should stay controlled and restricted, regardless of whether labourers utilize their gadgets or the association gives cell phones and tablets. Remotely cleaning a cell phone’s memory or following and locking a missing or taken device would require additional assurance. All the while, extra inquiries regarding client protection keep on arising; for instance, what are the drawn-out results of regulation authorization approaching the information put away on any PC seized as a feature of an examination? Portable endeavour figuring presents innovative, authoritative, and legitimate issues that the data centre should resolve eventually (Fig. 8).

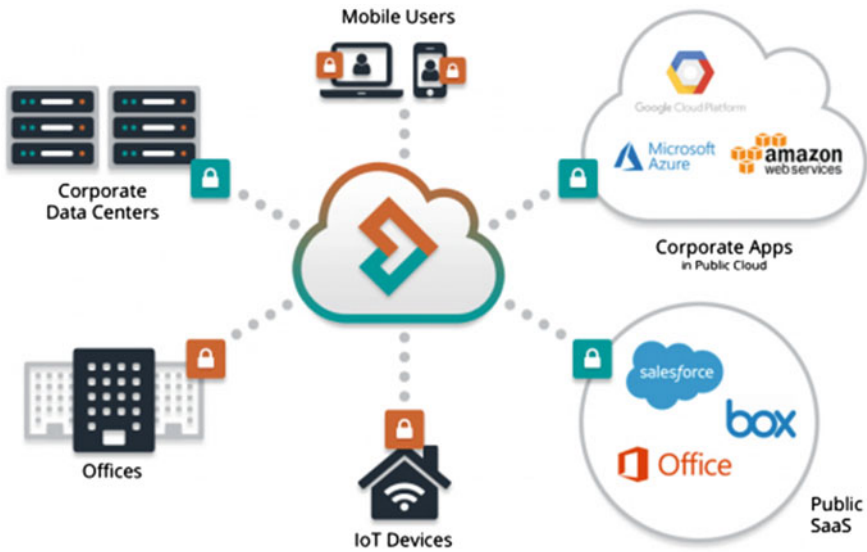


Fig. 8 Mobile Data Centres and the future of data centres [21]

5.6 Real-Time Reporting

The significance of ongoing information examination and revealing is developing. Not exclusively are DCIM apparatuses used to follow actual Data centre exercises, yet extensive information investigation empowers continuous checking of abnormalities or issues that might show a security break or other issue [21]. Moreover, ongoing checking following investigation propels us nearer to self-mending Data centres fit for starting a reaction, for example, disengaging a server or rerouting information traffic to a pre-characterized alert.

5.7 Balancing Cost Controls with Efficiency

Planning and cost administration are ongoing issues for each division; however, the Data centre’s expense control concerns are unique. While one must guarantee that the Data centres are thriving, inventive, and wealthy, one should likewise be aware of cost control. For instance, greening the Data centre is a consistent objective. Advancing energy effectiveness brings down functional expenses while promoting ecological obligation, so the IT directors control the productivity of force used. Different strategies, like virtualization, work on operational execution while controlling costs.

6 Threats Faced by Data Centres in India

The Data centre industry in India is evolving like never before. As per a report by MarketsandMarkets, the Indian Data centre market will arrive at \$1.5 billion by 2022 from \$1.0 billion, developing at 11.4%. This mushrooming of Data centres in India is mainly because of the ascent in web infiltration combined with the headways in distributed computing, cloud facilitating, Internet of Things (IoT), and Artificial Intelligence (AI). However, sadly, the Data centres in India have neglected to keep up with the prospering interest for information capacity, handling, and the board [22]. Their failure to work at ideal proficiency hampers the nature of their administration and squanders valuable assets.

6.1 *Inadequate Cognizance of Assets*

The resources like applications, associating links, capacity units, cooling frameworks, etc. reside in a data centre. With such countless complex frameworks working in parallel, it becomes bulky for data centre directors and administrators to screen and summarize the critical exhibition measurements close to ongoing. Constant measurements offer experiences in data centre activities, permitting the staff to act right away and make informed choices. Data centre administrators take manual readings without even a trace of continuous announcing. A conventional manual perusing does not hold much significance for a data centre where responsibility, utilization, and temperature generally change every several hours.

6.2 *Disproportionate Energy Exhaustion*

Data centres have continually been under the radar for being energy pigs. Data centres swallow a tremendous measure of energy inefficiently. An assessment indicates that data centres squander almost 90% of the power they pull off the matrix. Around the world, data centres drain 2% of the power delivered and transmit carbon dioxide equivalent to the aircraft business. With information traffic multiplying like clockwork, the circumstance can arrive at chaotic levels soon. Therefore, the Data centres need to execute extreme amplifications to mow down energy utilization to adequate levels and do their piece in lessening carbon impression.

6.3 Inefficient Capacity Planning

Most Data centres in India have no framework to decide whether their resources are running at the total limit. As a result, data centre chiefs will generally over-arrange assets to avoid any postponement or unscheduled vacation. While such a methodology guarantees higher uptime and accessibility, it also prompts the wastage of many assets such as unused space, power, and cooling [23].

6.4 Unfortunate Staff Productivity

In numerous Data centres, revealing manual frameworks are utilized. These frameworks require the staff to invest much energy in logging exercises into bookkeeping sheets. Such errands hamper the efficiency of Data centre administrators and keep them from zeroing in on other fundamental parts of the board. Supplanting generally utilized manual frameworks with mechanized frameworks can assist information with focusing staff work with higher effectiveness. They can invest energy in essential navigation and work on their contributions.

6.5 Long Recovery Periods

The majority of Data centres do not possess essential apparatuses to get data on what the resources in an organization are associated with and the location of these resources. Therefore, the data centre administrators take a great deal of time distinguishing and fixing issues when there is personal time. Such lengthy recuperation periods can be inconvenient to the drawn-out development of Data centres.

6.6 Growing Security Concerns

Data centres oversee and handle gigantic pieces of information. However, data centre offices are helpless against security gambles. Perhaps the greatest danger comes from humans. This can be from their workers, outsider clients getting into the organization, or favoured clients, for example, IT administrators. Data centres frequently neglect to safeguard their IT resources. Servers or hard drives at this point not being used often lie inactive and, on the off chance that not being disinfected as expected, can prompt spillage of essential data. Also, headways in IoT innovation acquire more gadgets and associations with the data centre network making new, unanticipated difficulties for administrators. With the multiplication of computerized devices and high-velocity organizations, India's development of data centres will proceed unabated. This will

fuel the previously mentioned difficulties and potentially bring more to the front. Data centres need to exploit well on schedule to forestall these from distressing their assets.

7 Security Threats of Data Centre

It is vital to comprehend that the data centre obliges the royal gems of secret information. This implies that when there is a digital assault, the individual data of the two clients and the organization's budget reports become uncovered. Data centre security, in any case, alludes to the actual practices and virtual advancements used to safeguard a Data centre from outer dangers and assaults. A Data centre is an office that stores an IT framework made out of organization PCs and capacity used to sort out, cycle, and store such information. Because of the special meaning of Data centres and the delicate data they hold, destinations must be carefully and truly got.

7.1 Classes of Data Centre Security

Under the intricacies encompassing Data centre security issues, parts ought to be thought about independently yet must follow one comprehensive security strategy. For the most part, security can be classified into physical and programming. Actual security envelops a broad scope of techniques used to forestall outside impedance. For example, programming security forestalls digital crooks from accessing the Network by circumventing the firewall, breaking passcode, or escaping clauses. Be that as it may, our consideration is on programming security. Hacking, malware, and spyware are Data centre security dangers or weaknesses. A Security information and event management (SIEM) offers an ongoing perspective on a Data centre's security centre. Before applications are sent, specific devices might be utilized to examine them for weaknesses that can be effectively taken advantage of and afterwards give measurements and intervening capacities.

With the ascent of distributed computing, permeability into information streams is a need because of malware stowing away within, in any case, genuine rush hour gridlock.

7.2 Who Needs Data Centre Security?

Nowadays, maintaining information is crucial for business. However, to guarantee improvement, it is vital to keep data safe and limit the gamble of potential dangers that may cause cash deficiency and notoriety.

Each datum community requires safety to ensure it proceeds with use. A few parts of “safety” comprise uptime focal points, such as various power sources and numerous climate control, and the sky is the limit. Data centre re-appropriating is a decent arrangement to ensure the information is put away as indicated by the best guidelines and gotten on each level.

Data centre network safety is exceptionally fundamental to any firm where private data reside in the data centres. Therefore, there must be thought from associations utilized in data centres either straightforwardly or through an accomplice to proffer answers for the high pace of digital assaults.

8 Cybersecurity Threats to Heed

Having underscored the enormous utilization of information, digital assailants are continually searching for new systems to cheat organizations. The vast majority of their procedures wait around the constant dangers to an association’s network safety. Notwithstanding, as one reads further, the majority of these dangers will be unveiled. The information on Data centre security dangers expands the insight concerning steps to forestall security issues. Coming up next are inescapable dangers of network protection.

8.1 Phishing Engineering Attacks

There has been a tremendous measure of phishing assaults against a wide range of targets in years. Phishing assaults are social designing assaults where the digital assailant creates a fake text, email, or site to deceive a casualty into delivering touchy data—which could include login qualifications for work, Visa subtleties, or watchwords to electronically connected records.

Among all digital assaults, phishing assault is one of the riskiest, as it tends to utilize a deluded representative to surrender authentic certifications and afterwards use the honour to wreck the organization’s framework.

8.2 Ransomware

Ransomware is the kind of malevolent programming intended to keep admittance to an association’s PC framework until an amount of cash is paid. These assaults, for the most part, include the aggressor contaminating an association’s Data centre using malware that encodes the entirety of the available information. In 2020, ransomware

assaults will be further uncontrolled than at any other time. Associations are being designated more than private residents because they have cash and the inspiration to pay ransoms [24].

8.3 Cyberattacks Against Hosted Services

The Data centre is essential due to business-basic and client-confronting applications. These applications can be delegated and taken advantage of in various ways, as discussed below:

Web and Application Attacks: Web applications are defenceless against a scope of assaults, incorporating those illustrated in the OWASP Top 10 and the CWE Top 25 Most Dangerous Software Weaknesses.

Dispersed Denial of Service (DDoS) Attacks: Service accessibility is fundamental for a positive client experience. DDoS assaults compromise accessibility, prompting loss of income, clients, and notoriety.

DNS Attacks: Data focuses facilitating DNS foundations are possibly defenceless against DNS DDoS assaults, reserve harming, and other DNS dangers.

Certification Compromise: Credentials penetrated through information breaks, qualification stuffing, phishing, and various assaults to ingress and take advantage of clients' Internet-based accounts. These and other assaults can disturb the accessibility, execution, and security of utilizations facilitated by a Data centre. Therefore, organizations should convey security arrangements that address these potential assault vectors.

8.4 IoT-Based Attacks

The utilization of savvy gadgets in homes and associations has expanded for the current year. Representatives are permitted to telecommute. The test is that not all ingenious devices have solid security introduced, making openings for digital assailants to capture these gadgets to invade business organizations. This assault uses a casualty's utilization of associated web gadgets to sneak malware onto an organization.

8.5 Internal Attacks

The cybersecurity threat by their employees is one of the biggest challenges for any organization. Some employees with vested interests may exploit their access to inflict

damage on the organization's Network. Though these attacks may be intentional or by unintentional human mistake, internal attacks remain the most significant risk to take care of due to their enormous amount of damage potential.

8.6 *Unpatched Security Susceptibility and Bugs*

An unexpected programming error in PC programming or working framework that digital assailants may avail to access frameworks unlawfully. As a rule, these imperfections may not emerge from a solitary working framework, yet communications from at least two unique projects make it hard to anticipate when a bug will show up.

9 How to Keep Data Centre Secure

Data centres store and deal with the touchy information in an association's ownership, contriving their security a centrepiece for any corporate information security procedure. Data centres ought to be gotten in light of the zero-trust security illustration, which cutoff points to ingress and authorize the base expected by business requirements. Successfully executing a data centre security technique requires conveying a scope of safety arrangements and carrying out accepted procedures. Nine of the primary contemplations for Data centre security include the following:

- **Forestall Vulnerability Exploitation:** Patch weak frameworks and applications and send an IPS to fix immediately upon a fix that is not yet accessible. IPS can likewise distinguish advantages against the DNS foundation or utilize DNS to bypass security insurances.
- **Execute Network Segmentation:** Network division forestalls parallel development and empowers the requirement of least honour access under the zero-trust security illustrations. Send security that can forestall east/west development between machines, notwithstanding security that forestalls north/south development enclosed by zones.
- **Secure Development Pipelines:** Implement secure coding and DevSecOps best practices and incorporate testing and strategy implementation into DevOps consistent joining and sending CI/CD pipelines.
- **Convey Web Application and API Protection (WAAP):** Use web application and API security answers to alleviate OWASP Top 10 dangers to web applications.
- **Use Cloud-Native Security Solutions:** In the half and half Data centre, specific responsibilities, compartments, and microservices with cloud-local security.
- **Safeguard Against DDoS Attacks:** Use on-prem and cloud DDoS securities to moderate DDoS dangers.

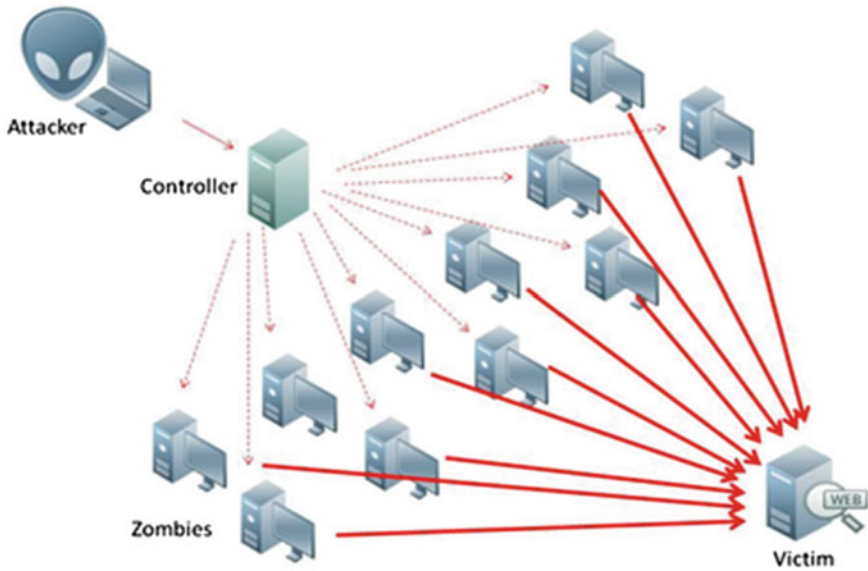


Fig. 9 Denial of service attack

- **Forestall Credential Theft:** Deploy against phishing insurances, for example, solid multifaceted verification (MFA) for clients to impede qualification taking assaults.
- **Secure the Supply Chain:** Detect and forestall complex inventory network assaults utilizing AI and ML-supported danger anticipation and EDR and XDR innovations.
- **Safeguard Sensitive Data:** Safeguard information very still, being used, and utilizing encryption, VPNs, and information misfortune counteraction (DLP) innovations (Fig. 9).

10 How to Curb These Attacks

Indeed, even as ongoing information breaks show that organizations are at a high gamble of digital assaults at some random time, there is an alleviation that these assaults can be checked.

10.1 Secure Your Hardware

The majority of organizations focus on the assurance of the product without concentrating on security. However, as this is disregarded, the organization will generally squander its gadgets to robbery, making it simple for private data to be controlled.

10.2 Encrypt and Backup Data

Organizations conceivably forestall admittance towards touchy information by hiding data through a code. Information encryption stays the “most productive fix” for information breaks assuming it may happen. This assists with keeping delicate data, including clients’ and representatives’ data and all business information.

10.3 Create a Security-Focused Workplace Culture

We have clarified before that workers can be a typical reason for information breaks, purposefully or accidentally; there will be a need to cause representatives to have a modest apprehension of the day-to-day activities that make an organization powerless against a digital assault. Associations ought to ensure enough security preparation and schooling for individuals from staff.

10.4 Invest in Cybersecurity Insurance

Cyber-criminals make a tireless attempt to identify progressive ways of breaching security defences. Therefore, companies should curtail risk by seeking a cybersecurity professional’s assistance to choose the first-rate protection for the organization because of the gamble of assault and the economic effect of such an occasion [25].

10.5 Physical Security

To restrain physical ambush, data centres employ routines such as the following:

1. CCTV security network: region and entrance points with 90-day video retention.
2. 24 × 7 on-site security defenders and network operations centre (NOC) Services and technical team.
3. Anti-tailgating/Anti-pass-back turnstile gate. Only allows one person at a time to get through the authentication.

4. Single avenue point into colocation dexterity.
5. Minimize load through dedicated data halls, suites, and cages.
6. Further entry restriction to private cages.
7. Three-factor authentication.
8. SSAE 16 compliant facilities.
9. Checking the provenance and design of hardware in use.
10. Minimize insider risk by overseeing activities and preserving the credentials.
11. Overseeing temperature and humidity.
12. Fire interception with zoned dry-pipe sprinkler.
13. Natural hazard risk-free regions.

10.6 Virtual Security

Virtual aggression can be avoided with the procedures such as the following:

1. Heavy data encryption during transfer or not: 256-bit SSL encryption for web applications. 1024-bit RSA public keys for data transfers. AES 256-bit encryption for files and databases.
2. Logs auditing activities of all users.
3. Secured usernames and passwords: Encrypted via 256-bit SSL, requirements for complex passwords, setup of scheduled expirations, and prevention of password reuse.
4. Entry based on the clearance level.
5. AD/LDAP integration.
6. Regulation based on IP addresses.
7. Encryption of session ID cookies to identify a unique user.
8. Two-factor authentication availability.
9. Third-party penetration testing performed annually.
10. Malware prevention through firewalls and automated scanner.

11 How to Secure Data Centres Against or After Cyberattacks

Being proactive in fixing the security suggests understanding the cyberattack lifecycle before it shows up at the data centre, how a break occurs, what happens once it is in, and how long it takes to decide it. Data centre security has inferred getting an affiliation's line for a significant time frame. Be that as it may, software engineers are getting keener. When they break the boundary, they move on a level plane to cause attacks on the enormous business and government associations. Also, software engineers these days are intentional and steady. It takes 24 days for the relationship to recognize and resolve an attack.

There is a relationship between risks and the applications running on the associations. These breaks use social planning techniques. Numerous association breaks start with an application, like an email conveying a disease. Exploiting a business connection gives the attacker permission to possibly an enormous number of clients and supplies data with immaterial effort. Whenever the aggressors are inside an association, they stay inconspicuous without genuinely trying, under the apparel of various applications, and continue with their vindictive activity subtle for weeks, months, or even seemingly forever.

Given the high risks, I am paying all due respect to security breaks after the attack will mean calamity. However, in light of everything, holding the attacks back from occurring regardless and making the attack expensive for a software engineer will urge them to progress forward.

11.1 Securing Different Regions Through Network Segmentation

When one secures their home, they do not simply connect the front and the back; you moreover set alerts for conceivable characteristics of segments like windows, parking space doorway, etc. It is a comparative idea for your data centre. Network division suggests various layers of safety that hold software engineers back from moving energetically inside the association. Would it be brilliant for them if they get past one layer? Think past the four dividers of affiliation and send security at section and leave concentrates yet, furthermore, at a more granular level. Plan and work for expectation:

It is great for organizations to guarantee that safety efforts are set up to shield data centres from disastrous assaults.

- As soon as possible, break down and recognize the basic alarms from harmless alarms, decreasing the reaction times required.
- Smooth out administration and paring down the number of safety approaches your association requires.
- Keep known and obscure assaults from happening by connecting designs that pinpoint destructive action.

11.2 Moving Beyond Segmentation to Cyber

Utilizing the organization's outskirts, conventional firewalls run as virtual machines. On the fringe, firewalling capacities are supplemented with an assortment of danger discovery and avoidance innovations like IDS/IPS against malware arrangements and web separating.

11.3 Advanced Attacks and Mature Attacks

The test is that data centres are not portrayed by their actual edges. A data centre will consistently encounter an attacker at a more grown-up time of an attack than the edge will and, in like way, will experience different kinds of risks and attack methodologies. Specifically, line risk evasion headways will, for the most part, be firmly based on perceiving a whole set out some reasonable compromise or defilement (for instance, exploits and malware). The issue is that aggressors will often move against the data centre after successfully compromising the edge. For example, the software engineer could have infiltrated various contraptions and taken client capabilities and, shockingly, the head's authorizations. Rather than exploits or malware, aggressors obviously will undoubtedly search for blade approaches to using theirs as of late obtained position of trust to access or mischief data centre assets. This infers that a data centre will habitually encounter attacks in a more grown-up time of the attack that could require clear signs of malware or exploits.

11.4 Behavioural

It is essential to recognize the total munitions stockpile in the programmer's tool compartment rather than simply abnormality. Penetrated chairman accounts, embedding indirect accesses, setting up secret passages, and RATS are indications of a continuous persevering assault. These procedures have let practices know that can make them stand apart from the regular traffic in the Network, giving you know what to search for. Rather than searching for a particular pernicious payload, one can search for what all loads would do.

11.5 Preempt the Silos

Recall that assailants do not adjust to limits by their actual nature. Digital assaults are an intricate trap of occasions and regarding the Data centre security as a different storehouse just aides the assailants. The more stages an aggressor needs to take to a break-in, the more secure Data centre climate is. We want to perceive that Data centres are extraordinary; however, they face general dangers.

12 Checklist to Help with Security Arrangements

1. **Secure the physical location:** A safe area implies sitting where the gamble of outside dangers, like flooding, is low. One likewise needs to consider the security of the supply of external assets like power, water, and interchanges.
2. **Data Centre should be wired:** Introduce reconnaissance cameras around the Data centre premises and eliminate signs that could give hints to its capacity. Data centre should be set as far as possible from the street as could be expected and it merits utilizing finishing to assist with keeping interlopers and vehicles under control. Simply have strong dividers without windows. Assuming there are windows, use those regions for regulatory purposes, as it were.
3. **Hire a security officer:** They should be a decent director of experts who can bear explicit undertakings to adjust to the security foundation and the job as the business needs to change. Extraordinary relational abilities are fundamental, alongside the capacity to assess and survey the effect of a danger on the company and impart it in non-specialized language.
4. **Restrict access:** Guarantee that actual access is confined to rare people who should be there. Characterize the circles that need admittance to the information vault. Confine admittance to the site and limit admittance to the primary entry and the shipment dock. Utilize two-factor verification, either a keycard/ideally biometric confirmation or an entrance code.
5. **Check who your people are:** You will do an intensive check of the individual, correct? Run an investigation application on representatives to cross-check issues, for example, addresses imparted to unwanted people. Get individuals' consent to run these checks: not exclusively will they favour checks to be run as it will add to their remaining inside the organization; it likewise implies those rejecting a look freely stand.
6. **Test your backup and security procedures** Test reinforcement frameworks routinely according to the maker's details. Test your fiasco recuperation plan by shortening a test region to the subsequent data centre. Characterize what you mean by a catastrophe and guarantee everybody knows what to do if one happens. Check to assume the recuperation plan works nevertheless permit you to meet your SLAs. Check whether available security systems are working accurately: for instance, honour levels ought to stay predictable with the jobs of every person. Check actual practices as well. For example, are fire entryways being set to open for the well-being of comfort? Are individuals leaving their PCs signed in and unprotected by secret word empowered screen-savers?
7. **Be smart about your backup:** Guarantee a reinforcement Data centre reflects the principal whenever the situation allows, so in case of a catastrophe closing down, the first, the second is online all the time. Construct your data centre as distant from the first as expected while staying associated by utilizing your picked broadband. You could involve it for load-adjusting and further developing throughput as well.

8. **Undertake a risk assessment:** Data centres are exceptionally similar to the business conditions. Many of the actions to safeguarding data centres are the presence of mind, yet you will not ever realize which are the savviest until you measure the expense against the advantages. This interaction will likewise permit you to focus on and centre your security spending where it makes the most significant difference. Get an outsider security appraisal organization to assess your security. Another pair of eyes regularly see things in-house staff might ignore. Do your verification first.

13 Benefits of Cybersecurity

Today, the data centre has not ever been more significant. The best network safety firm assists with keeping assaults from taking impact and guarantees that your organization's information stays private. The advantages of network safety cannot be overemphasized. The following are a couple of benefits of network protection:

1. Assurance of your business: Network safety arrangement gives advanced assurance to your business. This guarantees that your data is not at a gamble of likely dangers.
2. Expanded usefulness: Data centre security gives delayed down creation ability, preventing representatives from completing their positions. When network safety issues are dealt with, representatives will want to work really.
3. Move trust in your clients: Whenever you have demonstrated that your business is safeguarded against a wide range of digital assaults, this makes clients more positive about utilizing your administration.
4. Insurance of your clients: Guaranteeing that your business is protected from data centre security dangers assists with safeguarding your clients, who could be defenceless to a digital break as a substitute.

14 Conclusion

Data centres hold misleading information about associations. Consequently, it is essential to keep them secure. RSI Security's data centre security organizations help shield the data centre and assure the affiliation stays before harmful performers by offering the data expected to react to security breaks. Our skilled professionals ensure that your affiliation's private information is secured without devastating your association inside the IT office.

In the present expanded computerized peril scene, the standard data centre security model of boundary controls and acknowledgement-based models do not prevent sophisticated attacks on virtualized IT establishment or the OT firmware and programming that maintains it. That is because peril aversion developments are

consistently based on separating a hidden compromise rather than stopping an attack. Moreover, they are firmly subject to genuine resource noticing, which does hardly anything to diminish the danger.

This paper familiarizes one more philosophy with traditional data centre network insurance, known as cyber hardening, in which therapists attack surfaces and deny malware the consistency to spread. By cementing programming copies, data centre security gatherings can take out an entire class of cyberattacks. In the wake of perusing this paper, those liable for the uprightness, secrecy, and accessibility of data centres will be informed about digital solidifying and how the procedure gives more prominent assurance than edge controls and recognition-based models.

References

1. Lo, ai T., Darwazeh, N.S., Al-Qassas, R.S., AIDosari, F.: A secure cloud computing model based on data classification. *Elsevier Procedia Comput. Sci.* **52**, 1153–1158 (2015)
2. Tripathy, H.K., Mishra, S., Suman, S., Nayyar, A., Sahoo, K.S.: Smart COVID-shield: an IoT driven reliable and automated prototype model for COVID-19 symptoms tracking. *Computing*, 1–22 (2022)
3. Mishra, S., Thakkar, H.K., Mallick, P.K., Tiwari, P., Alamri, A.: A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection. *Sustain. Cities Soc.* **72**, 103079 (2021)
4. Bacon, J., Eyers, D., Pasquier, T.F.J.M., Singh, J., Papagiannis, I., Pietzuch, P.: Information flow control for secure cloud computing. *IEEE Trans. Netw. Serv. Manag.* **11**(1), 76–89 (2014)
5. Mishra, S., Panda, A., Tripathy, K.H.: Implementation of re-sampling technique to handle skewed data in tumor prediction. *J. Adv. Res. Dyn. Control Syst.* **10**, 526–530 (2018)
6. Nejad, M.M., Mashayekhy, L., Grosu, D.: Truthful greedy mechanisms for dynamic virtual machine provisioning and allocation in clouds. *IEEE Trans. Parallel Distrib. Syst.* **26**(2), 594–603 (2015)
7. Mishra, S., Mishra, B.K., Tripathy, H.K.: A neuro-genetic model to predict hepatitis disease risk. In: 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), pp. 1–3. IEEE (2015)
8. Tripathy, H.K., Mallick, P.K., Mishra, S.: Application and evaluation of classification model to detect autistic spectrum disorders in children. *Int. J. Comput. Appl. Technol.* **65**(4), 368–377 (2021)
9. Khan, A.U.R.: Mazliza Othman and Sajjad Ahmad Madani, “mobile cloud computing application models.” *IEEE Commun. Surv. Tutorials* **16**(1), 393–413 (2014)
10. Raja, H., Bajwa, W.U.: Cloud K-SVD: a collaborative dictionary learning algorithm for big, distributed data. *IEEE Trans. Signal Process.* **64**(1), 173–188 (2016)
11. Mondal, S., Tripathy, H.K., Mishra, S., Mallick, P.K.: Perspective analysis of anti-aging products using voting-based ensemble technique. In: *Advances in Systems, Control and Automations*, pp. 237–246. Springer, Singapore (2021)
12. Yang, K., Jia, X.: An efficient and secure dynamic auditing protocol for data storage in cloud computing. *IEEE Trans. Parallel Distrib. Syst.* **24**(9), 1717–1726 (2013)
13. Krithika, Dilipan, G.L., Shobana, M.: Enhancing cloud computing security for data sharing within group members. *IOSR J. Comput. Eng. (IOSR-JCE)* **17**(2), 110–114, Ver. V (2015)
14. Greveler, U., Justus, B. et al.: A Privacy preserving system for cloud computing. In: 11th IEEE International Conference on Computer and Information Technology, pp. 648–653 (2011)
15. Mohapatra, S.K., Mishra, S., Tripathy, H.K., Bhoi, A.K., Barsocchi, P.: A pragmatic investigation of energy consumption and utilization models in the urban sector using predictive intelligence approaches. *Energies* **14**(13), 3900 (2021)

16. Tripathy, H.K., Mishra, S., Thakkar, H.K., Rai, D.: Care: a collision-aware mobile robot navigation in grid environment using improved breadth first search. *Comput. Electr. Eng.* **94**, 107327 (2021)
17. Bohli, J.M., Gruschka, N., Jensen, M., Iacono, L.L., Marnau, N.: Security and privacyenhancing multi-cloud architectures. *IEEE Trans. Dependable Secure Comput.* **10**(4) (2013)
18. Wang, C., Chow, S., et al.: Privacy-preserving public auditing for secure cloud storage. *IEEE Trans. Comput.* **62**(2), 362–375 (2013)
19. Yu, R., Gjessing, S.: Toward cloud-based vehicular networks with efficient resource management. *IEEE Netw. J.* **27**(5), 48–55 (2013)
20. Surjijamol, R.: A social compute cloud: for sharing resources. *Int. J. Sci. Res. (IJSR)* **4**(2) (2015)
21. Mishra, S., Tripathy, H.K., Thakkar, H.K., Garg, D., Kotecha, K., Pandya, S.: An explainable intelligence driven query prioritization using balanced decision tree approach for multi-level psychological disorders assessment. *Front. Public Health*, 9 (2021)
22. Mallick, P.K., Mishra, S., Mohanty, B.P., Satapathy, S.K.: A deep neural network model for effective diagnosis of melanoma disorder. In: *Cognitive Informatics and Soft Computing*, pp. 43–51. Springer, Singapore (2021)
23. Mishra, S., Dash, A., Ranjan, P., Jena, A.K.: Enhancing heart disorders prediction with attribute optimization. In: *Advances in Electronics, Communication and Computing*, pp. 139–145. Springer, Singapore (2021)
24. Mishra, S., Tripathy, H.K., Mallick, P.K., Bhoi, A.K., Barsocchi, P.: EAGA-MLP—an enhanced and adaptive hybrid classification model for diabetes diagnosis. *Sensors* **20**(14), 4036 (2020)
25. Jia, G., Han, G., Zhang, D.: An adaptive framework for improving quality of service in industrial systems. *IEEE Access* **3**, 2129–2139 (2015)

Blockchain-Based Secure E-voting System Using Aadhaar Authentication



Ankit Kumar Jain, Sahil Kalra, Karan Kapoor, and Vishal Jangra

Abstract Election in any country is a very inflated and time-consuming process. Many countries have tried to implement numerous ways of conducting an election to make the process easy and time and cost-efficient. One such method is a blockchain-based secure voting system that has been in research for a long. However, it is still a challenge to implement it because of various reasons like security, availability of the internet to everyone, etc. Therefore, this paper presents a secure election system based on blockchain, tackling a few of these problems and making it more secure. The proposed system uses Aadhaar-based voter authentication and One-time password (OTP)-based verification. The proposed scheme uses a private blockchain that solves many security problems and even attacks like Sybil, Distributed denial of service (DDoS), etc. The proposed system can reduce the cost of elections to a large extent while increasing election security.

Keywords Election · Blockchain · Aadhaar · Authentication

1 Introduction

Voting is an important part of any democratic system as it helps in choosing the right leader. Thus it becomes important to conduct voting in a transparent and verifiable manner. However, voter turnout has diminished in recent years as elections are surrounded by ballot forgery, persuading voters to vote for a particular party by threatening them, etc. As a result, questions are growing about the current voting system's reliability and security [12]. Therefore, e-voting was adopted; however, it is not cost-effective and needs complete central authority monitoring. To fix these issues, blockchain technology can be used to build a secure and cost-effective voting system [12].

The idea of blockchain relies on a concept called “Distributed Ledger Technology”, which can store data on different servers. A blockchain is an immutable and

A. K. Jain (✉) · S. Kalra · K. Kapoor · V. Jangra
National Institute of Technology, Kurukshetra, India
e-mail: ankitjain@nitkr.ac.in

continuously growing list of records, called blocks, linked and secured using cryptographic algorithms. It consists of encrypted data blocks in the form of a chain that collectively makes the blockchain. Each block contains a number of transactions, and in the case of the voting system, each block contains information of the voters in the form of hash value [16].

In the blockchain, each transaction gets verified by all other participants, and for that, it must get successful verification from the majority of participants. It becomes almost impossible to change a block's data after adding it to the blockchain.

For securing blockchain technology, several cryptographic algorithms are applied. Blockchain technology makes use of cryptography in different ways for wallets, transactions, security, and privacy-preserving protocols. Cryptography in the blockchain is mainly used for two purposes:

- (a) To secure the sender's identity of transactions
- (b) To ensure that the past records cannot tamper

Therefore, a person can have complete confidence while using blockchain that if some information is recorded on a blockchain once, it is stored in a legitimate manner that preserves security. To increase blockchain security, proof of work is used by slowing down the creation of new blocks.

The main concern in voting these days is increasing the security of elections, whether online or offline. Security of blockchain-based voting systems is increased in many ways. Following are the security goals of a private blockchain are:

- (a) **Confidentiality:** Blockchain protects the voter identity by encrypting the block's data, so voters' information and vote are encrypted, ensuring confidentiality [16].
- (b) **Anonymity:** Each voter is assigned a private key that only the voter can see and a public key that anyone can see but is also hashed. Therefore, no voter can recognize one another in the blockchain, and so the voter's identity is protected.
- (c) **Non-repudiation:** Major problem in Electronic Voting Machine, or EVM-based election, is that it often claims the tampering of EVMs and unfairness in elections. Non-repudiation means denying like this. In a blockchain-based voting system, no one can deny the conduction of a fair election as it is transparent, and others verify every vote on the network to obtain consensus [16].
- (d) **Consistency:** Every node in a blockchain stores a copy of the complete blockchain. Therefore, consistency is maintained, and every node has a similar blockchain copy similar to the results of the election as well.
- (e) **Auditing:** As blockchain stores everything on blocks that cannot be changed so we can even inspect it any time after the elections to check if the election was conducted in an impartial manner [16].
- (f) **No possibility of Sybil Attack:** In a Sybil attack, an invader generally hacks a centralized system by taking control of the system by creating multiple nodes or accounts. We can eliminate this threat by using a private blockchain as only authenticated users have access to blockchain [3].

There can be two possibilities of conducting elections with a blockchain-based system. One is offline elections, and the other is online elections. Offline elections can be conducted using a blockchain-based system, and voters will have to visit the voting centre to vote. These systems will increase the security by adding all blockchain security features, and such systems cannot tamper like EVMs [5]. But still, in offline elections, voter gatherings remain low as many people cannot reach the voting centre for various reasons [5]. Hence, online elections can make elections very easy for voters. It can save everyone's time, and even voter turnout can be increased. However, online elections are difficult to conduct as online elections still face many challenges, like voters' devices could be compromised and thus can change voters' votes.

Many countries have tried to conduct online elections; Estonia was the first to conduct nationwide online elections [18]. It was reported that voter turnout increased largely, and it was cheaper than other methods of election, and each such election saved 11,000 working days [11]. However, security is still a concern. Some countries prefer a ballot-based election mechanism as it is the most secure to date. Some prefer electronic voting as it is easy to conduct and consumes less time than a ballot-based election. However, electronic voting lacks in terms of security because such machines can tamper.

There are many issues regarding the privacy of voters, delay in counting of votes, manipulation of results, and trust issues between different parties in paper-based elections [19]. In Turkey, there was a dispute between parties over election security and privacy and whether the votes counted were genuine or manipulated. This incident happened in many democratic countries, and as of this, many parties claim the manipulation of the result. Due to this, many issues occur, and in some cases, the parties also demand re-election. To overcome this, E-voting is considered by some countries in a variety of fields. The first county which tried e-voting on the country level was Estonia, in which elections were held between 2005 and 2007 [7].

However, blockchain-based elections are not applied to any country yet. Blockchain has been in the development process in recent times, and many countries have started taking an interest in this technology for the election process. South Korea has taken the initiative to conduct the blockchain-based election in the year 2017 [4]. Blockchain is a more secure and comfortable alternative to the traditional election system. Blockchain increases the security of the data such that it keeps the entire data in the form of blocks and removes the need for a central authority that can conduct the election. As previously stated, the paper election system takes time to count votes and make the results public, but blockchain solves this issue by providing instant results. Because the last node in blockchain holds all of the system's data, all we have to do is search the last node to get the results, cutting down on waiting time and eliminating the possibility of outcome manipulation. Therefore, the blockchain is considered to be the ideal solution for elections in democratic countries.

The proposed approach uses a private Ethereum blockchain using Ganache. A voter with only a valid Aadhaar number can vote as Aadhaar number is checked against the Aadhaar validator. Upon successful Aadhaar validation, OTP verification is done, and the OTP is sent to the voter's registered mobile number. The voter is

allowed to change the vote any number of times before the election ends. Metamask is used to submit voting transactions to the blockchain node. The last vote is considered the final vote. This method prevents the method of forced to vote, and provides voter anonymity, which provides better security for the election process. As the vote is stored in a block after hashing, its anonymity is maintained. Private blockchain eliminates the possibility of a Sybil attack in which a node can operate with multiple identities simultaneously and tries to take the power of the system. Such attacks are common in the centralized system. Auditing becomes easy with blockchain as each transaction can be verified.

The rest of the paper is organized as follows: Sect. 2 presents the related work. Sect 3 presents an overview of the proposed blockchain-based secure e-voting system. Sect 4 presents implementation details. Sect 5 presents the security analysis of the proposed e-voting system. Sect 6 compares the proposed system with the existing e-voting systems, followed by conclusion and future scope in Sect. 7.

2 Related Work

There are a lot of possible methods of conducting the election in democratic countries. Some of these methods are based on the paper-based election, and some use EVM in the process. One of the systems uses email-based authentication for the election where the verification code is sent to the person's email id and using that, the voter can cast a vote. However, since many systems can open the email, it is easy to hack the system, and certain people who register to use this system can use someone else's email address to vote on their behalf; for example, a parent vote can be cast by a son, resulting in election rules violations [14]. Researchers suggested a peer-to-peer blockchain-based voting system with the aim of protecting the privacy of ballots as well as the particulars of the individual casting votes to the blockchain system. They employ a one-of-a-kind vote pledge format in this strategy. Another method is to use a one-time ring signature, which helps to protect voters' identities by ensuring that only the individual whose id is used can vote [20]. However, in this election, each candidate requires a public key pair. Adding a new candidate would increase the difficulty of the signing process, requiring more Central Processing Unit power from all nodes involved in the system. This mechanism is not reliant on any central authority. Although in certain situations, the government will serve as a central authority to ensure a smooth and secure election.

McCorry et al. [14] implemented a decentralized voting system using blockchain with increased voter privacy. It used smart contracts to implement the blockchain. The advantage of the system is that it uses a decentralized voting system and increases privacy, but the system fails if the voter's device is compromised. It can only be used on a small scale.

Barnes et al. [4] also used blockchain for the online election system. The system was built in the United Kingdom and was named Net vote. It used Ethereum-based

blockchain and decentralized apps for various election processes. It can be integrated with other devices for voter authentication but currently doesn't have inbuilt voter authentication. The system is decentralized, which increases the security of the system. The system uses a public blockchain which increases gas costs. Attacks like the Sybil attack can be performed as blockchain is public.

Ayed et al. [3] has developed a blockchain-based voting scheme with an authentication system and maintains anonymity. The blockchain is publicly verifiable, but the method assumes that the voter uses a secure device to cast a vote. Moreover, this system doesn't allow to change the vote in case of voter's mistake, which increases the chances of forced vote.

Sadia et al. [16] proposed a blockchain-based voting system along with the smart contract. The protocol of the system is designed to reduce memory and time consumption which makes it faster. The system maintains anonymity, voter's privacy, confidentiality, transparency and is publicly verifiable and auditable. Manipulation of ballots is not possible in the system. Still, the system assumes voters use a secure device to cast votes. The system doesn't allow to change the vote, which increases forced voting.

Liu et al. [13] proposed a protocol where the choice was made safe using a random string and choice code. The length of the vote string varies depending on the election requirements. Anonymity is ensured by keeping voters identity private. Advantage of system is that it maintains auditability and transparency but ballots are visible when they are cast to the blockchain exposing the progress of election. Another problem is that there can be a ballot collision when same string is produced by different votes and one of the two ballots will become invalid although probability of the collision is very less.

Freya Sheer Hardwick et al. [9] proposed an e-voting system based on blockchain technology. System uses a central authority for voter authentication purpose. Advantage of the system is that voter can vote multiple times reducing the chances of forced vote but the system assumes voters use a secure device to cast votes.

In Anonymous voting by a two-round public discussion[8], a two-round protocol is suggested that computes the tally in two rounds without using a private channel. The protocol is efficient with regard to amputation and bandwidth consumption but is neither robust nor fair in some conditions.

Zcash [17] is a public blockchain system which uses a privacy protection system which shows transactions to encrypt sender, recipient and message content. To ensure data integrity and prevent fraud, Zcash uses anonymous evidence called zk-SNARK [17].

Several blockchain frameworks can be used for implementing the blockchain-based e-voting system. Exonum, Quorum, and Geth are some of the frameworks that can be considered and explained below:

- a. **Exonum:** Exonum blockchain is a scalable end-to-end blockchain that is developed using the Rust programming language. This system is used for private blockchain and achieves network consensus using a Byzantine algorithm. It can

Table 1 Blockchain framework evaluation [10]

Properties	Exonum [1]	Quorum [10]	Go-Ethereum [21]
Decentralized	Yes	Partially	User Preference
Language of smart contract	Rust	Rust	Solidity
Implementation language	Rust	JavaScript, C, Go	Go, C, JavaScript
Consensus	Byzantine Fault Tolerant or BFT Algorithm build on Custom	Istanbul Byzantine Fault Tolerant. Or IBFT, Quorum Chain	Proof of Authority, or POA, Proof of Work, or POW

handle 5000 transactions per second, but since Rust is the only programming language used in this framework, many users may find it difficult to use [1].

- b. **Quorum:** Quorum is a contract-based distributed ledger technology based on Ethereum that uses private policy and consensus mechanisms. This technology has altered the consensus system, aiming to provide an algorithm based on a consortium chain that allows for many transactions per second [10].
- c. **Geth:** Geth is one of the first Ethereum implementations, and it runs as expected with no risk of device downtime, fraud, or third-party intervention. Of all the frameworks used in this type of system, this is the most developer-friendly. If the system is implemented as a private or public network it determines the number of transactions per second. Table 1 compares various blockchain frameworks.

3 Proposed Work

The proposed approach implements a blockchain-based voting system with increased security with an Aadhaar validator, captcha, and OTP-based login. We have used the Ethereum blockchain and wrote smart contract for the election, which contains voting logic and the candidate's vote count. Blockchain itself is a secure technology that can handle many security issues than centralized systems. Adding the Aadhaar validator and OTP verification makes the system more secure. Elections can be conducted transparently, and the consistency of votes can be maintained wherein no block in blockchain can be changed, and every node has the same copy of the blockchain. In our system, the admin can start elections by defining the duration of elections and by starting the counter. The election ends when the counter reaches 0, and the election results can be displayed as soon as the election ends.

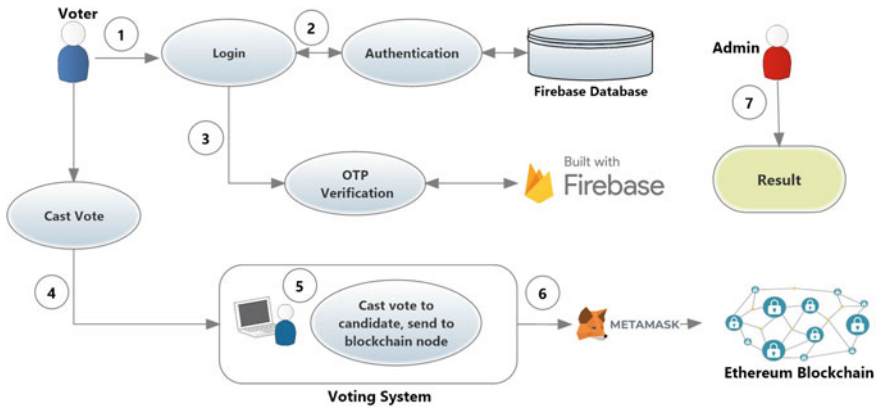


Fig. 1 Blockchain-based secure voting system architecture

3.1 System Architecture

Figure 1 presents the system architecture of the proposed voting system. When a voter approaches voting, and if the election is going on, the voter is redirected to the login page and asked to enter their Aadhaar number. Upon submission, the validity of the Aadhaar number is checked through the Aadhaar validator of the npm Aadhaar validator library. If the Aadhaar number is valid, the voter has to fill a reCaptcha to prove that the voter is a human being and not a computer program. After successful captcha submission, OTP is sent on registered mobile number. Registered mobile number is the one registered in government database corresponding to the Aadhaar number entered. After the OTP verification, the voter is logged in, the system checks if the voter has already voted, and if the voter has already voted, they are asked if the voter wants to change their vote and if no will be redirected to the home page. If the voter has not already voted or wants to change their vote, they can vote for the candidate of their choice. The most recent vote will be considered the final vote and will be added to the block, and the block will be added to the blockchain after the election ends. Figure 2 presents the voter interaction with the system.

A voter can select the candidate’s logo of their choice of the party or candidate. A voter can vote by double click on the vote button. Metamask pops up to confirm the voting transaction. After confirmation, the vote will be added to the blockchain. After voting, the voter is then redirected to the home page. When the election ends after the timer reaches 0, only the result page is displayed. The voter’s database is maintained in Firebase.

The smart contract of our blockchain created the genesis block of blockchain and is the basic logic behind voting. Algorithm 1 shows the smart contract of our blockchain. The smart contract contains the logic for voting, ending elections, and registration of new candidates. It contains a structure of the candidate, which consists of the candidate ID, candidate name, and voting count. The contract contains the list of voters, candidates list, and election state, which states if the election is proceeding

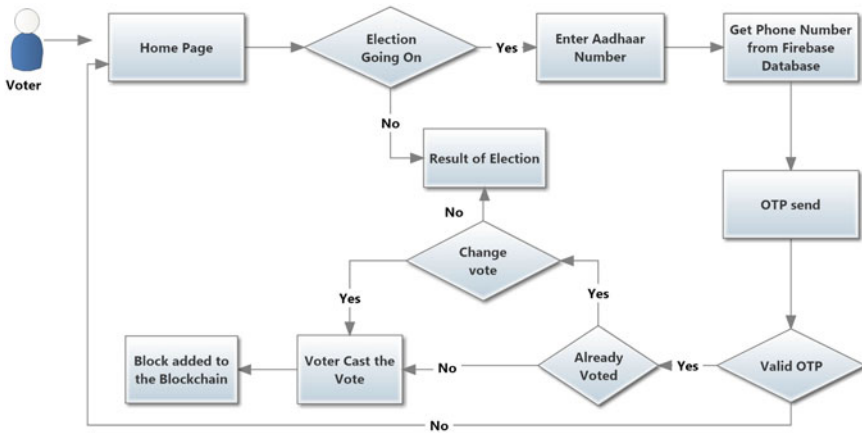


Fig. 2 Voter interaction with the system

or has ended. Register candidate function registers a candidate who can be voted when the election starts. Function castVote contains the logic for casting a vote. It checks if the election is proceeding currently and the candidate ID that the voter selected is valid and then increases the candidate vote count to which the voter voted.

Algorithm 1: Smart contract of proposed blockchain-based E-voting system

```

contract ElectionContract {
    struct cand
    {
        int cId ;
        string cName ;
        int vCount ;
    }

    int candCount ;
    event voted (int cId) ;
    mapping (address => bool) public votersList
    mapping (int => cand) public candidatesList
    boolean electiongoingon = true
    constructor() public
    {
        registerCandidate("Candidate A");
        registerCandidate ("Candidate B");
        registerCandidate ("Candidate C");
        registerCandidate ("Candidate D");
        registerCandidate ("Candidate E");
    }
    function registerCandidate (string name) private
    {
        candCount += 1 ;
        candidatesList[candCount] = cand(candCount, name, 0) ;
    }
    function castVote (int cId) public
    {
        require (cId <= candCount && cId > 0, " candidate is not valid " ) ;
        require (electiongoingon, "Election has been ended");
        votersList[msg.sender] = true;
        candList[cId].vCount += 1;
        emit Voted(cId);
    }
    function endElection()
    {
        electiongoingon = false ;
    }
}
    
```

4 Implementation Details

The setup consists of the blockchain on which we deployed our smart contract and then made the genesis block of the blockchain. The smart contract contains the list of candidates and each candidate’s ID and vote count. Smart contracts are written in solidity, which has a JavaScript-like syntax, as discussed in Algorithm 1. A smart contract is a contract between two parties to whom the transaction takes place. Once deployed on the blockchain, this agreement cannot be changed, and it is immutable. The smart contract used in the system contains the simple logic of voting where the voting function is obtained from the voter’s ID, and a candidate ID is sent by the voter after selecting the candidate. Then smart contract self-executes and increments the total count of votes of the candidate. The smart contract also checks if this voter has voted before and only considers the voter’s vote if it has not voted if not voted previously. It consists of the logic to end the election, which ends when the time for election ups.

Npm Aadhaar-validator's library [2] is used to validate the Aadhaar number, which checks against a structure of the Aadhaar provided by the Government of India. It uses the Verhoeff algorithm to validate the Aadhaar number as the United Identification Authority of India (UIDAI), uses it.

We have used Firebase's real-time database to maintain the records of voters, which stores voters' Aadhaar and phone numbers. A real database for the voters may contain details like name, age, family details, and all other information that the government has of the particular voter. To increase the security, we used Firebase's OTP verification too so that no person can vote with the Aadhaar number of another person. It also asks to submit Recaptcha so that no automatic system can vote.

For submitting voting transactions, we have used Metamask as a client and used Ganache to implement the blockchain as Ganache provides a private blockchain, an Ethereum blockchain. We can set the default account balance and even select the no of accounts that can vote. Ganache can provide >1 and <100 accounts. Ganache also provides forking in case the rules or agreement needs to be changed. One type of forking is soft forking which are very small changes in a protocol, such as maintenance and little changes, and the other is hard forking which even Ganache provides.

Forking is changing the main rules or protocol. Ganache provides five hard forks, which we can use to fork the blockchain if needed. We can even provide an option block number through which we want to fork our blockchain. Petersburg, Constantinople, Byzantium, Istanbul, and Muir Glacier are hard forks provided by Ganache. Hard forking is changing many rules, like increasing block size, increasing privacy or transaction speed, etc. For example, the Ethereum metropolis fork implemented a proof of stake algorithm and not a proof of work algorithm. PoW, in a broad sense, slows the creation of a new block because otherwise, a computer with a large power to generate a large number of hashes per second could alter the blockchain by calculating a new hash of every block on the blockchain. However, with PoW, hackers will now have to calculate the proof of work for each block, making the process slow.

Proof of stake (PoS) randomly chooses the node among miner nodes who have registered themselves as miners to mine the block instead of making all nodes compete to solve the hash first and wasting a large amount of power. Therefore, instead of miners, PoS uses validators, and the process is called minting or forging instead of mining. In production, we need miners or validators, but Ganache provides auto mining that calculates the transaction's hash and adds the block to the blockchain instantaneously. It also provides an option to select the time in seconds, after which Ganache will submit the new block to the blockchain. The new block contains all transactions that happened in that time or the amount of data that the block can store with the block's hash power. Then voter needs to enter their Aadhaar number which is checked against the Aadhaar-validator and database. If the Aadhaar number is valid, an OTP is sent to the voter's number, and the voter needs to enter OTP in prompt. If OTP verification is successful, the voting window appears.

5 Security Analysis of Proposed System

This section gives a brief overview of the security concerns, which are mitigated using the proposed method. Some of them are listed below:

- (a) **Server Tempering:** As the system is built with blockchain, it becomes almost impossible to hack the blockchain and change votes.
- (b) **Sybil Attack:** We have used a private blockchain, which eliminates the possibility of a Sybil attack in which hackers may try to take control of the whole system by creating a large number of nodes.
- (c) **Forced Vote:** When elections are conducted online, someone in power or family may influence voters to vote for the candidate of their interest. For such a scenario, we have allowed a voter to vote multiple times before the election ends and the last vote of the voter is considered.
- (d) **DDoS:** Distributed Denial of Service attack in which a malicious attack is performed to disrupt normal traffic of targeted server. Bootnodes are used for node discovery protocol and nodes are implemented with Byzantine fault tolerance algorithm, and the attacker is immediately located.
- (e) **Compromised Devices:** The system confirms voters stating the candidate they are voting for if the voter device is compromised and some other candidate gets selected because of the compromised device.

6 Comparison with Existing Techniques

Table 2 compares the proposed approach with various voting systems that have been used in the past. The United States of America uses a ballot-based system in which voters can cast a ballot in-person at the polling booth on the election day or cast a ballot by-mail [21]. Voter identification of voters who vote by mail is done through matching of signature, whereas in polling, voters are required to carry some government-issued photo identification. It is a most secure system as no hacking can be done, and anonymity is maintained too. Counting of votes is done by scanning ballots. However, security is more in such systems, but it is a long and time-consuming process. Moreover, voting and counting in such a system are difficult with a large number of voters.

Another very common voting system used these days is the EVM-based Voting [5]. Electronic voting machines are used for recording votes. These machines run on battery, and a voter has to select the candidate of their choice and vote that candidate. Voter Identification is done by verifying a government-issued voting card or other identification documents. VVPAT is also used with EVMs, which verify the vote by printing a slip containing the candidate's symbol [5], candidate name, and serial number, and slip fall in a sealed drop box of VVPAT. After the election ends, the counting of votes is done. Such a system is much faster than a ballot-based system, but recently, there have been many claims that such systems can be hacked.

Table 2 Comparison between different voting systems

Voting system	Architecture	Considerations for safety	User authentication
Paper-based voting [6]	Ballet-based voting	Hacking is not possible	Physical validation of voter card
EVM-based election [5]	The system has once written and read-only memory	i. Self-contained ii. Voter-verified paper audit trail, or VVPAT iii. No wireless components and interface	Physical validation of voter card [15]
Email address-based election [14]	Handshake Authentication Protocol	i. Confidentiality ii. Transparent	The email address of the voter
One-time ring signature election [20]	Smart contract on Ethereum	There is a degree of anonymity involved	Authentication based on asymmetry
Estonia election (online election) [18]	i. Authentication with two factors ii. Algorithm for encrypting data using public and private keys	i. The risk involved if the voter forgets his allocated ID and password ii. Regular monitoring for data integrity	Random passwords are generated at polling stations for voters
Securing e-voting based on blockchain in P2P network[22]	i. Authentication is based on Elliptic Curve Cryptography ii. Withdrawal model for allowing voters to change their vote before a pre-set deadline	i. Uses SHA-256 to generate the hash value of $H=Hash(ID+Vote+TImestamp)$ ii. Voter uses private key to generate a signature S, miner uses public key to verify S	User credential model based on Elliptic Curve Cryptography
Secure digital voting system based on blockchain technology[12]	i. Web-based interface with finger printing ii. Cryptographic hash of the transaction (ID) is emailed to the voter as a proof that the vote has been casted	i. Prevents double voting using Fingerprint ii. Tracks the vote casted using transaction ID	Voter logs into the system by providing their thumb impression
Proposed system	i. Smart contract on private Ethereum blockchain ii. Voter's data maintained on Firebase Database	i. Privacy ii. Transparency iii. There is a degree of anonymity involved	i. Aadhaar validation ii. OTP-based authentication

Another voting system is an online voting system using email-based authentication [14]. A handshake authentication protocol is used for authentication purposes. An email is sent to the registered email ID to identify the voter, and then the voter can vote after successful identification. Such a system has many security concerns as the email can be hacked or accessed by other family members. In addition, a vote can be cast only one time and cannot be changed. Estonia was the first country to use Internet-based voting [18]. Estonia has issued national smart ID cards that voters need to have to cast a vote. Voters can go to Estonia's election website to cast their vote, and with the help of a national smart ID card, voter identification is performed. To avoid a forced vote, a voters can change their vote multiple times before the election ends.

Online voting is the fastest among ballot-based and EVM-based voting systems, as election results can be displayed immediately after the election ends. Nevertheless, there are many security concerns in the online voting system, like servers' hacking and compromised voter devices. To implement the online voting system, advanced technology is required to secure the system.

One such existing comparable model is Secured e-voting-based on blockchain in the P2P network. It uses a synchronized model of voting records based on Distributed Ledger Technology to avoid forgery of votes. It uses a user credential model based on ECC to provide authentication and non-repudiation. It also uses a withdrawal model that allows voters to change their vote before a preset deadline.

Another such comparable model is the Secure Digital voting system based on blockchain technology. It uses the fingerprint of the voter for the authentication process. It uses hash for the transaction ID of the transaction to track the vote and so maintains transparency. Though the system authenticates the voter, it doesn't protect against forced vote because voting is web-based.

7 Conclusion and Future Scope

The paper implemented a blockchain-based secure voting system with smart contracts, which helps conduct a secure election with a minimized cost. It discussed how to conduct the elections in online and offline mode. An offline blockchain-based system would increase security, but an online blockchain-based system will help save the cost of the election and time and make elections easy.

We have used OTP-based authentication to increase security. Correctness and privacy have been a predominant concern in the election process; hence the blockchain-based system can be the future of elections. There are still areas of security where this system may lack, like candidates can bribe voters to ask for their OTP and vote on their behalf to themselves. We can include face detection authentication for such activities so that no other person can log in with another's person OTP. Also, we can extend this system to allow voters to change their vote as many times as they want so that even if a candidate bribes a voter to cast a vote of their choice, the voter can later change their vote to a candidate of their choice.

For places where very few people have internet, the election can be conducted both online and offline. Pooling booths with blockchain-based machines to vote could be set up so that the elections can be conducted easily over the area where the people have fewer smart devices or face internet connectivity issues.

References

1. About Exonum, available at: <https://exonum.com/about> Last accessed on 27 Jan 2022
2. aadhaar-validator, available at: <https://www.npmjs.com/package/aadhaar-validator>. Last accessed on 27 Dec 2021
3. Ayed, A.B.: A conceptual secure blockchain-based electronic voting system. *Int. J. Net .Sec. Appl.* **9**(3), 1–9 (2017)
4. Barnes, A., Brake, C., Perry, T.: Digital voting with the use of blockchain technology. In: Team Plymouth Pioneers-Plymouth University (2016)
5. ETOline (2020) <https://economictimes.indiatimes.com/news/elections/lok-sabha/india/what-are-evms/articleshow/68807699.cms?from=mdr>
6. Hall, T. E.: Primer on the US election system. In: International Foundation for Electoral Systems (2012)
7. Hanifatunnisa, R., Rahardjo, B.: Blockchain based e-voting recording system design. In: 2017 11th International Conference on Telecommunication Systems Services and Applications (TSSA), pp. 1–6. IEEE (2017)
8. Hao, F., Ryan, P.Y.A., Zielinski, P.: Anonymous voting by two-round public discussion. *IET Inf. Sec.* **4**(2):62–67 (2010)
9. Hardwick, F.S., Gioulis, A., Akram, R.N., Markantonakis, K.: E-voting with blockchain: An e-voting protocol with decentralisation and voter privacy. In: 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp. 1561—1567. IEEE (2018)
10. Hjalmarsson, F. Þ., Hreiðarsson, G. K., Hamdaqa, M., Hjálmtýsson, G.: Blockchain-based e-voting system. In: 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), pp. 983—986. IEEE (2018)
11. i-voting, available at: <https://e-estonia.com/solutions/e-governance/i-voting/> Last accessed on 27 June 2021
12. Khan, K.M., Arshad, J., Khan, M.M.: Investigating performance constraints for blockchain based secure e-voting system. *Futur. Gener. Comput. Syst.* **105**, 13–26 (2020)
13. Liu, Y., Wang, Q.: An e-voting protocol based on blockchain. *IACR Cryptol. ePrint Arch.* **2017**, 1043 (2017)
14. McCorry, P., Shahandashti, S.F., Hao, F.: A smart contract for boardroom voting with maximum voter privacy. In: International Conference on Financial Cryptography and Data Security, pp. 357—375. Springer, Cham (2017)
15. Ofori-Dwumfuo, G.O., Paatey, E.: The design of an electronic voting system. *Res. J. Inf. Technol.* **3**(2), 91–98 (2011)
16. Sadia, K., Masduzzaman, M., Paul, R.K., Islam, A.: Blockchain based secured e-voting by using the assistance of smart contract (2019). arXiv preprint [arXiv:1910.13635](https://arxiv.org/abs/1910.13635)
17. Sasson, E.B., Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E., Virza, M.: Zerocash: Decentralized anonymous payments from bitcoin. In: 2014 IEEE Symposium on Security and Privacy (SP), pp. 459–474. IEEE (2014)
18. Springall, D., Finkenauer, T., Durumeric, Z., Kitcat, J., Hursti, H., MacAlpine, M., Halderman, J.A.: Security analysis of the Estonian internet voting system. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp. 703—715(2014)

19. Vassil, K., Solvak, M., Vinkel, P., Trechsel, A.H., Alvarez, R.M.: The diffusion of internet voting usage patterns of internet voting in estonia between 2005 and 2015. *Gov. Inf. Q.* **33**(3), 453–459(2016)
20. Wang, B., Sun, J., He, Y., Pang, D., Lu, N.: Large-scale election based on blockchain. *Proc. Comput. Sci.* **129**, 234–237 (2018)
21. Wicaksana, A.: Towards secure and auditable e-voting system with go ethereum. *Turk. J. Comp Mathe Edu (TURCOMAT)* **12**(10), 3006–3012 (2021)
22. Yi, H.: Securing e-voting based on blockchain in P2P network. *EURASIP J. Wirel. Commun. Netw.* **2019**(1), 1–9 (2019)

DevOps Tools: Silver Bullet for Software Industry



Divya Srivastava, Madhushi Verma, Shashank Sheshar, and Madhuri Gupta

Abstract DevOps is a development and operations practice where engineers collaborate throughout the full-service lifecycle, from design to development to production support. DevOps is a newly emerging field in the field of Software application. Most of the giant companies have now shifted towards DevOps practices as it channelizes the development and operation process for any software development. The present chapter aims to provide the DevOps information for the basic to reach out to maximum people working in or planning to shift towards DevOps. Starting from the definition, the phases, tools, and security are discussed in the present paper. Each phase of the DevOps life cycle is discussed with the tools used in that phase. The DevOps practices are powered with several tools to provide end-to-end automation in software development. The present paper presents the basic knowledge of prevalent tools available for DevOps practices. Along with the DevOps automation, the chapter also gives an overview of DevOps security, its need, and its tool. The chapter covers the software phases and the tools used to automate it. It also provides information regarding the tool platform, availability, and usage. To emphasize more on DevOps, the chapter has also summarized the industrial and academic opportunities in DevOps.

Keywords DevOps · DevOps tools · Docker · Ansible · Git · Jenkins · DevOps security · CICD

These authors contributed equally to this work.

D. Srivastava (✉) · M. Verma · M. Gupta
CSET, Bennett University, Greater Noida, Uttar Pradesh, India
e-mail: divya.srivastava@bennett.edu.in

M. Verma
e-mail: madhushi.verma@bennett.edu.in

M. Gupta
e-mail: madhuri.gupta@bennett.edu.in

S. Sheshar
CSED, Thapar University, Patiala, Punjab, India

1 Introduction

DevOps has emerged as a “silver bullet” for improving software development efficiency and its operations. In 2009, Patrick Debois gave the word DevOps, which is a combination of two words: Development and Operation. It is an umbrella that describes all the phases of software development, coding, building, testing, delivery, and deployment, as shown in Fig. 1.

DevOps practices [1] help the organizations produce a quality product in less time. The state-of-DevOps report 2020 shows that even during the tough times in the year 2020, DevOps practices have yielded excellent results in “software delivery” and “change management.” DevOps practices have helped IT organizations faster and easy delivery of quality products on a secured platform.

DevOps is an advancement of Agile Software Development [2]. DevOps practices advocated CICD (Continuous Integration and Continuous Delivery) that makes it fast and more efficient as compared to pioneer agile techniques [3, 4]. DevOps principles work to collaborate people towards a common business goal by focusing on People and Culture, Process and Practices, and Tools and Technologies. Thus, it covers all the areas required by any organization for speedy and efficient product delivery. Traditional software development practices focused only on the development part, and there used to be a chain for every phase of software development. The traditional practices were slow, and there has been an imbalance between delivery and unstable production in most cases. DevOps was thus created to balance by bringing everyone involved in software development and deployment under one umbrella. It encourages collaboration between the development and operational team; therefore, a more collaborative and efficient product is produced. DevOps is not only for developers and operators but also a combined effort of managers, engineers, and administrators. Every single person has an important role to play and contribute to DevOps.

DevOps is no doubt acting as a silver bullet in IT Industry but it comes along with a cost of security. DevOps Security also termed DevSecOps is a collection of practices that integrate software development (Dev), IT operations (Ops), and security (Sec) to improve an organization’s ability to run and deliver applications and services quickly and securely, while DevOps practices bring tremendous advantages to IT sector

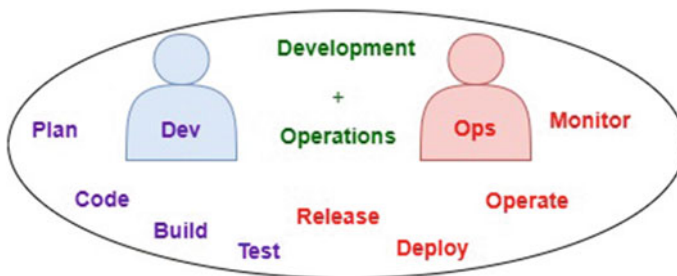


Fig. 1 DevOps: development and operation

by keeping the applications safe and secure. DevOps environments need to protect services, applications, and tools along with traditional development environments.

1.1 Background

Many renowned companies like Google, Microsoft, Amazon, etc., are now providing different tools and platforms for DevOps practices. The main reason behind DevOps’ popularity is that people interested in system administration and software development find space while working in DevOps. They work together towards a common goal and achieve an efficient product quickly. Amazon refers to DevOps as a “combination of cultural philosophies” because it combines approaches used by software developers, testers, managers, operations, and other expertise. There has been great growth in DevOps practices in the recent few years, both in the field of industry and research [5]. The graph shown in Fig. 2 shows that the research in the field of DevOps geared up from 2015.

DevOps practices work in a cross-functional environment and thus require a set of tools that allow this. It is equipped with automated tools, making it a “SILVER Bullet” for software industries. Sometimes these tools are even termed “Toolchains” as DevOps practices focus on a collaborative work environment. The paper aims to build knowledge of various DevOps tools based on their respective job. It also provides the platform for the tool and is also available in terms of Open Source (Os), Enterprise (En), Paid (Pd), Free(Fr), and Free+paid (Fm (Free trial version and later on payment basis)).

The rest of the paper is organized as follows: DevOps Lifecycle is discussed in Sect. 2. DevOps Tools are discussed in Sect. 3. Section 4 discusses the use of DevOps in the Industry and Education Filed. Section 5 concludes the chapter along with the future prospects of the DevOps market all over the world.

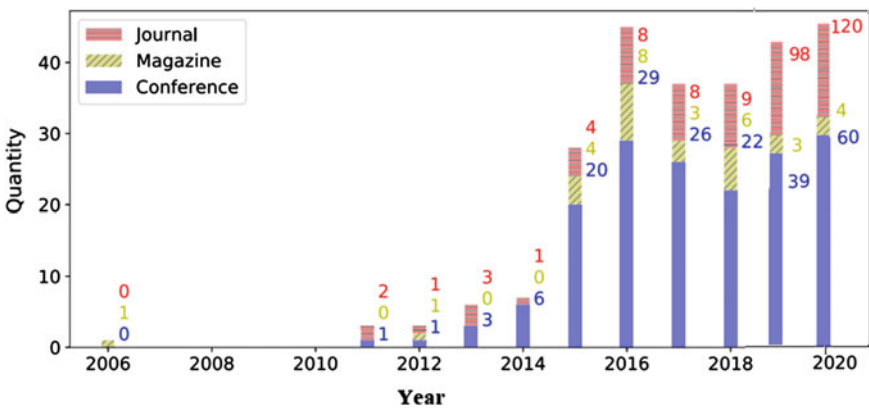


Fig. 2 DevOps Publications based on type and year [5]

2 DevOps Life Cycle

The DevOps lifecycle enhances advancement processes from beginning to end and connects with the association in a nonstop turn of events, bringing about quicker conveyance times. This cycle fundamentally comprises of the accompanying seven phases.

2.1 *Continuous Development*

Continuous Development includes arranging and coding the product. The whole improvement process becomes separated into more modest advancement cycles. This interaction makes it simpler for the DevOps group to speed up the general programming advancement process. This stage is instrumental in planning the vision for the whole improvement cycle, empowering engineers to comprehend project assumptions completely. Through this, the group begins imagining its ultimate objective too. There are no DevOps instruments needed for arranging; however, numerous adaptation DevOps tools are utilized to keep up with code. This course of code upkeep is called source code management or source version control system. Famous DevOps tools for source code upkeep incorporate JIRA, Git, Mercurial, and SVN. Also, there are various DevOps tools for bundling the codes into executable documents, like Ant, Gradle, and Maven. These executable records are then sent to the next phase of the DevOps lifecycle.

2.2 *Continuous Integration*

Continuous Integration (CI) incorporates various advances connected with the execution of the test interaction. Alongside this, customers also give feedback/data to add new elements to the application. Most changes occur in the source code during this stage. CI turns into the center for settling these regular changes every day or month-to-month premise. Developing code is a blend of the unit and reconciliation testing, code survey, and bundling. Since developers roll out continuous improvements, they can rapidly detect issues (if any) and resolve them at the beginning phase. This stage encounters persistent new code functionalities with the current source code. Jenkins is one of the most well-known instruments for continuous combination. It helps in getting the refreshed code and setting up an executable form.

2.3 *Continuous Testing*

Next in the DevOps lifecycle is the Continuous Testing stage, wherein the created code is tried for bugs and mistakes that might have advanced into the code. This is

where quality examination (QA) assumes a significant part in really taking a look at the ease of use of the created programming. Fruitful consummation of the QA interaction is essential in deciding if the product meets the customer's details. DevOps automated tools, like JUnit, Selenium, and TestNG, are utilized for continuous testing, empowering the QA group to examine various code bases simultaneously. Doing this guarantees no blemishes in the usefulness of the created programming. Also, Docker compartments are utilized in continuous testing to copy the real test environment. A Docker holder is an independent, lightweight executable container with everything to run an application: framework apparatuses, framework libraries, runtime code, and settings. Automated testing is done on DevOps tools like Selenium, after which the reports are produced on another computerization apparatus, for instance, TestNG. Automation of the whole testing stage additionally becomes conceivable with the assistance of the ceaseless combination apparatus Jenkins. Automated testing assumes an indispensable part in saving time, work, and exertion.

2.4 Continuous Deployment

Continuous Deployment (CD) ensures easy application deployment without influencing the application's presentation. It is important to guarantee that the code is sent unequivocally to all suitable waiters during this stage. This cycle disposes of the requirement for planned deliveries and speeds up the input component, permitting developers to resolve all the issues more rapidly and with more noteworthy exactness. The containerization concept in DevOps assists with accomplishing consistent deployment through various configuration management tools. A containerization device like Vagrant executes consistency across test, improvement, arranging, and creation conditions. Containerization manages to carry virtualization to the level of a working framework over various operating systems.

Continuous Deployment ensures to help the association to have a reliable testing environment setup. Configuration management tools hold significant worth in the persistent deployment stage. It includes arranging and keeping up with consistency in the useful prerequisite of the application. Well-known DevOps devices utilized for design the executives incorporate Ansible, Puppet, and Chef that assist with executing a speedy arrangement of new code.

2.5 Continuous Monitoring

Monitoring the execution of a product is essential as it helps to decide the general adequacy of the product/application. This stage processes significant data about the developed application. Through nonstop observing, developers can recognize general trends and ill-defined situations in the application where more exertion is required. Consistent observing is a functional stage where the goal is to improve the general productivity of the product application. Additionally, it screens the exhibition of the

application also. In this way, it is one of the essential periods of the DevOps lifecycle. Different framework mistakes, for example, “server not reachable”, “low memory”, and so forth, are settled in the consistent monitoring stage. It likewise keeps up with the accessibility and security of the administrations. Network issues and different issues are consequently fixed during this stage at the hour of their discovery. Tools like Nagios, Splunk, Sensu, ELK Stack, and NewRelic are used by the monitoring team to screen client activities for inappropriate conduct. Accordingly, during continuous monitoring, engineers can proactively look at the general strength of the framework. Proactive checking works on the framework’s dependability and usefulness and decreases upkeep costs. In addition, significant issues are straightforwardly answered to the development team to be adjusted in the underlying stages.

2.6 Continuous Feedback

Continuous Feedback is an essential input to learn and investigate the ultimate result of the application. It establishes the vibe for working on the current form and delivering another adaptation in view of feedback given by the customers. This stage plays a vital role in improving the existing version of the application based on the feedback input. This input is the data assembled from the customer’s end. The data is huge, as it conveys all the information about the existing version of the product and its related issues. It additionally contains ideas given by end clients of the product.

2.7 Continuous Operations

The last stage in the DevOps lifecycle is the briefest and least demanding to handle. Continuity is at the core of all DevOps tasks that automate release processes, permit developers to recognize issues rapidly, and assemble better forms of programming items. Continuation is critical to redirections and other additional means that block improvement. Development cycles in ceaseless tasks are more limited, permitting associations to publicize continually and speed up the general chance to showcase the application.

Thus, all the stages in DevOps Lifecycle improve the worth of applications by improving them and being more proficient. Therefore, it has emerged as a silver bullet in the software industry.

3 DevOps Tools

DevOps Tools have been categorized based on key aspects they perform. The categorization of tools is fuzzy i.e. one tool may fall into one or more categories. The tools are broadly categorized as:

- Code
- Build
- Test
- Delivery
- Deployment
- Monitor

3.1 Code

In this phase, the developers decide the toolkit to be used and install the plugins required to develop the application. Coding Nomenclature, clean code practices, etc., are part of this phase. Code Development, Merging of Codes, Code Review, Source Code Management, etc. Git is one of the most popular Source Code Management (SCM) used to maintain software versioning systems.

3.1.1 Code Management Tools

Git is the most popular tool for code repositories. It is an open source (Os) and is used for SCM. It operates on a client–server basis, where a central code repository is maintained. The central repository acts as a hub for both clients and developers, and both of them can simultaneously download the code from there. Linus Torvalds [6] is the creator of Git. Git makes it easier for remote teams to collaborate.

Git and its variants are used for version control as shown in Fig. 3, and all of them are OS. The other popular version control tools are Atlassian Bitbucket[7] whose trial version is available free, whereas, after that, it allows the user after payment (Free + Premium (Fm)). Similarly, another popular tool for version control is subversion (SVN) [8] and is made available as OS by Apache License.

3.2 Build

Once the developer shares the code on the repository and requests to merge their new code with existing ones on the repository, this is where the build operation comes in role. Building the codebase, Running end-to-end series, Integrating the code, Unit Testing, etc. If the build operation fails at any point of time, the developer is notified to resolve the issue. It minimizes the integration challenges that arise while working on a common codebase. The errors are exposed in the early stages of the development lifecycle by continually verifying code changes into a common repository and executing builds and tests.

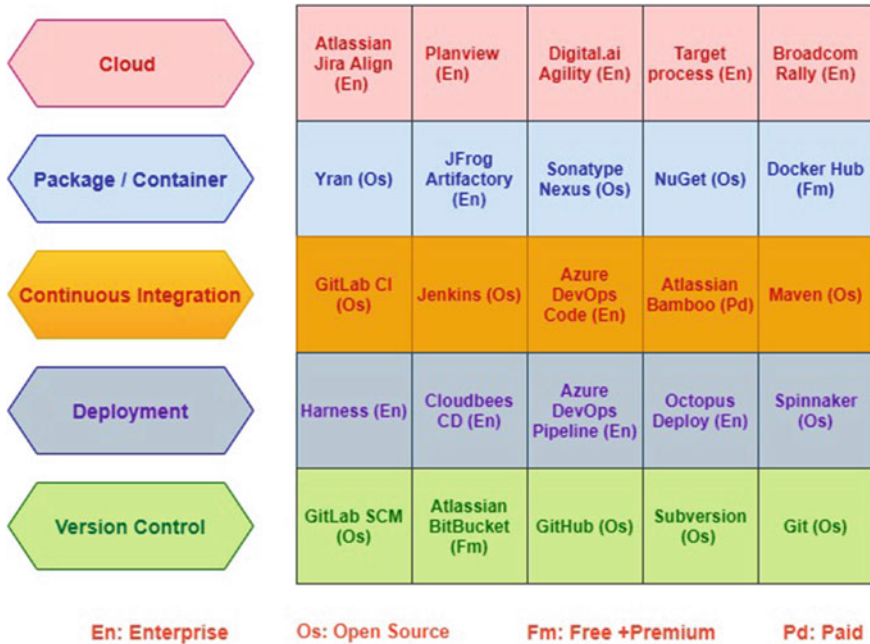


Fig. 3 DevOps tools











3.2.1 Build Management Tools

There are different tools available for the building phase. Some of the popular tools, such as Gradle [9], Jenkins [10], Docker [11], Kubernetes [12], are listed in Table 1. All of them are OS and are easily available.

3.3 Test

After the code is built, it is moved to the test environment. Test operations have been performed both ways, Manual as well as Automated. One of the most common forms of Manual testing is User/Client Acceptance Testing. People use the app as a client to indicate any bugs or enhancements that need to be fixed before going live. In the case of Automated Testing, security scanning against the application, checking for changes to the infrastructure and compliance with hardening best practices, evaluating the application’s performance, and load testing are all examples of automated tests. The type of testing done during this phase depends on the organization and what is relevant to the application, but this stage may be thought of as a testbed where you can plug in new features.

Table 1 Popular DevOps tools and their usage

DevOps tools			
Tool name	DevOps phase	Availability	Icon
Git	Code, Build	Os	
Gradle	Build	Os	
Selenium	Test	Os	
Jenkins	Build, Test, Deploy	Os	
Puppet	Deploy, Operate	Os	
Chef	Deploy, Operate	Os	
Docker	Build, Deploy, Operate	Os	
Kubernetes	Build, Deploy, Operate	Os	
Ansible	Deploy, Operate	Os	
eG Enterprise	Monitor	Fm	

3.3.1 Test Management Tools

Several tools are available for automated tests. Selenium [13] is one of the most popular open-source tools for automated tests. It is used for automation testing of web-based applications. Selenium IDE is a simple tool for anyone to use, and it's built to operate with Firefox. The Firefox plugin allows users to record and replay test actions using a GUI (Graphical User interface). It also supports the creation and execution of test cases without knowing scripting languages. It supports Java,

C#, Perl, Python, Ruby, and Groovy. Another tool available for test automation is JUnit [14], which is the unit testing tool for Java programming language and is freely available over the internet. NEOTYS NEoload [15] is a paid testing tool used by fortune companies to do the performance and load testing. It is one of the best tools covering all the enterprise testing requirements. Native hybrid and Mobile based applications can perform the test by using the enterprise version of Appium [16] automated tool. The other tools available are Tricentis Tosca [17], Sauce Labs [18], Parasoft [19], JMeter [20], Cucumber [21], etc.

3.4 Delivery

It is also called a Release phase in DevOps. It is when the application is ready to be delivered or released for deployment over the production environment. Thus, the first milestone is achieved in DevOps Pipeline. Depending on an organization's DevOps maturity, any build that reaches this pipeline level may be deployed automatically. Feature flags allow developers to switch off new features so that users don't notice them until they're ready to use. This is the nirvana of DevOps, and it's how companies manage to release numerous versions of their product at the same time.

3.4.1 Delivery Management Tools

The process of managing, planning, scheduling, and controlling the entire process of producing a piece of software through various phases of development and environments—such as testing and distributing software releases—is known as release management. There are several tools for release management. Some of the common tools are XL Release [22], IBM UrbanCode Release [23], Plutora Release [24], CA Release Automation [25], etc.

3.5 Deployment

The process of releasing the application to the production is deployed. There are several tools and methods that can be used to automate the release process and ensure that releases are reliable and have no downtime.

3.5.1 Deploy Management Tools

To automate the deployment tools, the common tools available are Jenkins, Puppet [26], Docker [26], Kubernetes [26], Chef [27], Ansible [28], etc.

3.6 *Monitor*

This phase observes the application being launched in the market. It is based on the market review, application performance, customer feedback/reviews, etc.

3.6.1 **Monitor Management Tools**

Several tools are available for monitoring the delivered product. Top in the list are Sensu [29], Pager Duty [30], Datical Deployment Monitoring Console [31], Splunk [32], etc.

Sensu is both infrastructure as well as application monitoring solution. It can be used to analyze, measure, and monitor the infrastructure, service health, application health, and business KPIs, among other things. Sensu seeks to answer modern-day difficulties in modern infrastructure platforms by combining static, dynamic, and ephemeral infrastructure at scale. Sense isn't a SaaS solution, but it does provide you with complete control over the availability of your monitoring system. It is widely used as it facilitates dynamic addition and deletion of the client, sends notifications and alerts, works with multi-tiered networks, and most importantly, fulfills the automation needs.

DevOps practices are implemented using the tools mentioned above, having functionalities like Configuration Management, Version Control, Test, Deploy, etc. The knowledge of these tools opens the areas for the engineer for various designations in Industry. Some of them are shared in the section below.

4 **DevOps in Industry and Education**

DevOps has helped industries maximize production in terms of quality and quantity. The DevOps practices following CICD help in continuous integration and delivery based on feedback. The popularity of DevOps in the industry is very well described in [33] and the authors termed it Industry 4.0 Industries are providing internships and recruiting people for several designations in the field of DevOps like:

1. DevOps Architecture
2. DevOps Tester
3. DevOps Developer
4. Release Manager
5. DevOps Security Expert

DevOps has recently gained so much popularity among the industries that now it is being run as a specialization course in some of the reputed universities both at undergraduate as well as postgraduate levels. Some of the postgraduate programs available are:

1. DevOps and Continuous Software Engineering by the University of Limerick [34]
2. DevOps by Letterkenny Institute of Technology, Ireland [35]
3. Software Engineering (Cloud Computing) by Torrens University, Australia [36]
4. Cloud Computing and DevOps by IIT Roorkee & Wiley, India [37]
5. Cloud and DevOps by EICT Academy IIT Guwahati & Microsoft, India [38]
6. Software Development—Specialization in DevOps by IIIT Bangalore, India [39]
7. DevOps Postgraduate Program by Caltech CTME & Simplilearn [40]

5 Conclusion and Future Perspective

The present paper discussed various DevOps tools, their usage, availability, etc. DevOps practices have improved the continuous delivery of valuable software. It supports modern interaction techniques that help in quick and reliable feedback for fast and efficient software delivery. DevOps has changed the way it needs to achieve IT security. When moving from long-term planned delivery of monolithic applications to an agile development environment, security needs to be deeply integrated into the development and operational processes. DevSecOps begins with a secure development lifecycle for the services and applications created and defined security patterns and ends with automated security for automated operations.

The paper has presented its importance in the industry as well as academics. The Global market insight predicts the DevOps Market growth by 2026 shown in Fig. 4. Japan and the U.S have also given their prediction for the development of DevOps as shown in Figs. 5 and 6, respectively. The market prediction shows that DevOps has an excellent future perspective and will build a better software world.

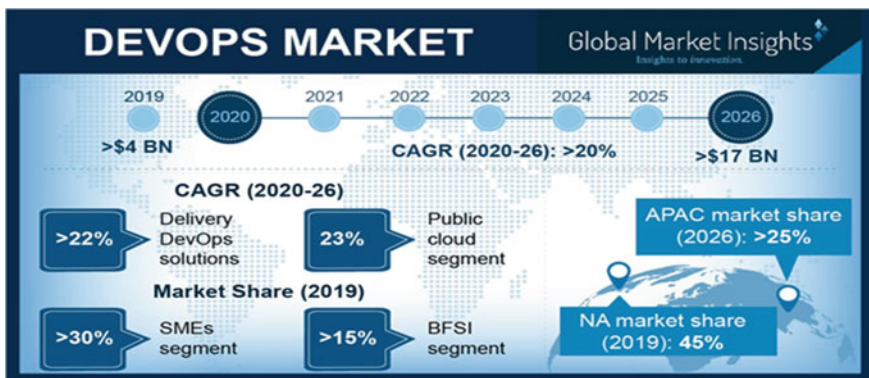


Fig. 4 DevOps market [41]

Rising trend of self-hosting automation processes augmenting the demand for on-premise deployment model in Japan

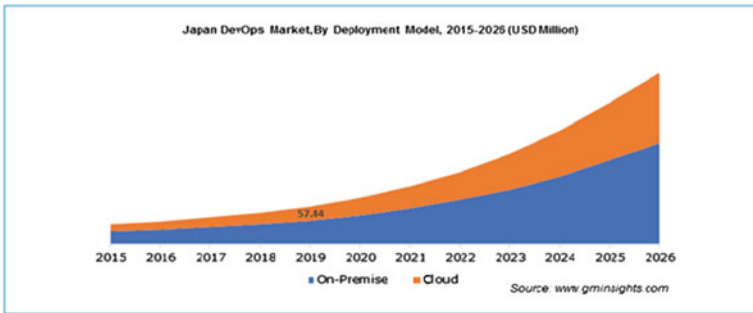


Fig. 5 DevOps market Japan [41]

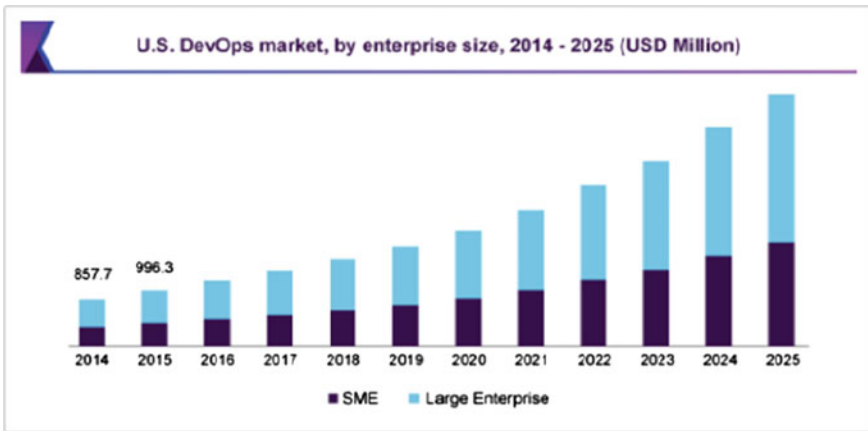


Fig. 6 DevOps market [41]

References

1. <https://www.netsparker.com/devops-security-tools/>
2. Christensen, H.B.: Teaching DevOps and cloud computing using a cognitive apprenticeship and story-telling approach. In: Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education (2016)
3. Kroll, P., Kruchten, P.: The Rational Unified Process Made Easy: A Practitioner’s Guide to the RUP: A Practitioner’s Guide to the RUP. Addison-Wesley Professional (2003)
4. Pressman, R.S.: Software Engineering: A Practitioner’s Approach. Palgrave Macmillan (2005)
5. Leite, L., et al.: A survey of DevOps concepts and challenges. ACM Comput. Surv. (CSUR) **52**(6), 1–35 (2019)
6. <https://www.weave.works/blog/15-years-of-git>
7. Light, J., Pfeiffer, P., Bennett, B.: An evaluation of continuous integration and delivery frameworks for classroom use. In: Proceedings of the 2021 ACM Southeast Conference (2021)

8. Pingrong, L., Xiaoquan, S., Junqin, Y.: Research on the application of DevOps in the smart campus of colleges and universities. *J. Phys.: Conf. Ser.* **1883**(1). IOP Publishing (2021)
9. Jokinen, O.: Software development using DevOps tools and CD pipelines: a case study (2020)
10. Agarwal, V., Krishnappa, H.K.: Robot framework integration with Jenkins: a blessing for automation (2021)
11. McKendrick, R.: *Mastering Docker: Enhance Your Containerization and DevOps Skills to Deliver Production-Ready Applications*. Packt Publishing Ltd (2020)
12. Paavola, E.: Managing multiple applications on kubernetes using GitOps principles (2021)
13. García, B., et al.: A survey of the selenium ecosystem. *Electronics* **9**(7), 1067 (2020)
14. Batra, P., Jatain, A.: Software Quality Enhancement Using Hybrid Model of DevOps, Intelligent Systems, pp. 281–288. Springer, Singapore (2021)
15. Srivastava, N., Kumar, U., Singh, P.: Software and performance testing tools. *J. Inf. Electr. Electron. Eng.* **2**(01), 1–12 (2021)
16. da Silva Lima, G.B., et al.: Devops methodology in game development with unity3D (2020)
17. Atesogullar, D., Mishra, A.: Automation testing tools: a comparative view. *Int. J. Inf. Technol. Sec.* **12**(4) (2020)
18. Jayasri, A.S.V.: An extensive risk-mitigating framework for continuous testing using DEVOPS (2020)
19. Angara, J., Prasad, S., Sridevi, G.: DevOps project management tools for sprint planning. *Estim. Execut. Matur., Cybern. Inf. Technol.* **20**(2), 79–92 (2020)
20. Koltun, A., Pańczyk, B.: Comparative analysis of web application performance testing tools. *J. Comput. Sci. Inst.* **17**, 351–357 (2020)
21. Haver, T.: Cucumber 3.0 and beyond, PNSQC proceedings (2018)
22. Solouki, S.: *Knowledge Management Practices in DevOps*, Diss. Université d'Ottawa/University of Ottawa (2020)
23. Österberg, G.: A systematic literature review on DevOps and its definitions. *Adopt. Benef. Chall.* (2020)
24. Parashar, R.: Path to success with CI/CD pipeline delivery. *Int. J. Res. Eng. Sci. Manag.* **4**(6), 271–273 (2021)
25. Singh, R.: DevOPS now and then (2020)
26. Uphill, Thomas, et al.: *DevOps: Puppet, Docker, and Kubernetes*. Packt Publishing Ltd (2017)
27. Vadapalli, S.: *DevOps: continuous delivery, integration, and deployment with DevOps: dive into the core DevOps strategies*. Packt Publishing Ltd (2018)
28. McAllister, J.: *Implementing DevOps with Ansible*, vol. 2. Packt Publishing Ltd (2017)
29. Bovina, S., Michelotto, D.: The evolution of monitoring system: the INFN-CNAF case study. *J. Phys.: Conf. Ser.* **898**(9). IOP Publishing (2017)
30. Mishra, R.M., More, M.: A qualitative study of DevOps. **7**(1) (2021)
31. Kubryakov, K.: *Deployment and testing automation in web applications: implementing DevOps practices in production* (2017)
32. Subramanian, K.: *Introducing the Splunk Platform. Practical Splunk Search Processing Language*, pp. 1–38. Apress, Berkeley (2020)
33. Hasselbring, W., et al.: Industrial devops. In: 2019 IEEE International Conference on Software Architecture Companion (ICSA-C). IEEE (2019)
34. <https://www.ul.ie/gps/devops-and-continuous-software-engineering-graduate-diploma>
35. https://www.lyit.ie/CourseDetails/D202/LY_KDVOP_M/DevOps
36. <https://www.torrens-international.com/programs/postgraduate-degrees/graduate-diploma-of-software-engineering-cloud-computing/>
37. <https://www.wileyx.com/iit-roorkee-wiley-post-graduate-certification-in-cloud-computing-and-dev-ops>
38. <https://intellipaat.com/post-graduate-certification-cloud-and-devops/>
39. <https://www.iiitb.ac.in/executive-post-graduate-programme-in-software-development>
40. <https://www.simplilearn.com/pgp-devops-certification-training-cours>
41. <https://www.gminsights.com/industry-analysis/devops-market>

Robust and Secured Reversible Data Hiding Approach for Medical Image Transmission over Smart Healthcare Environment



K. Jyothsna Devi, Priyanka Singh, José Santamaría, and Shrina Patel

Abstract With the rapid progress of cloud computing, there has been a marked improvement in the development of smart healthcare applications such as Internet of Medical Things (IoMT), Telemedicine, etc. Cloud-based healthcare systems can efficiently store and communicate patient electronic healthcare records (EHR) while allowing for quick growth and flexibility. Despite the potential benefits, identity violation, copyright infringement, illegal re-distribution, and unauthorized access have all been significant. To address all these breaches, in this paper, a reversible medical image watermarking scheme using interpolation is proposed. The medical image is partitioned into Border Region (BR), Region of Interest (ROI), and Region of Non-interest (RONI) regions. BR is used for embedding integrity checksum code generated from ROI for tamper detection. RONI is used for embedding watermark. To ensure complete recovery of ROI and high embedding capacity, ROI is compressed before embedding. To ensure high-security compressed ROI, hospital emblem and EHR merged and then encrypted using a random key generated from Polybius magic square to get higher security. The proposed scheme is proved to take less computational time as there are no complex functions used in the embedding. The experiments performed on the proposed scheme is proved to have high imperceptibility, robustness, embedding capacity, security, and less computational time. All these confirm that the proposed approach is a potential candidate for suitable in smart healthcare environment.

K. Jyothsna Devi (✉) · P. Singh

Department of Computer Science and Engineering, SRM University, Amaravati 522502, Andhra Pradesh, India

e-mail: jyothsna_devi@srmmap.edu.in

P. Singh

e-mail: priyanka.s@srmmap.edu.in

J. Santamaría

Department of Computer Science, University of Jaén, Jaén, Spain

S. Patel

Department of Computer Science and Engineering, Sardar Vallabhbhai Patel Institute of Technology, Vasad 388306, Gujarat, India

e-mail: shrinapatel.comp@svitvasad.ac.in

Keywords Random key · Reversible data hiding · NMI · Smart healthcare

1 Introduction

The Internet of Things (IoT) has infiltrated nearly every aspect of life. One of the major areas that makes substantial use of IoT infrastructures and solutions is smart healthcare. With the usage of wearable and mobile devices, IoT-based smart healthcare systems have greatly brought value to the health sector [1]. Smart healthcare systems intelligently manage patient monitoring infrastructures and positive healthcare will grow as a result of recent advancements in cloud computing, blockchain, and big data. Also, smart healthcare adoption lowers costs and improves healthcare. Smart healthcare is a platform that brings together a diverse group of participants and stakeholders, including doctors, patients, hospitals, and healthcare diagnosis laboratories. Figure 1 illustrates smart healthcare infrastructure.

Smart healthcare was used to assess patient electronic healthcare record (EHR) in smart cities. Medical image management, health record keeping, and patient–doctor communication are all important factors. Managing medical image transmission from diagnosis laboratories to diagnosis practitioners is a critical task that requires higher security. Maintaining confidentiality, handling authentication, and assuring

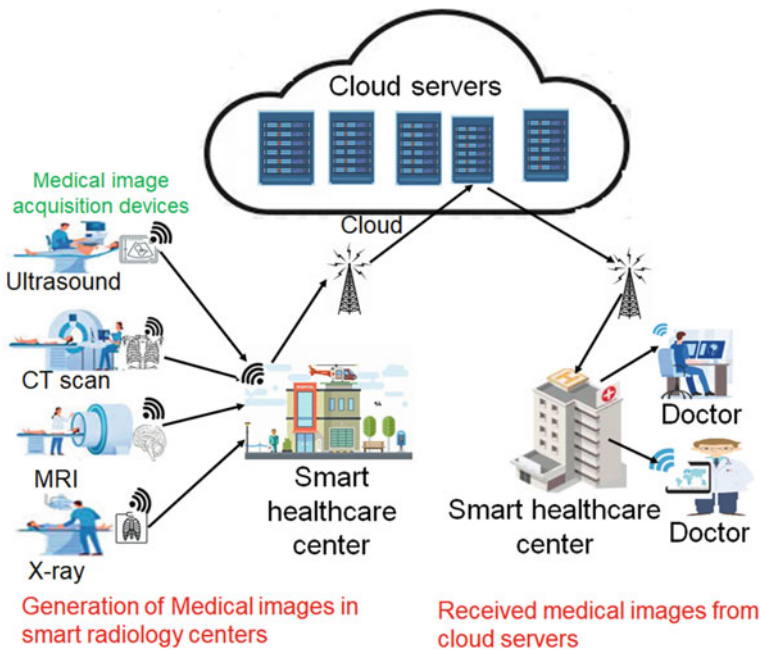


Fig. 1 Secure medical image transmission iSmart healthcare infrastructure

data integrity are all required for secured transmission of medical images. When it comes to the exchange of EHR, authentication is a vital part of smart healthcare. For smart healthcare, robust secure frameworks are essential. Therefore, confidentiality, integrity, authentication, and authorization are thus the security criteria for smart healthcare in medical image transmission. To address all these security requirements, medical image watermarking (MIW) is the prominent solution. MIW is a method of concealing a medical image with a hospital emblem or significant patient information in order to recognize the transmitter. The hospital logo is incorporated into the medical image in smart healthcare applications to identify the origin hospital. EHR is also integrated into medical images in order to determine patient information. In smart healthcare, both the hospital emblem and the EHR serve as a watermark. Furthermore, building a MIW scheme suitable for smart healthcare has a number of challenges, including lower computational costs, higher embedding capacity, and greater security [2]. To overcome these obstacles, to ensure low computational cost, high embedding capacity, authenticity, and good security, a robust highly secured region-based reversible MIW scheme using interpolation is proposed. Reversible MIW is a data hiding technique in which the watermark is implanted in the image and the original image can be restored by extracting the watermark in the same way [3]. Firstly, the cover image is separated into three regions: Border Region (BR), Region of Interest (ROI), and Region of Non-interest (RONI). BR is used for embedding integrity checksum code (IC). ROI is critical for diagnosis; any alterations or distortions lead to an incorrect diagnosis. As RONI is unimportant to diagnosis, compressed ROI is integrated into RONI in the suggested scheme to maintain ROI integrity. Therefore, in the proposed scheme, the hospital emblem, compressed ROI, and EHR are merged and then encrypted before being embedded in RONI to provide authentication, integrity, confidentiality, and security. Embedding can be done with the use of neighbor mean interpolation (NMI).

2 Related Work

In recent years, a lot of research have progressed towards the secured transmission of medical images over the internet, owing to the innovation of 5G. Researchers have developed a number of MIW schemes for embedding and extraction that are using the spatial and transform domains. To ensure total reversibility of the ROI, the scheme given in [4] suggested a DWT-SVD region-based MIW scheme in the transform domain. Yet, this scheme is limited by the lack of EHR security. To attain good EHR security and tamper detection, [5] proposed IWT-based reversible MIW scheme. To provide high robustness of EHR, the scheme proposed in [6] suggested an NSCT-RDWT-SVD-based hybrid transform MIW. For EHR security, this scheme used a chaotic map. However, the transformation techniques proposed in [4–6] meet MIW's imperceptibility, robustness, and integrity characteristics while having a significant computing cost for the embedding and extraction procedure. Low computational cost is a crucial challenge for real-time applications like smart healthcare. The scheme

in [7] proposed a high-security-hybridized compression and then encryption region-based MIW in the spatial domain. This scheme used SHA-256 to generate a digital signature for tamper detection. A reversible MIW employing OPR is proposed in [8] to assure high embedding capacity, although this technique lacks robustness. Some of the schemes related to the suggested scheme were addressed in the literature review. In most of the schemes reviewed above, there is no balance between watermarking characteristics. Each one of the works discussed above has its own unique strengths and weaknesses. This inspired us to suggest a scheme that would ensure all aforementioned watermarking characteristics while also being appropriate for use in a smart healthcare environment.

The following are the key contribution of the proposed scheme:

1. **Higher Security:** Using double encryption, the proposed scheme improves the security of EHR and ROI. The ROI is first compressed using LZW, and then compressed ROI, hospital emblem, and EHR are merged to get watermark. Finally, the Polybius magic square (PM) technique is used to encrypt watermark.
2. **Higher Imperceptibility:** The NMI embedding procedure ensures high imperceptibility.
3. **Low Computational cost:** To ensure linear time complexity, spatial domain embedding with NMI is used.
4. **High embedding capacity:** The use of NMI embedding and ROI compression with LZW allows higher embedding capacity.
5. **High ROI integrity:** The use of LZW compression for ROI and NMI embedding allows complete recovery of ROI and RONI regions.

3 Proposed Work

The proposed reversible region-based scheme can be described in two modules: (1) Watermark embedding and extraction and (2) Watermark encryption and decryption.

3.1 Watermark Embedding and Extraction

In the proposed scheme, the cover medical image is taken and partitioned into three regions, Border Region (BR), Region of Interest (RIO), and Region of Non-interest (RONI), manually by the radiologist. BR and RONI of the medical image are not used for diagnosis by the radiologist. Taking advantage of this fact, BR and RONI are used for watermark embedding. The integrity checksum code (IC) is embedded in LSB of the BR for detection of ROI tampering. In order to recover the entire ROI, the proposed approach compresses ROI before embedding. The binary Hospital emblem is used for the authenticity of the image that has been shared can be traced back. Therefore, compressed ROI, hospital emblem, and EHR are merged to generate a

watermark, which would then be encrypted with a random key. Further, the encrypted watermark is embedded in RONI using NMI. The process of watermark encryption is explained in Sect. 3.2. The watermark embedding and extraction process are as follows:

Watermark Embedding

A cover medical image (CM) of size MN and binary watermark (W) of size AB is considered in the proposed scheme for embedding and extraction process. Firstly, CM is partitioned into three regions, namely BR, ROI, and RONI. The block diagram for the watermark embedding is shown in Fig. 2. BR and RONI used for watermarks embedding. Integrity checksum code (IC) of 32-bit is generated from ROI using image features [9, 10] for tamper detection and embedded in BR of LSB. Further, ROI is compressed using LZW [11]. Finally, compressed ROI, hospital emblem, and EHR are merged to form a binary watermark (W) and then encrypted using PM. The encrypted watermark (W') is embedded in RONI using NMI [12] in spatial domain. The process of NMI is shown in Fig. 3. Firstly, RONI is partitioned into 33 non-overlapping blocks as shown in Fig. 3a, then select corner pixels to form 22 block as in Fig. 3b. NMI is applied on each 22 block to get 33 block as shown in Fig. 3c. Watermark pixel bits are embedded in 33 block of pixel positions (1, 2), (2, 1), (2, 2), (2, 3), (3, 2) as in Fig. 3d. For each 33 block, maximum of 5 bits are embedded in spatial domain as shown in Fig. 3 of d. For handling overflow condition by addition of 1 bit, 255 intensity pixels are removed from RONI before embedding and removed pixels are replaced back after embedding. Finally, BR, ROI, and RONI are combined to get watermarked medical image. The algorithmic steps for watermark embedding is explained in Algorithm 1. The watermark extraction and recovery of original cover medical image are as explained below:

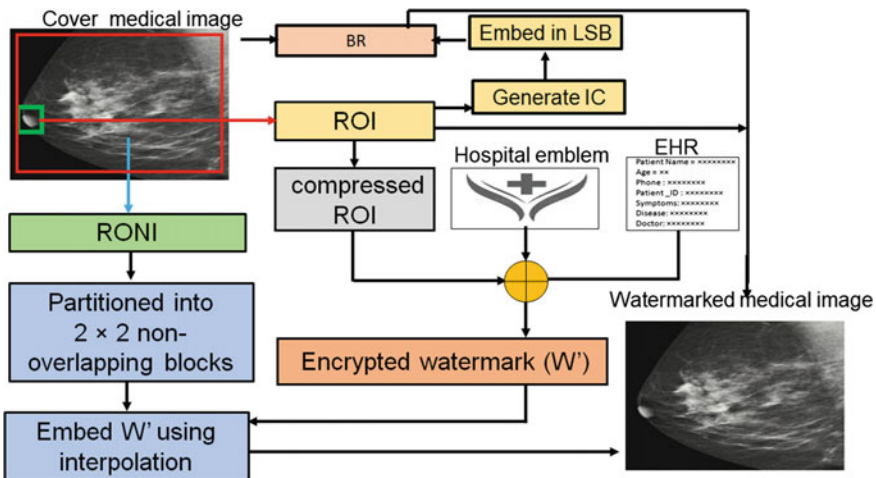


Fig. 2 Block diagram for embedding process

Algorithm 1: Watermark embedding process

- Require:** Cover medical image (CM), binary hospital emblem and EHR
Ensure: Watermarked medical image (WM)
- 1: Partition CM into BR, ROI, RONI manually.
 - 2: Generate IC code for ROI and embed in BR of LSB bit position
 - 3: Compress ROI using LZW
 - 4: Merge compressed ROI, EHR and hospital emblem to get watermark. Then encrypt watermark using random key generated from PM.
 - 5: Embed encrypted watermark in RONI using NMI
 - 6: Finally combine BR, ROI and RONI to get watermarked medical image (WM).

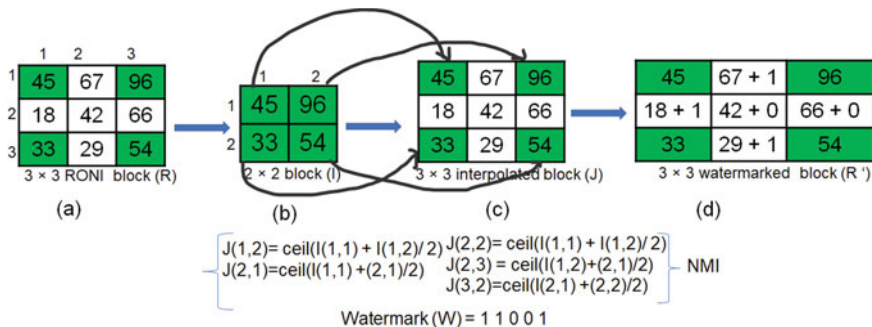


Fig. 3 Watermark embedding using NMI

Watermark Extraction

The embedding watermark is extracted from watermarked medical image by applying embedding process in reverse. To extract watermark, watermarked medical image is partitioned into three regions (BR, ROI, and RONI), and the block diagram for the extraction process is shown in Fig. 4. Applying NMI on RONI to extract encrypted watermark is shown in Fig. 5. Then watermark is decrypted using a secretly received random key to get compressed ROI, hospital emblem, and EHR to verify integrity, confidentiality, and authorization. To verify integrity of the ROI, IC is generated for ROI (IC) and extracted from BR (IC'). If both IC and IC' are equal, ROI is no tampering, otherwise decompress extracted ROI and replace ROI with decompressed ROI. Finally, adjust the modified RONI extracted pixel positions to get the original RONI, and combine ROI, RONI, and BR to get cover medical image.

3.2 Watermark Encryption and Decryption

In the proposed scheme compressed ROI, hospital emblem, and EHR is encrypted using secret random key generated from Polybius magic square (PM) approach to ensure high security. To achieve integrity, confidentiality, and complete recovery of

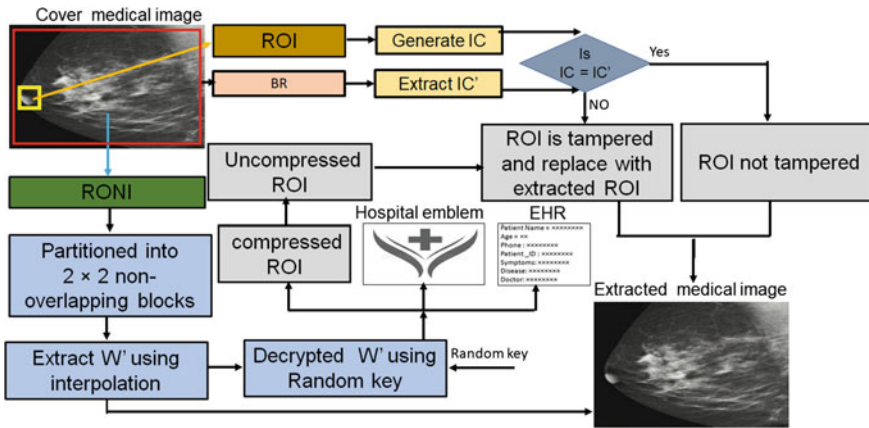


Fig. 4 Block diagram for extraction process

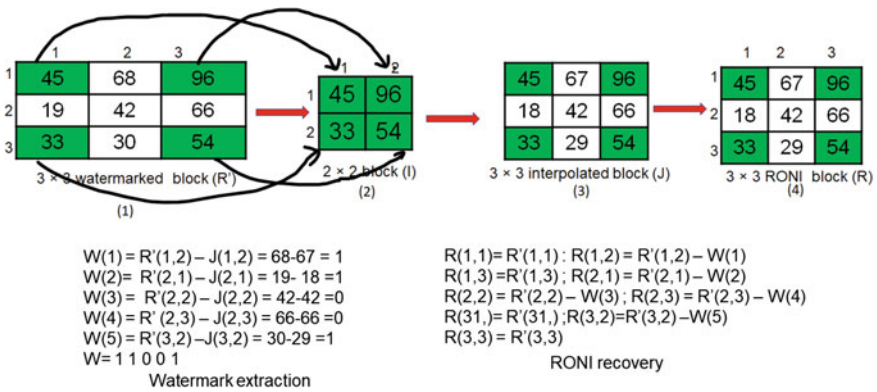


Fig. 5 Watermark extraction using NMI and recovery of original RONI

ROI, the selected ROI portion is compressed using Lempel–Ziv–Welch (LZW). LZW compression is effective in recovering original image and has less computational time for compression and decompression processes. Further binary compressed ROI, binaryEHR, and binary hospital emblem are merged to get binary watermark (W). To ensure higher security, W is encrypted using binary random key with XOR operation. The generation of random key using PM is elucidated below:

Random Key Generation (R_{key})

Binary R_{key} is generated from 66 PM in the suggested scheme. A magic square is structured in such a way that the sum of all the distinct numbers in each row, column, and diagonal elements is same [13]. With an exception for order 2, there are the magic squares for all the orders of N. A 66 magic square is used in the recommended scheme for encryption process. The sum is given as an input and then

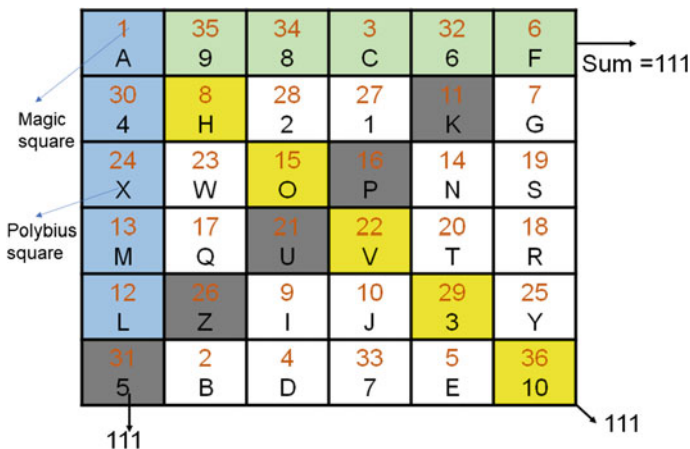


Fig. 6 PM process

the corresponding elements in the rows and columns are arranged accordingly as shown in Fig. 6. Polybius square is used in the proposed encryption process which is an order of 66. Here, there are 36 alphanumeric characters present in this square [14]. These alphanumeric characters are structured are in square by assigning each character A–Z and 1–9 in a sequential way from 1 to 36 as shown in Fig. 6.

Further to generate binary random key (R_{key}), the magic square numbers and Polybius alphanumeric characters are scanned from raster scan fashion from left to right and top to bottom and each scanned value is converted into corresponding 8-bit binary format. Thus, it generates 2424 binary key (original). Now to generate R_{key} , the 2424 binary key is rotated by 900, 1800, and 2700 that gives us the three distinct 2424 binary keys. All the four of the generated 2424 binary keys are combined to get the final binary R_{key} . If the size of the watermark is larger, then geometrical operations (scaling, resize, and translation) are applied on the original 2424 binary key to get new 2424 binary keys. At the receiver end is generated the final R_{key} from the original 2424 binary key. Therefore, 2424 binary key acts as the secret key.

4 Experimental Results and Discussion

Test images for the proposed scheme are taken from the benchmark sources OASIS data set [15]. rescaled a cover medical image to 256256 pixels and an EHR to 120120 pixels and hospital emblem 6464 pixels, respectively. For evaluation, both grayscale and color images medical images with different modalities are used as shown in Fig. 7. The proposed scheme is run on a Windows 10 processor i5 with MATLAB R2017b. The proposed scheme is evaluated in terms of imperceptibility, robustness, security, and computational cost with all of the test images.

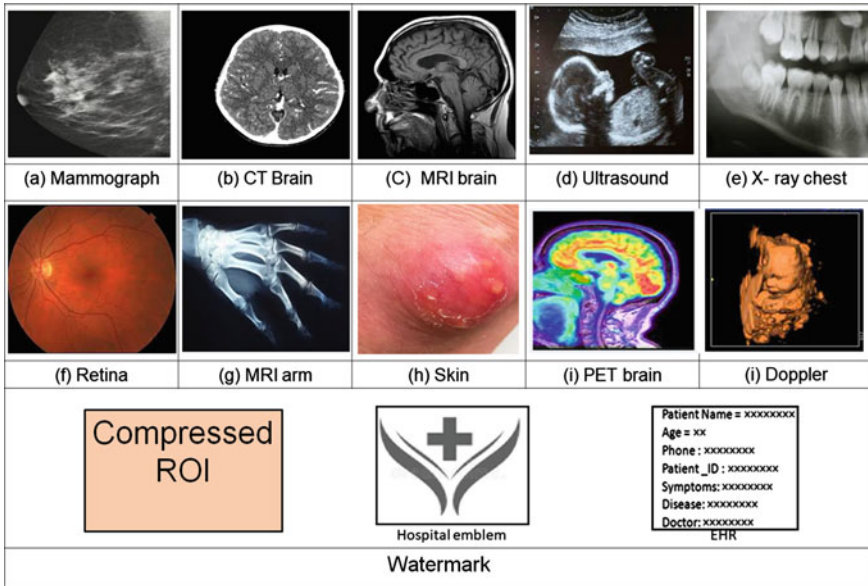


Fig. 7 Grayscale cover medical images (a–e), color cover medical Images (f–i), and Watermark (Compressed ROI + hospital emblem + EHR)

4.1 Imperceptibility Test

The suggested scheme’s imperceptibility performance is evaluated both subjectively and objectively. Watermarked images and the corresponding extracted EHR and hospital emblem are shown in Fig. 8. The cover medical images in Fig. 7 and the watermarked images in Fig. 8 are remarkably similar. The embedding watermark is unnoticeable. Furthermore, we objectively assess the imperceptibility of the proposed scheme utilizing the qualities PSNR and SSIM, as shown in Table 1. Table 1 shows that, for all grayscale and color images, the suggested scheme’s PSNR and SSIM are above ideal values, i.e., 37dB and 1, respectively. Therefore, in objective and subjective studies, the proposed scheme is highly imperceptible for different image modalities.

4.2 Robustness Test

One of the primary goals of the smart healthcare environment is the reliable transmission of medical images. Table 1 displays the recommended scheme’s robustness performance in terms of NC and BER metrics under zero attacks. In Table 1, NC and BER for all grayscale images are = ideal values 1, 0 respectively. For all color

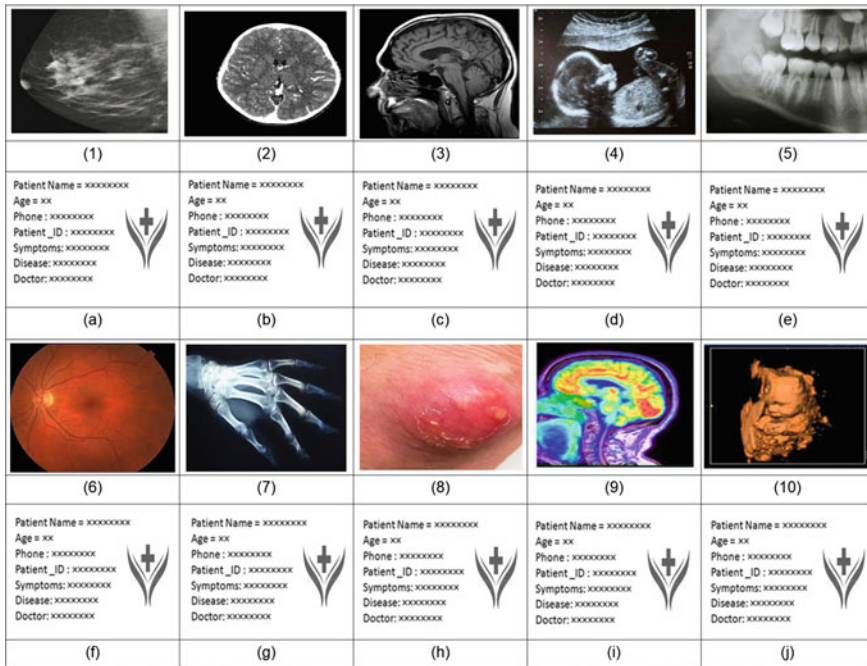


Fig. 8 Grayscale watermarked images (1–5), color watermarked images (6–10), and corresponding extracted EHR, hospital emblem (a–j)

Table 1 PSNR, SSIM, NC, and BER (under zero attacks) of grayscale, color images (NCOR-NC between original and recovered images)

Grayscale image	PSNR	SSIM	NC	BER	NCOR	Color image	PSNR	SSIM	NCOR	BER	NCOR
Mammograph	40.21	0.9785	1	0	1	Retina	42.91	0.9952	0.9999	0.0002	1
CT Brain	39.97	0.9631	1	0	1	MRI arm	41.63	0.9919	0.9984	0.0083	1
MRI Brain	41.62	0.9989	1	0	1	Skin	43.10	0.9841	0.9991	0.0008	1
Ultrasound	41.28	0.9991	1	0	1	PET brain	40.61	0.9879	0.9986	0.0019	1
X-ray chest	39.85	0.9769	1	0	1	Doppler	41.22	0.9982	0.9992	0.0006	1

images, NC and BER are nearly close to ideal values. The embedded EHR and hospital emblem are successfully extracted in all image modalities under zero attacks, as can be seen in Figs. 7 and 8. Further to assess the performance of the proposed scheme under zero attacks, various image processing attacks (noising, filtering, and compression) are applied to watermarked images, and the resulting NC and BER are reported in Table 2. As shown in Table 2, the proposed scheme is highly resistant to median filter, gaussian filter, and butter worth filter attacks, and the corresponding NC and BER are equal to the ideal value. From this observation, the recommended scheme is highly resistant to the majority of filtering attacks. With salt and pepper,

Table 2 NC, BER for mammograph image under attacks (NCOR-NC between original and recovered images)

Attacks	NC	BER	NCOR	Attacks	NC	BER	NCOR
Gaussian noise (0.002)	0.6429	0.0429	0.9814	Butterworth filter	1	0	1
Salt & pepper (0.002)	0.9997	0.0004	0.9999	Gaussian filter 33	1	0	1
Poisson noise	0.8083	0.0402	0.9815	Translation	0.7391	0.0510	0.9842
Median filter 33	1	0	1	Rotation (25 ₀)	0.5127	0.0618	0.9617
Sharpening	0.8206	0.0291	0.9919	JPEG compression (50%)	0.9862	0.0391	0.9838

Poisson noise, sharpening, and JPEG compression, the proposed scheme exhibits good robustness, with NC and BER values that are close to ideal. For Gaussian noise, translation, and rotation attacks, the proposed scheme performs moderately; improvements in these attacks can be seen in future research. Moreover, the proposed scheme is a reversible watermarking scheme; to demonstrate the reversibility of the original image from the watermarked image after extracting the watermark, NC is calculated between the original and extracted cover medical images under zero, under attacks as shown in Tables 1 and 2. The NC between the original and recovered images is 1 in Table 1, implying that the proposed scheme proved successful in restoring the original cover medical image. Similarly, NC under attack is equal to or almost equal to the ideal value; therefore, the recommended approach can recover the cover medical image under attack as well.

4.3 Security Test

To analyze the watermark image’s security correlation test is performed between the original and encrypted images, original and decrypted images. The correlation coefficient (CC) between the original and encrypted images, as well as the original and decrypted images of different binary images, is recorded in Table 3. Horizontal, vertical, and diagonal directions are used to compute CC. Drawn from Table 3, it seems to be that, for all binary test images, the relative CC between original and decrypted images is equal to 1. This implies that the suggested encryption method decodes its original image from the encrypted images correctly. Furthermore, in all stated orientations, the CC between the original and encrypted images is nearly equal to zero, implying that the proposed scheme generates a secure cipher image. It could be inferred from this discussion that the proposed random key for encryption is a highly secure key.

Table 3 CC of different binary images (H-Horizontal, D-Diagonal, V-Vertical)

Test Image	CC between original and encrypted image			CC between original and decrypted image		
	H	D	V	H	D	V
EHR	0.0082	0.0074	0.0112	1	1	1
Hospital emblem	0.0092	0.0091	0.082	1	1	1
CT brain	0.0081	0.0117	0.0271	1	1	1
X-ray chest	0.0161	0.0185	0.0118	1	1	1

Table 4 Computational time for different medical images (ET-Embedding time, Ext-Extraction time)

Grayscale image	ET (s)	ExT (s)	Color image	ET (s)	ExT (s)
Mammograph	0.3018	0.2981	Retina	0.4182	0.4128
CT brain	0.3829	0.3781	MRI arm	0.3981	0.3941
MRI brain	0.3165	0.3073	Skin	0.4029	0.4002
Ultrasound	0.2976	0.2917	PET brain	0.4171	0.4126
X-ray chest	0.3176	0.3128	Doppler	0.3819	0.3803

4.4 Computational Cost

The ability to communicate data in real time is one of the most crucial aspects of the smart healthcare application. The suggested scheme's performance is examined on different image modalities with the i5 process and the computational time is reported in seconds as shown in Table 4. Table 4 shows that the embedding and extraction time for grayscale images is less than 0.39 s and less than 0.42 s for color images. Therefore, the suggested approach can be shown to be suitable for real-time applications.

5 Conclusions

In this paper, a region-based reversible MIW technique for secure medical image transmission in smart healthcare environments such as IoMT, Telemedicine, and Telehealth is suggested. The proposed scheme is tested on a variety of grayscale and color images of various modalities, and the results obtained are effective in maintaining the watermark's imperceptibility, robustness, confidentiality, and security at low

computational cost. The suggested scheme for geometrical attacks such as rotation, resizing, and scaling assaults is minimal robustness. In the future, we'd like to be able to efficiently handle geometrical attacks.

References

1. Elhoseny, M., et al.: Secure medical data transmission model for IoT-based healthcare systems. In: IEEE Access vol. 6, pp. 20596–20608 (2018)
2. Anand, A., Singh, A.K.: watermarking techniques for medical data authentication: a survey. *Multimedia Tools Appl.* **80**(20), 30165–30197 (2021)
3. Giri, K.J., et al.: Survey on reversible watermarking techniques for medical images. In: *Multimedia Security*. Springer, Singapore, pp. 177–198 (2021)
4. Alshanbari, H.S.: Medical image watermarking for ownership and tamper detection. *Multimedia Tools Appl.* **80**(11), 16549–16564 (2021)
5. Nazari, M., Maneshi, A.: Chaotic reversible watermarking method based on iwt with tamper detection for transferring electronic health record. *Sec. Commun, Netw* (2021)
6. Thakur, S., et al.: Chaotic based secure watermarking approach for medical images. *Multimedia Tools Appl.* **79**(7), 4263–4276 (2020)
7. Aparna, P., Kishore, P.V.V.: An efficient medical image watermarking technique in E-healthcare application using hybridization of compression and cryptography algorithm. *J. Intell. Syst.* **27**(1), 115–133 (2018)
8. Kaw, J.A., et al.: A reversible and secure patient information hiding system for IoT driven e-health. *Int. J. Inf. Manag.* **45**, 262–275 (2019)
9. Dugelay, J.-L., Roche, S.: Process for marking a multimedia document, such an image, by generating a mark, Pending patent EP 99480075.3, EURECOM 11/12 EP (1999)
10. Rey, C., Dugelay, J.-L.: Blind detection of malicious alterations on still images using robust watermarks. In: *Secure Images and Image Authentication Colloquium*. IEE Electronics and Communications, London (2000)
11. Sangeetha, M., Betty, P., Nanda Kumar, G.S.: A biometric iris image compression using LZW and hybrid LZW coding algorithm. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS). IEEE (2017)
12. Hassan, F.S., Gutub, A.: Efficient image reversible data hiding technique based on interpolation optimization. *Arabian J. Sci. Eng.* **46**(9), 8441–8456 (2021)
13. AL-Hashemy, R.H., Mehdi, S.A.: A new algorithm based on magic square and a novel chaotic system for image encryption. *J. Intell. Syst.* **29**(1), 1202–1215 (2020)
14. Arroyo, J.C.T., Dum Dumaya, C.E., Delima, A.J.P.: Polybius square in cryptography: a brief review of literature. *Int. J.* **9**(3) (2020)
15. <https://www.oasis-brains.org>

Advancements in Reversible Data Hiding Techniques and Its Applications in Healthcare Sector



Buggaveeti Padmaja, Maharana Suraj, and V. M. Manikandan

Abstract Among all the approaches, Digital watermarking is the most widely implemented approach for copyright protection and authentication of data. In this technique, a unique piece of information is known as a watermark. Then the watermark gets into an image, later, to achieve its objective the watermark will be extracted. For the transmission of medical images, digital watermarking schemes are mostly used to ensure that the image has not gone through any unauthorized or illegal modifications during the transmission. Since conventional watermarking schemes alter the pixels in the original image, it is not suited for watermarking medical images. In medical images, permanent modifications may adversely affect the diagnosis process at the receiver side, caused by watermarking, especially when we are using some computer-aided diagnosis tools. This motivated computer scientists to work on reversible watermarking schemes. The reversible watermarking technology makes it possible to recover the required medical image from the watermarked image, while extracting the hidden watermark. So, the reversible watermarking technique does not affect the diagnosis in any way since the recovered image will be equivalent to the original image. This recovered image will be used by the user. The use of reversible watermarking techniques to send patient reports along with medical images is also explored, with the patient reports being embedded in the medical picture itself rather than the watermark. These techniques are commonly known as reversible data hiding techniques. This book chapter gives a brief overview of reversible data hiding techniques, reversible watermarking methods, and the major applications in medical image transmissions. In addition, the chapter addresses contemporary reversible data hiding and reversible watermarking algorithms intended specifically for medical picture transmission. The

B. Padmaja · M. Suraj · V. M. Manikandan (✉)
Department of Computer Science and Engineering, SRM University-AP, Amaravati, Andhra Pradesh, India
e-mail: manikandan.v@srmmap.edu.in

B. Padmaja
e-mail: padmaja_buggaveeti@srmmap.edu.in

M. Suraj
e-mail: maharana_suraj@srmmap.edu.in

chapter also discusses some of the obstacles that must be overcome when developing a reversible watermarking system for healthcare applications.

Keywords Data security · Medical image security · Data hiding · Watermarking · Reversible watermarking · Clinical data transmission · Health care

1 Introduction

Information security is one of the emerging domains in the communication system. It is a set of techniques and methodologies which are designed to ensure the security of electronic, confidential, and private data. It prevents breaches of data. It helps the healthcare industry to keep health records safe from unauthorized use. In the healthcare industry, data is confidential and sensitive as it carries medical information and drug-related data. Nowadays, those data are stored in electronic form. Information security approaches can be applied to this information to ensure its integrity, facilitating secure connection among healthcare providers, and improvising safety of medication, reporting, and tracking. The benefits of health information technology include improved access to and compliance with guidelines, as well as improved healthcare quality [1].

Between 2009 and 2020, around 3,709 data of health care reported breach of 400 or more records. Theft, loss, exposure, or unauthorized disclosures of 268,190,493 hospital records resulted because of those breaches. In the year 2020, on average 1.75 breaches took place per day. One cannot risk losing the clinical data of a patient in the process of transferring it from one source to another. It may cause major issues like loss of that information and can lead to inaccurate treatment and many such losses.

For healthcare organizations, one of the most useful data protection methods is encryption. Encrypting the data in transit and at rest, by healthcare authorities and businessmen, makes it impossible for attackers to decode the patient's information even if they manage to gain access to any sensitive data. Data hiding is an important method of ensuring secure communication. It is a technique where the sender embeds private information in a medium like an image, video, watermark, etc. When the data hiding techniques are applied to medical information, we must ensure that there is no loss of data when the receiver extracts it. One cannot risk losing the clinical data of a patient in the process of transferring it from one source to another. It may cause major issues like loss of that information and can lead to inaccurate treatment and many such losses. To avoid these, reversible data hiding (RDH) techniques come into play [2–6]. This process of securely sending data to the receiver from the sender with zero loss of information and distortion in the image is called RDH. These techniques came into existence in the year 2002.

Digital watermarking is a method where confidential information is embedded into an image [7]. A watermark is a logo, text, or pattern that is almost transparent in an image. Digital watermarking schemes are used in many applications like copyright

protection, fingerprinting and digital signatures, data authentication, protection, etc. The objective of this process is to save the integrity of information. Medical data of a patient needs to be taken care of when it is transferred between two sources. Watermarking is one of those methods which ensures the security of data. So, this method is also used in the transfer of clinical data.

There are two things to be considered when the digital watermarking technique is applied:

- There should be no loss of information.
- Original image must be restored perfectly without any distortion in pixels.

The process of perfectly restoring the real image after extraction of the watermark information is said to be reversible digital watermarking. When the technique is applied to medical information, the user must ensure that it is reversible.

In the further sections of the chapter, we will discuss various methods for secure data transmission, data hiding schemes in detail, the eversible data hiding approaches, reversible watermarking, some related work, and the challenges in this area.

2 Methods of Secure Communication

One of the two important strategies for secure communication is cryptography. Cryptography's purpose is to render data unreadable to a third party. Network security protocols are separated into symmetric (secret-key) and asymmetric (public-key) cryptography techniques. By employing the same key, symmetric algorithms are utilized to cipher and decrypt original messages. Asymmetric algorithms, on the other hand, use a public-key cryptosystem to exchange keys and then apply quicker secret key algorithms to assure stream data secrecy. A pair of keys is used in public-key encryption techniques, one of which is known to the public and is used to encrypt data to be transferred to a receiver who has the corresponding private key. Both private and public keys are distinct and require a key exchange. Encryption is one of the efficient ways to provide confidentiality to the content by changing them into an unreadable form. Encryption of medical images and healthcare data before transmission is very common in the healthcare industry which will help secure the data from unauthorized access (Fig. 1).

The data is disguised in a cover file and transferred across the network which is known as Data Hiding. The fundamental benefit of data concealing techniques over encryption is that they can hide the existence of secret information. On the other hand, one can easily recognize an encrypted data file by seeing the content even though they cannot infer anything from that. Data hiding is a useful approach to safeguard data since it allows you to conceal data in a host (cover) without losing its value. Data hiding methods may be used to hide secret messages in photographs in an undetectable fashion, such that the original image and the image with embedded data appear to be the same. Every difference between the tagged and original image is considered noise, making it harder for a third party to decode the encoded data.

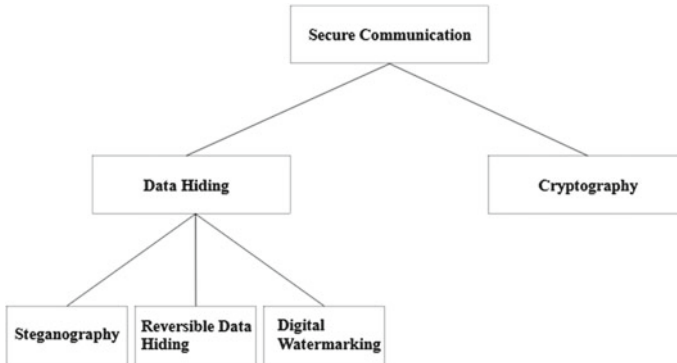


Fig. 1 Classification of secure communication techniques

Existing data concealing methods, on the other hand, have a relatively limited embedding capacity (the number of bits that can hide in an image). As a consequence, the embedded message is relatively brief (very). To increase picture embedding capacity while keeping the peak signal-to-noise ratio (PSNR) below allowable levels, several data concealment strategies have been applied. To assess the image's visual quality, PSNR is used. There are 2 types of data hiding methods: RDH and non-RDH. In RDH approaches, the actual cover medium can be restored while extracting the hidden details. In non-RDH methods, the original cover medium is irrevocably corrupted and cannot be recovered later. Military and intelligence communication, private communication, and protecting civilian speeches from attackers all use RDH algorithms. The RDH schemes are very popular in medical image transmission to hide patient reports in the medical image. This provides a way to transmit the medical image text files as a single entity instead of sending them as two different entities.

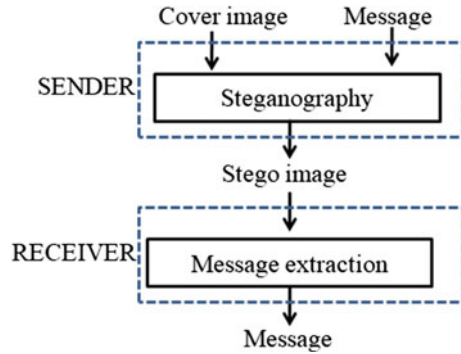
Data hiding techniques include steganography, digital watermarking, and reversible data hiding (RDH). These topics are detailed in subsequent subsections.

2.1 *Steganography*

Steganography is a secret communication technique [8, 9]. The technique of concealing information is referred to as steganography. Any type of digital file, particularly image files, is most commonly used to conceal data. The host files (which contain hidden information) can then be transferred via an unsecured channel without anyone knowing what's inside. The existence of the message is known in steganography but in cryptography the meaning is unknown. Information is hidden using steganography software. The overview of an image steganography scheme is shown in Fig. 2.

By considering the nature of the cover item, steganography is divided into five types: audio steganography, image steganography, network steganography, text

Fig. 2 Overview of image steganography scheme



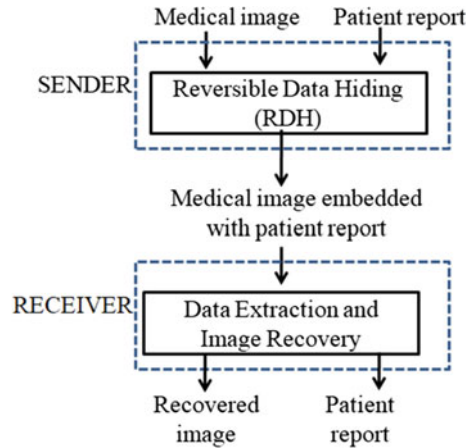
steganography, and video steganography. We will go through the various forms of steganography briefly. The practice of concealing information inside text files is called text steganography. Image steganography is the process of concealing information by using an image as the cover object. In audio steganography, the confidential information is encrypted in an audio signal that alters the binary sequence of the corresponding audio file. When compared to other methods, hiding hidden messages in digital sound is far more challenging. You may use Video Steganography to hide any sort of data in a digital video. This kind has the benefit of being able to store a vast quantity of data and being a moving stream of pictures and sounds. The process of embedding information in data transmission network control protocols, such as Internet control message protocol (ICMP), transmission control protocol (TCP), and user datagram protocol (UDP), is known as network steganography. In several covert channels included in the open systems interconnection (OSI) model, steganography can be used. All of these technologies appear to be of tremendous assistance, but criminals and terrorist organizations are taking advantage of them. Understanding how to utilize steganography to obscure data and prevent it from being abused may be incredibly beneficial in both attack and defensive scenarios.

2.2 Reversible Data Hiding (RDH)

The RDH enables you to embed a huge quantity of data inside an image, allowing you to extract the hidden data while recovering the original image [2–6]. This makes it a good choice for instances where metadata has to be preserved in the cover signal, but the original signal needs to be retrieved without a loss following data extraction. These schemes are used for many applications. Some of the important applications are

- For authentication purposes, include authentication data in medical photographs or other very sensitive images.

Fig. 3 Overview of RDH scheme in medical image transmission



- Before digital material can be uploaded to a cloud service provider, metadata must be incorporated into it.
- For the purpose of concealing some inquiry details in forensic pictures.
- Embedding and transferring clinical data into a medical image.

The overview of an RDH scheme for patient transmission along with the medical image is illustrated in Fig. 3.

2.3 Digital Watermarking

The copyright of digital files can be protected via digital watermarking techniques. To protect digital assets such as music, photos, and formal documents, a number of watermarking systems have been proposed. Digital watermarks can be visible, and they might be in the shape of logos or pictures in the corner of the digital content [7]. A number of invisible watermarking schemes are also there. Digital watermarking is frequently used in E-commerce to provide conditioned and consumer access to specified resources. As a result, digital watermarking allows artists to use the Internet to reach a larger audience for their work. Over the past few years, there have been many watermarking techniques that were introduced. The classification is shown in Fig. 4.

Techniques for watermarking may be classified into two groups: human perception-based and domain-based. The watermark information in digital watermarking might be fragile, semi-fragile, robust, or hybrid. Watermarks that are fragile are used to detect tampering, whereas robust watermarks are used to withstand standard image processing procedures. Semi-fragile watermarks are strong against friendly attacks but fragile against malicious attacks, while hybrid watermarks have mixed features.

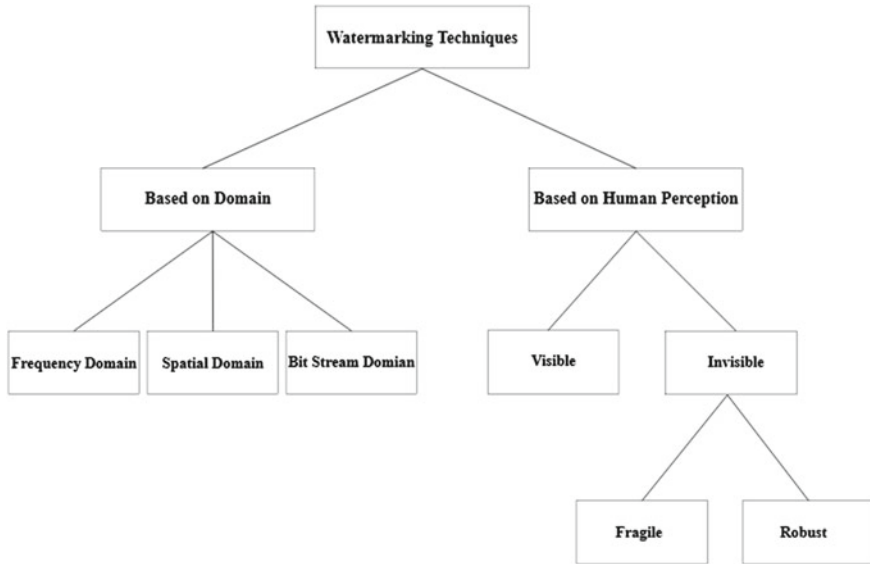


Fig. 4 Classification of watermarking Techniques

Fragile watermarking schemes are widely used for data authentication during medical image transmission. The sender and receiver will agree with a watermark at the beginning itself. When the sender wants to send some medical images from one place to another, the watermark will be embedded in the medical image. At the receiver side, the receiver will attempt to extract the embedded watermark using the watermark extraction procedure. Further, the extracted watermark will be compared with the original watermark. If these two are highly correlated then it indicates less alterations in the content during transmission. The changes in the content may be due to loss of information during transmission or due to an attack by an unauthorized person.

The conventional digital watermarking methods modify the original content permanently while embedding the watermark. This is a major concern when using digital watermarking techniques in medical image transmission since the changes in the medical images during watermarking may lead to a wrong diagnosis on the receiver side. This motivated the researchers to work on an area called reversible watermarking in which the receiver can extract the watermark and also the recovery of the original image is possible at the receiver side.

Because of its increasing applications in many different fields, in recent years, reversible watermarking has received a lot of interest. Reversible watermarking techniques were developed to be used primarily in situations where the validity of a digital image must be guaranteed, and the original content must be decoded. Patients’ privacy must be secured due to the ever-increasing volume of medical digital photos and the necessity to send them between experts and hospitals for more accurate and better treatment. On the concept of ensuring the visual quality of the image, this method

embeds the information into a carrier image. The goal is to recreate the host image without loss after the watermark has been extracted. As a result, the amount of embedding information is more demanding than typical watermarking methods. Hence, it has more broad study and application value in the sectors of legal, military, medical, and other fields that demand high-image authenticity and integrity. The research on reversible image watermarking techniques aims to produce the highest embedding capacity of effective information with the least amount of distortion.

The next section of the book chapter introduces you to the efforts and research that researchers have done in recent years in the area of reversible watermarking.

3 Related Work

This section provides an overview of recent related research on reversible watermarking and reversible data hiding in medical image transmission, both of which have received widespread recognition in the scientific world. Before discussing the related works, the efficiency parameters used to analyze reversible data hiding schemes or reversible watermarking schemes are listed below. This discussion will help the readers to understand the further discussions in a better way.

3.1 Efficiency Parameters

In the process of analyzing different works, we will come through the following efficiency parameters:

- **Bit Error Rate:** The quality of data extraction is checked using this efficiency measure. In general, if the bit error rate is 0, it signifies that the data encoded in the image has been correctly retrieved, meaning that no data bits have raised an error.
- **Embedding Rate:** The ratio between the greatest number of bits that can be contained in a picture and the complete number of bits in the image is used to compute it. Schemes having a high embedding rate are frequently used in order to embed the greatest number of bits.
- **Peak Signal-to-Noise Ratio (PSNR):** This parameter determines the visual quality of a picture. A high PSNR number indicates that the picture quality is rather excellent and that the mistake or corrupting noise that affects the image's visual quality is minimal. The PSNR of infinity indicates that the original and restored images are identical.
- **Structural Similarity Index (SSIM):** The SSIM values vary from 0 to 1, with 0 being the lowest and 1 being the highest. When the SSIM value is 1, the recovered image is a perfect match to the original. For any RDH scheme, it is suggested to make sure that the SSIM is 1.

- **Natural Image Quality Evaluator (NIQE):** It determines the image quality without a reference and compares the restored image to the original. The lesser the NIQE value, the better the perceptual quality.
- **Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE):** BRISQUE calculates the image quality without a reference and compares it to the original image, just like the NIQE parameter. Its value ranges from 0 to 100 in most cases. The lower the score, the better the image quality.

3.2 *Related Works on Reversible Data Hiding*

Reversible Data Hiding methods have been used on both natural and encrypted images. We will look at some of the more extensive efforts that have been done on natural images. As we've seen, it's critical to ensure that, regardless of the algorithm used, the original image is restored following data extraction.

A RDH scheme based on a prediction error histogram is discussed in [2]. The method worked by shifting pixels into black and white groups depending on prediction error histogram shifting. The pixels were classified using a checkerboard pattern. A black pixel will have four 4-neighbor pixels, as shown in the checkerboard pattern (top, right, bottom, and left). An average of three pixels from a four-neighborhood region that is extremely close to the center pixel value is used to estimate the value of the black pixel in the middle. The prediction error is then computed by comparing the value of the predicted pixel to the value of the actual pixel. The prediction error that correlates to all of the black pixels in the image is shown by the histogram of the prediction error. The histogram of prediction error is investigated for additional data concealment using the histogram shifting technique. Overflow and underflow, on the other hand, constitute a major issue in the RDH process. The results, on the other hand, show that image retrieval was successful and that image quality was improved. The embedding rate is high according to the approach, implying that the number of bits that may be embedded is similarly considerable. The BER is also included, resulting in a zero error rate. The images used in the approach were taken from the USC-SIPI image collection. The theoretical temporal complexity of the preceding approach was $O(N)$, which may be decreased for better results.

A RDH strategy on the basis of a histogram of the blocks of the host image is discussed in [10]. To ensure the accurate recovery of the original picture, each block's maximum intensity value is utilized for data hiding, and the grayscale value is used for data concealment by altering the least significant bits of eight selected pixels; each block is embedded in the same block. To hide the secret data in a cover image I , according to this approach, the I should be divided into sections. Every piece will be $B \times B$ pixels in size. Only if the peak intensity value's height is greater than $(F + 8)$, in which F is the number of 254 s and 255 s in that block, and one block will be used for data concealing. Whether a given block is suitable for data concealing or not is determined in the first phase, and a status sequence for embedding, M , is generated. If M_i is 0, it means that the i -th block is not utilized for data concealing,

and if M_i is 1, it means that the i -th block is. It's worth noting that the blocks are accessed and/or numbered in row-by-row linear order, ignoring the initial block. By altering the LSBs of the initial (N_B-1) pixels, the details about the embedding status will be buried in the first block of the picture, where N_B is the total number of blocks in the image. Because the original picture must be retrieved before the LSBs may be changed, the LSBs are encoded in the image itself, together with a coded message. Overflow pixels (initial pixels with pixel value 255) and pixels taken 255 after histogram shifting are differentiated with help of marker bits. The pixel with marker information 0 was originally 254, whereas the pixel with marker information 1 was originally 255. All of these details will be presented in a single image block. At the receiver's end, the overhead information may be retrieved, and this data will allow the recipient to recover the actual image to its original state. The visual quality of the recovered image was assessed using the SSIM and PSNR efficiency criteria. The PSNR is high, and the SSIM is near to but not quite one. This may be improved so that the system can be utilized in everyday situations.

The introduction of an effective overflow handling solution based on histogram shifting is discussed in [4]. Images from the USI-SIPI dataset were used to test the approach. Pixel values in a grayscale picture generally range from 0 to 255. As a result, histogram shifting converts a pixel value of 255 to 256, which is theoretically impossible. The term "overflow" is used to characterize this condition. To deal with the overflow, this technique was implemented. In this procedure, the histogram of the original picture I is displayed first. The peak of the histogram is located, and P_k is assigned to it. Pixel values greater than the histogram's peak are displaced, resulting in a gap in the histogram immediately following the peak. For the data hiding approach, all of the pixels are available in a predetermined sequence. If the pixel value is P_k , the data embedding method is performed on that pixel. To embed a message in a pixel, bit b , with P_k intensity value, b adds P_k . As a result, the altered image seems to be similar to the original. This picture contains fragments of a hidden message. The name of this image is Stego Image. Each pixel is examined after the data concealing method, and if its value exceeds the histogram's peak by b , those pixels are decremented by b , and message bit b is extracted. If the value of the pixel is the same as the actual value, the value is set to 0 and the pixel remains unchanged. As a consequence, the original pixel values have been restored, and the Stego image's message has been retrieved. Although two images produced a high embedding rate, which is laudable, this method asserts that the embedding rate is only determined by the histogram's peak. This approach may fail if the peak value is 255 since we can't increase it by one, resulting in a pixel value of 256.

In [6], a separable reversible data hiding mechanism is found. The suggested method is divided into three phases: data embedding, picture encryption, and data extraction/image recovery. To create an encrypted image, the content owner encodes the original uncompressed image with an encryption key. The data-hider then uses a data hiding key to compress the encrypted image's least important bits, resulting in a thin region that may take the extra data. The data encoded in the generated space may be simply retrieved at the receiver's end using the data hiding key from the

encrypted picture containing extra data. As the data embedding modifies only the LSB, the decryption with the help of the encryption key might result in a picture that appears exactly like the original. When both the data hiding keys and encryption are employed, the additional embedded data can be effectively recovered, and the actual picture then can be flawlessly reproduced by making use of the spatial correlation found in natural images.

As a result of the foregoing approach, the PSNR of a natively decrypted picture has been observed to be high but not infinite. The recovered picture has a PSNR of infinity, suggesting that it has been fully restored. However, the embedding rate for this technology may have been higher in order for it to be more useful in a wider range of applications.

Following a review of important publications on RDH, we may conclude that a few considerations must be made:

- To guarantee that the maximum amount of message bits may be embedded, the embedding rate must be as high as feasible.
- The image's visual quality must be excellent, which implies the PSNR and SSIM must be acceptable at the end.
- The bit error rate should be 0 at all times. Because the data is precisely extracted, the algorithm becomes more efficient.

3.3 Related Works on Reversible Watermarking

Before looking at the related works, let us understand the basic and important terms that we will come across in the watermarking techniques for a better understanding.

In general, the regions in medical images are divided into two sections: ROI and RONI. ROI stands for the region of interest and RONI stands for the region of non-interest. The ROI area is a diagnostically important part of the image that must be kept as much as feasible. The non-critical component of the picture is included in RONI like the background. When the ROI component is used to hide the watermark, the pixel intensities in this region may be deformed, resulting in misperceptions and, as a result, misdiagnosis. Watermarking techniques developed by RONI incorporate data in areas that are deemed insignificant in medical examinations. However, this can only be done if RONI is installed. The quantity of data that can be implanted is greatly dependent on the RONI size. The algorithm used by the researcher determines if the picture should be divided into ROI and RONI. Only a few methods don't need picture segmentation, and only a few do.

A novel robust reversible watermarking scheme for protecting medical image authenticity and integrity is presented in [11]. Robustness refers to a watermarking algorithm's capacity to survive operations like compression, filtering, and geometric assaults like rotation, scaling, and translation. A verification method for medical photos was included in this technique. It is broken down into four stages, as follows:

- Watermark Generation Phase: Watermarks, as well as integrated and authenticity data, are created during this phase.
- Phase of embedding watermark: Embedding of a watermark is done with the help of SLT at this phase (slantlet transform). The usage of single value decomposition (SVD) improves the watermarking robustness by making the value invariant to diverse attacks.
- Watermark extraction phase: At this phase, the watermark is removed, which is the opposite of the embedding process.
- Security verification phase: Integrity of the information and the ability to recover from tampering are checked in this phase.

The approach makes use of the SLT-SVD hybrid transform. In this design, the ROI and RONI are simply employed to create the watermark. The previous method reconstructs ROI and RONI pictures without loss using an RDM-based reversible function. By integrating the watermark into the full medical image by not separating ROI and RONI, the suggested watermarking technique solves the privacy issues associated with geographically partitioned picture partitioning, exceeding conventional ROI-lossless watermarking. The PSNR values of the recommended design are higher than those of the other schemes, and the SSIM was nearly one. However, because it is being used for medicinal purposes, the outcomes may be more favorable. The Bit Error Rate (BER), which is not zero, is accustomed to determining robustness.

A technique that uses the Integer Discrete Cosine Transform (InDICT) and Difference Expansion for reversible watermarking (DE) [12]. Frequency domain watermarking includes this notion. Using this method, the image is divided into non-overlapping chunks. The difference in expansion inserting and extraction process is applied to the separated blocks. In the transform processing stage, the integer-based DCT transform is applied to increase correctness and computation performance. For each block, the technique starts with an 8×8 image that is subjected to 2D-DCT. A zigzag scan is used to convert the watermark image into a 1D vector, which may also be done after the watermark image has been encrypted. The zigzag scan is used because it spreads out the neighboring pixels and strengthens the picture. To evaluate whether there is any overflow or underflow, a reconstructive stimulation method is performed on each block. According to the experimental results, PSNR is high and provides reversibility, which is vital for medical applications. The algorithm's embedding capacity is represented by the number of blocks with the proper energy. The amount of blocks varies depending on the images. It's possible that the rate of embedding was higher.

Another reversible watermarking scheme is discussed in [13]; it's been disclosed that there's a reversible invisible digital watermarking method for medical image ownership security and content authentication. This solution is designed in the geographical domain using meaningful data from patients as ownership data and a SHA256 hash function built using an adaptive prediction algorithm and additive predictor error technique. Hash functions are used for mapping. The hash function H for a hash value h getting a variable-sized input x can be written as $H = h(x)$. Hash functions are called collision-free when no messages x and y have the same

$H(y) = H(x)$. SHA256 is one of the widely adopted collision-free hash functions. It has also been adopted by NSIT. In the proposed RDH algorithm, the idea is to create 64 hexadecimal-size authentic watermark information for healthcare pictures, by using SHA 256 hash function. In this work, there has been a discussion about both, the extraction of watermark and medical images. In the receiver end, at the decoder, watermark from the clinical image and the watermark of ownership, both can be extracted. The use of the additive prediction error technique is done to extract both watermarks.

This algorithm has been implemented on MRI and X-ray images. Anything unique to the patient, such as a mobile number or an Aadhar number, might be used as ownership information. The ownership information is then incorporated 10 times in the image as a watermark. The content authentication watermark is created using the hash algorithm. The study goes into great detail on how PSNR varies when watermark hiding parameters vary.

A scheme for medical image watermarking with tamper detection and recovery using reversible watermarking with LSB Modification and run length encoding (RLE) compression is explored in [5]. This method uses the LSB modification to detect and recover tampering in the ROI. The ROI and RONI of a photograph are first separated. To detect and recover tampering, watermarking would be completed in ROI. RONI will be put to use to integrate the image's actual LSBs, making the watermark reversible. Only the LSBs from the ROI are used, rather than all of the LSBs in the image. Before adding the watermark, the picture's original LSBs are eliminated, and the LSB of each pixel is set to zero. After that, the LSBs that were removed will be compressed and placed into the RONI. The original LSBs of the image will be compacted and saved in the RONI. The LSB is saved and can later be used to restore the image's pixel values to their original state. A watermarked image's PSNR is around 46 dB, implying that the watermarking method may create a watermarked image with the least amount of deterioration and that appears extremely same as the initial one. During the trial of watermark reversibility, the actual LSBs of the watermarked picture are restored with no modification. The compacted LSBs from the RONI would be retrieved, decrypted, and recovered to every pixel. This procedure's outcome, the recovered picture, is then compared to the original image. While the recovered image is nearly similar to the initial one, the PSNR is in the range of 56 and 61 dB, showing that there is still a minor difference. The PSNR might be upgraded to make it more useful in medicine.

In [14], in CT SCAN images, a method for reversible watermarking was discovered. This process employs the ROI and RONI watermarking techniques. With the help of a segmentation algorithm, the procedure segments the actual image into ROI and RONI. It embeds a fragile watermark in the ROI to maintain the fairness of the captured image, while a compound robust watermark is implanted in the RONI to safeguard client data and copyright of medical pictures. The method uses prediction-based reversible watermarking to integrate ROI repair data into RONI. Following the embedding step's segmentation of ROI and RONI, the LSBs of ROI are segregated and preserved in an isolated storage. Watermarks are made that are both delicate and durable. The LSBs of ROI are replaced with delicate watermarks to

create watermarked ROI. This process is repeated for each RONI pixel until all of the watermark data is implanted. The watermarked picture is now segmented into ROI and RONI with the help of the segmentation procedure. Watermarked ROI's LSBs are extracted. The approach determines that the restored picture has not been altered in any manner. The PSNR values were good, and the technology permits CT scan images to be transferred from one source to another without fear of losing medical information.

A few works and its characteristics are summarized in Table 1.

Following the analysis of various situations, it can be stated that the majority of works were primarily focused on the following:

Table 1 A few schemes for reversible data hiding/reversible watermarking

Reference No.	Secure communication technique	Algorithm used	Quality measures
[2]	Reversible Data Hiding	Prediction error-based histogram shifting	0.0948bpp PSNR—Infinity SSIM—1
[6]	Reversible Data Hiding	Separable Reversible Data Hiding in Encrypted Image	ER—0.017 bpp PSNR—Infinity SSIM—1
[15]	Reversible Data Hiding	High capacity using histogram shifting of each image between minimum and maximum frequency	0.045–0.14 bpp PSNR—49–53 dB
[10]	Reversible Data Hiding	Block-wise histogram shifting	ER-0.058–0.11bpp PSNR—50–52 dB SSIM—0.99
[16]	Reversible Watermarking	Difference Expansion Method	BER—0.39–0.49 PSNR—Infinity RMSE—0
[17]	Reversible Data Hiding	Interpolation-based Reversible Data Hiding (IRDH)	PSNR—39–47 dB SSIM—0.85–0.92
[5]	Reversible Watermarking	LSB modification and Run Length Coding Compression	PSNR of watermarked image—46 dB PSNR of extracted image—56–61 dB
[18]	Reversible Watermarking	Modulation mode of Discrete Cosine Transform coefficients	PSNR—70 dB SSIM—1
[19]	Reversible Watermarking	Difference expansion technique	PSNR—Infinity SSIM—1 RMSE—0
[3]	Reversible Watermarking	Recursive Dither Modulation (RDM) technique	PSNR—45–66 dB SSIM—0.97–0.99

- Reversible data hiding schemes are more popular for clinical data transmission along with medical images.
- Reversible watermarking is an extension of reversible data hiding scheme in which a watermark will be embedded in the image using a reversible data hiding technique. The watermark can be extracted at the receiver side for ensuring data authenticity.
- The quality of the watermarked image and embedding rate is a major concern
- The image recovery process is expected to be lossless at the receiver side.
- Robust reversible data hiding schemes are the recent advancements in this domain in which the image can be recovered at the receiver side even though some distortions happened in the image during transmission.

4 Medical Image Datasets for the Research Work

In this part, we'll go through the key datasets that were considered for various methodologies.

- Natural photos were acquired from a dataset named USC-SIPI, which is supervised by the University of Southern California [20]. It's a collection of photographs that are mostly kept for the purposes of image processing, image analysis, and machine learning. There are four volumes in this database, namely
 - Textures
 - Aerials
 - Miscellaneous
 - Sequences.
- Several strategies were also carried out using DICOM image collections. These datasets are only available for use in research and education. Digital Imaging and Communications in Medicine (DICOM) is a standard for the sharing and administration of medical imaging data and related information. DICOM files can be shared between two organizations that can receive DICOM-formatted pictures and patient data. A set of DICOM images are available in OsriX DICOM image set [21].
- Open Access Series of Imaging Studies (OASIS) is an initiative whose goal is to make neuroimaging datasets publicly available for medical study. The most recent release is OASIS-3, which is a longitudinal neuroimaging, clinical, cognitive, and biomarker dataset for normal aging and Alzheimer's disease [22].
- National Biomedical Imaging Archive (NBIA) is a repository that provides access to imaging resources that will increase the use of imaging in biomedical research and practice today by improving the efficiency and repeatability of cancer detection and diagnosis using imaging. Imaging can be used to offer an objective assessment of a patient's reaction to treatment. Ultimately, this will enable the development of imaging resources that will improve clinical decision support [23].

5 Research Challenges

In this part, we'll look at some of the research problems that came up throughout the implementation phase. Every algorithm or scheme, when it is being implemented for practical applications, must make sure that the following points are considered:

- When the technique is used for medical purposes, one must ensure that the medical image is perfectly restored on the receiver side. This can be verified using a parameter PSNR. Peak Signal-to-Noise Ratio is shortly called PSNR.

The formula of PSNR is $10 \log_{10} ((PEAK)^2/MSE)$ dB.

PEAK denotes the image's highest pixel intensity, while MSE denotes the noise-measuring metric known as mean square error. When the mean square error is 0, it means that the image is perfectly restored without any distortion and thus PSNR results to be infinity.

- Similar to the PSNR, Structural Similarity Index (SSIM) is also considered to verify the image reversibility. To understand that image retrieval is perfect, SSIM should be 1. It generally ranges from 0 to 1.
- The data that is embedded in the images are confidential. So, the factor Bit Error Rate is taken into account to check the accuracy of the message retrieved from the picture at the receiver end. When the Bit Error rate is 0, it suggests that there is no message bit that has raised an error during the extraction process.
- Embedding rate should be higher, which means that the amount of data users can embed should be high for better usage. It is computed in Bits per Pixel (bpp).

Along with reversible data hiding, reversible watermarking techniques are also commonly utilized. There are a few obstacles to overcome throughout the implementation of this process:

- It's critical to have a strong embedding capability that can allow embedding the necessary legitimate data.
- Even after inserting the watermark or information, the quality of the tagged photos must be maintained. The imperceptibility factor is reflected in the quantity of invisibility of the watermark.
- The picture's durability should be considered. Misdiagnosis might occur if the picture is altered during transmission.
- The receiver must have the flexibility to restore the original image without distortion after removing the encoded watermark.

The elements listed above are frequent research problems that every researcher encounters throughout implementation. Only if the aforesaid variables are taken into account will a scheme be entirely acceptable for all purposes.

6 Conclusion

In this book chapter, we have discussed the importance of secure communication, especially during medical image transmission. A detailed discussion is done on various ways to transmit data securely through an unsecured channel. This chapter also discussed the motivation to work in the domain of reversible data hiding. The extension of reversible data hiding called reversible watermarking schemes is also discussed in detail in this chapter. Various well-known reversible data hiding schemes are briefed along with their merits and demerits. The well-known datasets available for the study are also summarized in the chapter. The researchers working in this domain can use those datasets for the experimental study. A number of research challenges are also discussed in this chapter so that the researchers can focus on these areas to design and implement new reversible data hiding schemes suited for the healthcare sector.

References

1. Box, D., Pottas, D.: Improving information security behaviour in the healthcare context. *Procedia Technol.* **9**, 1093–1103 (2013)
2. Padmaja, B., Manikandan, V.M.: A novel prediction error histogram shifting-based reversible data hiding scheme for medical image transmission. 2021 4th International Conference on Security and Privacy (ISEA-ISAP), pp. 1–6 (2021). <https://doi.org/10.1109/ISEA-ISAP54304.2021.9688572>
3. Ribina, B., Jeena, R.S.: Recursive dither modulation based reversible watermarking scheme for medical images. 2015 International Conference on Control Communication & Computing India (ICCC), pp. 375–379 (2015). <https://doi.org/10.1109/ICCC.2015.7432924>
4. Manikandan, V.M., Renjith, P.: An efficient overflow handling technique for histogram shifting based reversible data hiding. 2020 International Conference on Innovative Trends in Information Technology (ICITIT), pp. 1–6 (2020). <https://doi.org/10.1109/ICITIT49094.2020.9071553>
5. Tjokorda Agung, B.W., Adiwijaya, Permana, F.P.: Medical image watermarking with tamper detection and recovery using reversible watermarking with LSB modification and run length encoding (RLE) compression. 2012 IEEE International Conference on Communication, Networks and Satellite (ComNetSat), pp. 167–171 (2012). <https://doi.org/10.1109/ComNetSat.2012.6380799>
6. Zhang, X.: Separable reversible data hiding in encrypted image. *IEEE Trans. Inf. Forensics Secur.* **7**(2), 826–832 (2012). <https://doi.org/10.1109/TIFS.2011.2176120>
7. Katzenbeisser, S., Petitcolas, F.A.P.: *Digital watermarking*. Artech House, London 2 (2000)
8. Kahn, D.: *The history of steganography*. International Workshop on Information Hiding. Springer, Berlin, Heidelberg (1996)
9. Cheddad, A. et al.: Digital image steganography: Survey and analysis of current methods. *Signal Proc.* **90**(3), 727–752 (2010)
10. Murthy, K.S.R., Manikandan, V.M.: A block-wise histogram shifting based reversible data hiding scheme with overflow handling. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6 (2020). <https://doi.org/10.1109/ICCCNT49239.2020.9225552>

11. Liu, X. et al.: A novel robust reversible watermarking scheme for protecting authenticity and integrity of medical images. *IEEE Access* **7**, 76580–76598 (2019). <https://doi.org/10.1109/ACCESS.2019.2921894>
12. Gao, L., Gao, T., Sheng, G., Cao, Y., Fan, L.: A new reversible watermarking scheme based on Integer DCT for medical images. *2012 International Conference on Wavelet Analysis and Pattern Recognition*, pp. 33–37 (2012). <https://doi.org/10.1109/ICWAPR.2012.6294751>
13. Kunhu, A., Al-Ahmad, H., Mansoori, S.A.: A reversible watermarking scheme for ownership protection and authentication of medical Images. *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pp. 1–4 (2017). <https://doi.org/10.1109/ICECTA.2017.8251971>
14. Memon, N.A., Alzahrani, A.: Prediction-based reversible watermarking of CT scan images for content authentication and copyright protection. *IEEE Access* **8**, 75448–75462 (2020). <https://doi.org/10.1109/ACCESS.2020.2989175>
15. Fallahpour, M., Megias, D., Ghanbari, M.: High capacity, reversible data hiding in medical images. *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 4241–4244 (2009). <https://doi.org/10.1109/ICIP.2009.5413711>
16. Qasim, A.F., Meziane, F., Aspin, R.: A reversible and imperceptible watermarking scheme for MR images authentication. *2018 24th International Conference on Automation and Computing (ICAC)*, pp. 1–6 (2018). <https://doi.org/10.23919/ICAC.2018.8749000>
17. Wahed, M.A., Nyeem, H., Elahi, M.F.: An improved interpolation based reversible data hiding for medical images. *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–6 (2019). <https://doi.org/10.1109/ECACE.2019.8679278>
18. Bahrushin, A.P., et al.: *J. Phys. Conf. Ser.* **1399**, 033025 (2019)
19. Qasim, A.F., Aspin, R., Meziane, F., et al.: ROI-based reversible watermarking scheme for ensuring the integrity and authenticity of DICOM MR images. *Multimed Tools Appl* **78**, 16433–16463 (2019)
20. USC-SIPI by University of Southern California. <https://sipi.usc.edu/database/database.php?volume=rotate>
21. DICOM Image Library. <https://www.osirix-viewer.com/resources/dicom-image-library/>
22. OASIS Brains—Open Access Series of Imaging Studies. <https://www.oasis-brains.org/>
23. National Biomedical Imaging Archive—NBIA. <https://www.re3data.org/repository/r3d100012650>

Security Issues in Deep Learning



Shrina Patel, Parul V. Bakaraniya, Sushruta Mishra, and Priyanka Singh

Abstract Deep learning has created substantial improvements for industries and set the tempo for a destiny constructed on artificial intelligence (AI) technology. Nowadays, deep learning is turning into an increasing number of vital in our everyday lifestyles. The appearance of deep learning in many applications in life relates to prediction and classification such as self-driving, product recommendation, classified ads and healthcare. Therefore, if a deep learning model causes false predictions and misclassification, it may do notable harm. This is largely a critical difficulty inside the deep learning model. In addition, deep learning models use big quantities of facts inside the training/learning phases, which in corporate touchy facts can motivate misprediction on the way to compromise its integrity and efficiency. Therefore, while deep learning models are utilized in real-world programs, it's mile required to guard the privateness facts used inside the model. The countless opportunities and technological abilities that system learning has added to the arena have concurrently created new safety dangers that threaten development and organizational development. Understanding system learning safety dangers is one of every of our contemporary technological time's maximum vital undertakings due to the fact the results are extraordinarily high, mainly for industries along with healthcare in which lives are at the line. We talk about the forms of system mastering safety dangers that you may stumble upon so you may be higher organized to stand them head-on.

S. Patel (✉) · P. V. Bakaraniya
Department of Computer Engineering, Sardar Vallabhbhai Patel Institute of Technology, Vasad,
Gujarat, India
e-mail: shrinapatel.comp@svitvasad.ac.in

S. Mishra
School of Computer Engineering, KIIT(Deemed to be) University, Bhubaneswar, Odisha, India

P. Singh
Department of Computer Engineering, SRM University, Amravati, Andhra Pradesh 522508, India

1 Introduction

Deep learning has created substantial improvements for industries and set the tempo for a destiny constructed on artificial intelligence (AI) technology. Nowadays, deep learning is turning into an increasing number of vital in our every day lifestyles. The appearance of deep learning in many applications in life relates to prediction and classification such as self-driving, product recommendation, classified ads and healthcare. Therefore, if a deep learning model causes false predictions and misclassification, it may do notable harm. This is largely a critical difficulty inside the deep learning model. In addition, deep learning models use big quantities of facts inside the training/learning phases, which in corporate touchy facts can motivate misprediction on the way to compromise its integrity and efficiency. Therefore, while deep learning models are utilized in real-world programs, it's mile required to guard the privateness facts used inside the model. The countless opportunities and technological abilities that system learning has added to the arena have concurrently created new safety dangers that threaten development and organizational development. Understanding system learning safety dangers is one of every of our contemporary technological time's maximum vital undertakings due to the fact the results are extraordinarily high, mainly for industries along with healthcare in which lives are at the line. We talk about the forms of system mastering safety dangers that you may stumble upon so you may be higher organized to stand them head-on

1.1 Implementations of Deep Learning

An in-depth study added new ways of looking at technology, AI as well as its offshoots. Deep Learning has impacted the way we live and will continue to influence how we look forward to the future. DL holds the date of the marketplace with the help of using the date. There are various Deep Learning applications available, some of the popular applications are Photo Resetting, Face Reconstruction, Automatic Coloring, Photo Caption, Advertising, Earthquake Prediction, and the Discovery of brain cancer. In-depth learning also enhances each aspect of lifestyle with the help of problem-solving solutions and adding new dimensions to research. In-depth learning's remarkable performance is within the reach of current security mechanisms. Every small and major institution faces a significant issue today; hundreds of thousands of computer viruses and security threats are being generated, and large groups like banks and governments are being targeted. However there are many security solutions, and security is an existing research topic. In-depth learning has provided new features within the cybersecurity environment with the help of network detection, eliminating malware, detection, and system security.

2 Background

2.1 Deep Learning

The in-depth study allows over-the-counter computer models that combine more than one layer of processing to test the presentation of information beyond the range of invisible layers. These procedures have enormously worked in the domain of voice acknowledgment, visual acknowledgment, object disclosure, and various regions, as well as the improvement of illness medication and genomics. An inside and out investigation of fake sensor networks frequently integrates extra. An in-depth study of artificial sensor networks often incorporates additional professional model parameters compared to the sample scope in which they are trained. However, the number of those models shown significantly decreased circular error, i.e., the difference between training error and terrorist error. It's just clean to achieve standard systems with a smaller cycle. So separates ordinary neural networks correctly from those who now not do now? The top-notch way to this query will now not assist to make the neural networks greater descriptive but may cause greater dependable and dependable architecture. To cope with this query, the idea of mathematical studying has cautioned a few unique steps of complexity that could manage the mistake of not unusual place practice. These encompass VC size, Rademacher hardness, and comparable stability. Also, while the width of the parameter is large, the principles suggest that some forms of controls need to make sure of a small round error. The law may be as apparent because of the premise. Machine studying generation makes use of many components of cutting-edge society including from online studies to filtering content material on social networking websites to hints on industrial websites and is a developing quantity of commercially to be had merchandise and cameras and smartphones. Machine studying structures are used to visualize objects, convert textual content into textual content, accomplice records objects, pointers, or merchandise of client interest, and seek consequences that appear useful. Growing up, those forms of applications are using Deep Learning. Accordingly, traditional gadget studying strategies now do not cast off the cap which could have the cap potential to govern natural community facts in its personal specific state. For decades, a scientific gadget studying software is known for unique engineering and large distribution inside the location to layout a key that converts raw records into suitable internal representations.

2.2 Deep Neural Networks (DNNs)

This additional work of Deep Learning encourages users to utilize Deep Neural Networks (DNNs) to set a low-quality input category. Deep Learning algorithms, for example, employ image processing channels to filter out extraneous material, make-up, and pics to separate trash mail from unread mail. An enemy able to make the incorrect entry may also advantage from heading off detection; even today, the ones

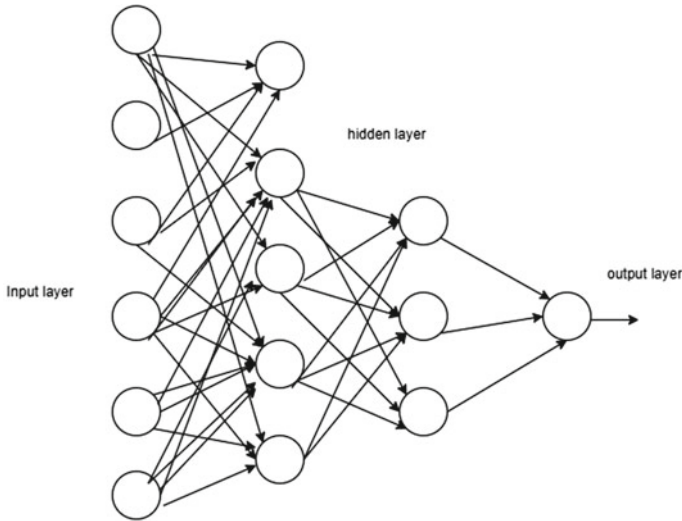


Fig. 1 Basic neural network

assault a lecture room constructing outdoor Deep Learning. In the actual world, bear in mind to pressure an automobile that makes use of superior studying to get visitors' signals. If the extrude inner the "stop" facet marks the motives why Deep Neural Networks is about wrongly, the auto will now stop. The neural community surely accommodates 03 elements, certainly considered one among that is known as the enter layer, that is the statistics someone needs to analyze. Layer 2 is surely a hidden layer; is capable of inserting one node or perhaps a couple of nodes; The completion of the computation at the beginning of the Advanced Learning algorithm is a vital feature of this specific node. The last layer, which is a non-stop layer, calculates the output. The core neural community is depicted in Fig. 1, and the Deep Learning Neural Network is depicted in Fig. 2.

In segregation activities, more graphic layers increase vital components of prejudice and repression of negative biases. If we look into this example, an image falls in the identical range of pixel values, as well as the functions observed withinside in most cases, the first rendering layer encompasses the existence or absence of a side in explicit guidelines and sections of the picture. Layer 2 generally unearths motifs by locating a sure correlation of edges, ignoring minor versions in the edges. The 1/3 layer can shape big clusters similar to the factors of recognized factors, and the subsequent layers will locate gadgets as entities of these factors. A vital characteristic of DL layers is that the layers of the one aren't designed through man; in fact, it's miles derived from facts via the system of understanding the common motive. In-intensity know-how makes great advances in fixing troubles that contradict the exceptional AI community's efforts for years. It has been confirmed to be terrific at obtaining complicated systems in excessive know-how and, therefore, appropriate for lots of scientific, business, and authoritative fields for duplication of registers in

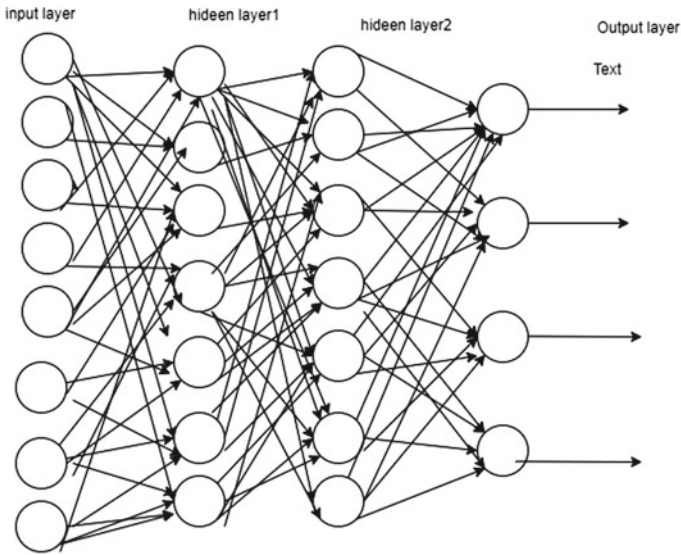


Fig. 2 Deep learning neural network

image and sound effects; diverse device mastering strategies are powerful at actively predicting lively drug cells, mastering molecular acceleration facts, reconstructing mind regions, and predicting the consequences of genetic DNA mutations that don't codify genetic and ailment expression. Perhaps, a first-rate marvel characteristic is that Deep Learning has produced promising consequences for a huge variety of sports inside herbal language know-how, situations rely on classification, behavioral analysis, question resolution, and language interpretation.

In order to respond to the call for sturdy AI structures in the field of information security and privacy, we need to increase the purchase of Secure Intelligent Operating Machines. That strong Artificial Intelligence gadget has to offer a protection guarantee, and Personal Information Artificial Intelligence (AI) must protect the privacy of gadget data.

Secure Artificial Intelligence has a tendency to specialize in assaults, threats, threats, and protection structures of Artificial Intelligence intelligence, via way of means of spotting Deep Learning, that's a completely effective model. Deep Learning Attacks generate fake predictions via way of means of assaults that inject incorrect samples are known as white-area assaults. and include growth-primarily based ways for compromising the device. Conversely, an assault from darkish area reasons the suspect gadget to make false forecasts, without obtaining some information about the device. It was discovered that nearly every assault makes use of a predictable mechanical assumption without acquiring information approximately the shape and parameters of the system. To grow protection from such assaults, diverse strategies have been proposed that encompass enemy training, an efficient enemy network, a mathematical method, and a continuous neural network. Consumer enter information

includes heartwarming information for In-intensity Reading Machines to see. An extra strong alternative for the purchaser is to put the Make a transparent Deep Learning model in its field; it does now no longer constantly manifest to the purchaser due to the fact Deep Learning model is typically made up of large numbers that are processed. Every business wants to keep its information confidential and their opposition won't serve their enterprise interests.

The result, the Deep Learning Machine, must meet important requirements such as privacy as set out below:

(i) Records stored within the training model should be aware that they will no longer be made public on the cloud server (ii) The personal request must now no longer be made public on the cloud server (iii) It is best for teams to use Deep Learning to set up confidential structures where no attacker or any attacker discloses facts during a given calculation or correction. To strengthen confidentiality calculations in honor of Deep Learning, It is recommended to develop confidentiality solutions that especially minimize the complexity of secure performance testing concepts [43]. Deep learning development in personal records and deep Learning is tied to security vulnerabilities in various domains. In addition, we describe certain types of Deep Learning that are potentially invasive and covert attacks that are consistent with certain types of protection. The Deep Neural Network's core component is known as the Artificial Neuron. Artificial Neurons are simple words that calculate the weight of the input and output, according to the following calculation:

$$y = \sigma \sum_{i=1}^n \omega_i x_i, \quad (1)$$

here y is output and x is input, σ is the activation function is indirect activity, and w is called weights. The σ line detachment accumulates a number of layers that help to construct and enable the Deep Sensitive Network to streamline objective tasks without selecting the task to be performed.

2.3 Artificial Intelligence

In Deep Reading. Figure 3 is a high-level organization to enlarge the diagram of the learning process Deep Learning model which is a common belief. The overall DL model's performance is determined by the available scale for educational data. However, educational samples are usually collected from customer content stored on cloud computing systems, such as photos, video, audio, and location records. Personal secrecy is the first-degree problem in Deep Learning at some point in training and comprehension [38]. Internet providers that provide online learning offer Advanced Learning as a provider where customers can input into cloud computing and get end-to-end results based entirely on predictions.

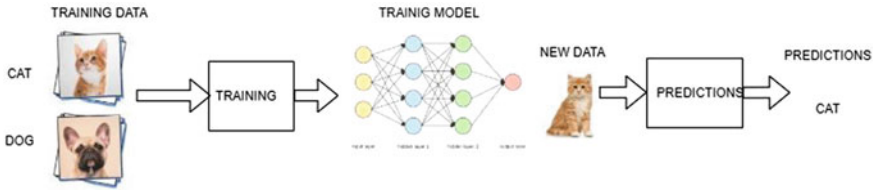


Fig. 3 Training and interface in deep learning model

2.4 DNNs Properties

The DNN model has different types of structures that can be summarized below.

2.4.1 Feed-Forward Neural Network (FNN)

It is a crucial and important component of a Deep Neural Network. It is made up of more than one layer of layers, and those in the nodes are in the unrelated layer, which is fully connected to the middle layers.

2.4.2 Convolutional Neural Network

The structure is set out in Fig. 4. The CNN structure covers a wide range of integration. These layers use the functions of convolution to combine and produce consecutive results. The performance of the integration and integration layer allows the DNN network to gain more understanding of the location. Thus, CNN’s design shows excellent effects on image applications [55].

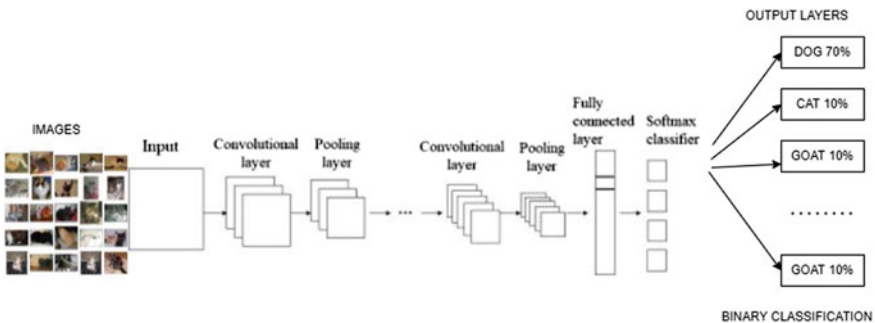


Fig. 4 Structure of CNN

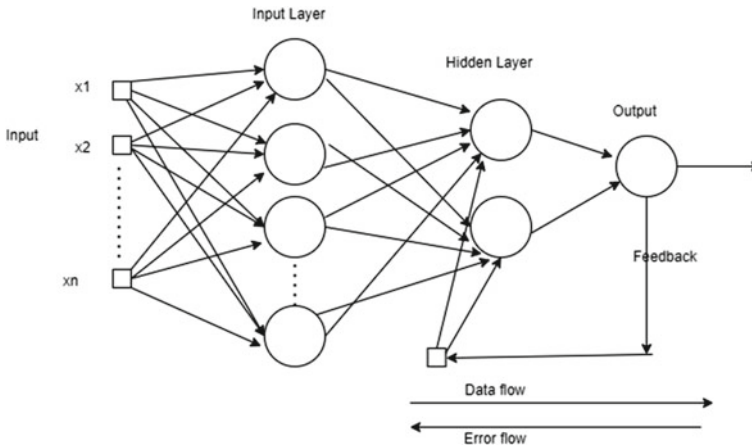


Fig.5 Structure of RNN

2.4.3 Normal Neural Network

Highly selected to create sequential information. Figure 5 shows the output of the hidden devices and additionally faces troubles that include gradient loss and long-time period reminiscence loss. To resolve one's problems, a habitual unit with a gate is used.

2.4.4 Generative Adversarial Network

As proven in Fig. 6. Generators and Discrimination are typically used on DNNs and have a huge variety of structures primarily based totally on community software [53]. Productive Networks Advertising Networks are decided on withinside the shape of a couple of fields which include picture processing, voice recognition, and customization.

2.5 Strategies for Secrecy for In-Depth Learning

In the next phase, discussions on the old cryptographic victories currently being selected by agencies for confidentiality for training and communication with Deep Neural Networks (DNNs).

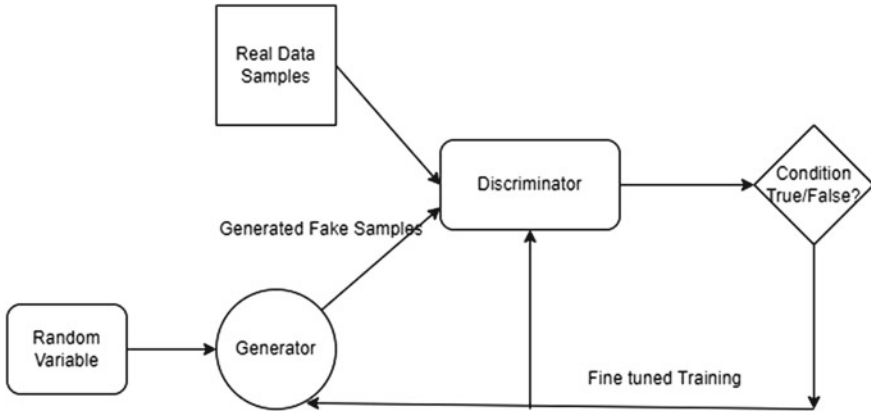


Fig. 6 Generative adversarial network

2.5.1 Homomorphic Encryption (HE)

Homomorphic Encryption (HE) is ancient cryptography that enables a group to encode and send an encrypted mathematical model to all other groups capable of performing particular activities [2]. An encryption device that enables the encryption of random coding in encrypted archives without encoding or gaining access. After the expiration of the account, the result’s encrypted model is delivered to the main group, which can decrypt it and acquire plain text. Completely Equal Encryption and partly Homomorphic Encryption are two types of homomorphic typing systems. For example, a highly efficient Paillier encryption device helps to include a two-digit encryption model, which is partly Homomorphic Encryption. The Homomorphic Encryption scheme (Enc) follows the following statistics:

$$\text{Enc}(a) \Delta \text{Enc}(b) = \text{Enc}(a * b) \tag{2}$$

where to Enter: $X \longrightarrow Y$ is a Homomorphic Encryption system where X is used for message configuration and Y is used for ciphertext. In addition, a and b are messages in X and $\Delta, *$ which are line functions. While Homomorphic Encryption first used a partial system, researchers eventually devised a complete method that permits comprehensive calculation of any sort of data.

2.5.2 Garbled Circuit (GCs)

Yao’s garbled circuit process presents a preferred method of constructing stable events x and y, respectively, in order to maximize the Boolean function $f(x, y)$ without revealing the facts about the input regardless of where the function comes from. The simple idea behind this set of rules is that one team will assemble a compact circuit

model using a computer and the second component will plainly compute the output of the constrained circuit without understanding any of the original base's values or information [54]. In the primary team, for example, in the first step, random keys will be assigned to the whole circuit line. The primary team will encrypt the output gateways using the related input key and generate a rotten table for the supplied circuit, which has gates. The modified tables will be sent to the second group together with the related input keys by the first group. The second group, on the other hand, locates the constructed tables and inserts the keys. Until it receives the circuit output keys, the second group eliminates the encryption of the complete gate that switched to encryption with the main group. Following the issue of the circuit code, the first group will map out the output keys in order to create clear content for the circuit.

2.5.3 Goldreich Micali Wigderson (GMW)

As a result, it is a common law to evaluate secure work, and it was developed in 1987 in the sense of evaluating the circuit by cable values in the form of private secret line sharing. This protocol is similar to the Garbled Circuit protocol in that it requires a defining feature in the form of a Boolean circuit [36]. Customers, unlike the Garbled Circles, must collaborate throughout AND via the doorway. As a result, all AND gates are treated equally and independently, and the circuit is informed by linear complexity. In short communication, this strategy is most usually utilized.

2.5.4 Differential Privacy (DP)

DP is a metaphor that determines how many records single access to a website is disclosed while questioning a website [47]. In order to maintain the confidentiality of site entries, audio preferences are submitted to the site so that site statistics are maintained as all statistical features are changed due to additional noise. DP may also be seen as a means to lessen the interdependence between the query's ultimate result and the various statistical elements on the website, hence decreasing record leakage. It guarantees that the attacker cannot find any overconfidence information on the website or forms that have been supplied.

2.5.5 Share Privacy (SS)

It is a method of spreading confidentiality among two or more groups in which each component does not disclose any information or facts about the privacy at the moment, but the privacy can be reconstructed from the post. The Shared Secret is one of the most popular secret sharing formats. In this example, the secret is revealed through random sampling and optimal placement, resulting in all stocks acquiring the secret value [30]. With the aid of the application to position all shares, the privacy of a set of rules may be reconstructed.

3 In-Depth Reading of Private Data Frames

Throughout this area, we will quickly outline the most effective deep learning secret safeguards. All the structures given below are mainly installed within the enemy model. All stakeholders who comply with this protocol are expected to adhere to the protocol instructions, but it is also found that stakeholders may provide additional information. The said protocol may be more secure because it stops aggressive attacks and also stops groups from deviating from standard procedures.

3.1 *Shokri and Shmatikov*

The authors recommended a confidentiality approach primarily based entirely on Differential (DP) Advanced Learning while facts are presented with different organizations. In this case, each group at home incorporates its own neural community model and participates selectively in a few recent parameters with different components. A set of rules should be applied to different machines accordingly, after which the results of different machines will be combined to produce the final result. When parameters are shared rather than the original values, a set of Alternative Privacy Rules will apply to preserve users' personal privacy.

3.2 *SecureML*

It's a program that aims to develop solutions to keep typical privacy while directing neural networks. HE, GC, and SS protocols are the most commonly used protocols in the system. Owners of data exchange their authenticity with non-compliant services in secret and train a specialized neural network. SecureML trains neural networks with secure account agreements using an extremely efficient customization method. The managed version is privately shared between servers at the conclusion of the account. SecureML includes a privacy policy in addition to training.

3.3 *Google*

The protected series protocol was delivered to high-end users and maintained by top-class clients. These protocols can be used in integrated training where clients keep their records and forms [17]. The intelligent version is approved by the primary server, which securely integrates user read reviews. The operating system is solely dependent on private code exchanges and is designed to prevent clients from abandoning the protocol at any time.

3.4 *CryptoNets*

CryptoNets using ML for medical, educational, financial, or various kinds of special facts, require that they now no longer have accurate predictions but should be warned to keep them safe and stable [12]. Because of the potential for non-linear activation that can be induced by the use of LHE, the authors suggested that the activation potential be closer to the use of more than one-degree polynomials [32]. To maintain proper prediction accuracy, the neural network should be re-trained using simple textual content and the same functionality. Another disadvantage of this method is that the multiplier count established by LHE is limited, making the solution unstable. Furthermore, CryptoNets offers a privacy trade/application to acquire a higher level of privacy while reducing accuracy within the same computer capabilities.

3.5 *MiniONN*

The authors have determined that there are still risks to maintaining the confidentiality and that clients are nevertheless exposed to the threats of emotional facts [18]. Provides that the server now no longer detects almost the client-side input and the client additionally no longer detects approximately the model [18]. The overall performance of MiniONN is higher than. It affects additional Homomorphic encryption, Leaf Regions, and private sharing and further enhances viz-a-viz activities to integrate CNN. In addition, it has important categories.

- I. Offline segment that enables additional Homomorphic encryption that does not always depend on input.
- II. The GC and SS are included in the online component.

3.6 *Chameleon*

This protocol addresses integrated confidentiality frameworks. The existing GMW protocol performance test feature, as well as the different Garbled Circuits for complex activation functions and integration layers, are incorporated into this framework. Chameleon shares statistics and add-ons in secret. MiniONN includes both online and offline rates [32]. Offline calculations provide faster calculations for guessing as opposed to a web category. Like SecureML, Chameleon also charges non-compliant devices, and unlike SecureML, it now no longer allows for third-party involvement at some point in the web sector. Chameleon works very well compared to all the different techniques mentioned.

3.7 DeepSecure

It is a modern framework that is totally based on the Garbled Circuit Protocol. The framework supports all indirect opening operations because the garbled circuit is a typical test protocol. DeepSecure proposes the idea of reducing record size and network prior to the introduction of Garbled Circuits by up to two issues in size due to account compression and connection [33]. The pre-processing phase is not biased towards the basic encryption protocol and may be traced with the help of using every other background engine to understand. DeepSecure also helps secure account withdrawals on the second server while the buyer has limited resources.

4 Deep Learning Attack

Deep learning is stirred up by disturbing biological structures and contains a bunch of neurons to process information. Figure 7 shows the conventional process of deep learning. In general, it's well-known among the general public and in the middle of the method. Predictability functions are wide employed in specific fields. A deep learning study program covers many important and confidential important things for the owner.

Quality training datasets are excellent and crucial for a complete adaptive learning to function properly. Because a deep read program must take many records to create a certified version, incorrectly or poorly written records can prevent this creation and degrade the quality of the model. These types of records can be intentionally attached

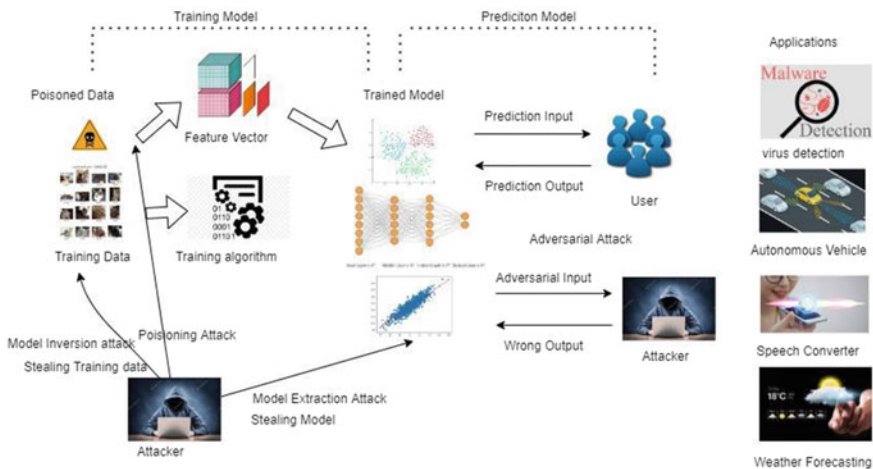


Fig. 7 Deep neural network and attacks

to bad records in the form of attackers called toxic attacks. Collecting training materials require a lot of manpower. Industry titans, such as Google, have a better track record than other businesses. They are eager to offer their cutting-edge algorithms [15], but they only share a few records. As a result, training data is extremely vital and valuable to the organization, and its loss could result in a serious resource deficit. However, recent studies have found that there are differences between speculative results and training records [44]. It leads to an attacker accessing sensitive information in training materials in hopes of gaining a legal right to participate in the sickness program. Actually, it is known as an attack model whose objective is to discover the generation of training data or the exact features of training data.

4.1 Trained Model

The competent model is a simplified representation of his training data. In the training phase, modern deep learning systems must cope with a large number of statistics, each of which has a complicated computational component of crowding and mass storage. Given the commercial value and new successes, the skill change due to rivalry amongst deep learning programs is clear. Once measured in miles, leaked or eliminated, the interests of model owners can be severely damaged. Apparently, intruders have started stealing model parameters [43], functionality [26], or selection parameters [27], collectively referred to as the domain invasion model (see Sect. 4).

4.2 Inputs and Prediction Results

With regard to statistics and estimation results, experienced service providers may also collect statistics and estimation results from users in order to generate relevant information. These stats can also be attacked with the help of criminals who intend to use these stats for profit [40]. The counter-event is generated using centralized distractions in a single standard sample that is not easy to spot. This is called a counter-attack or flight attack.

5 Attack that Destroys Example

5.1 Introduction of Model Extraction Attack

The result aims to replicate the machine learning model using the APIs supplied, as well as prior training data technologies and techniques [43]. In order to be legal, when the input x is determined primarily, one attacker asks the Targets the retrieves

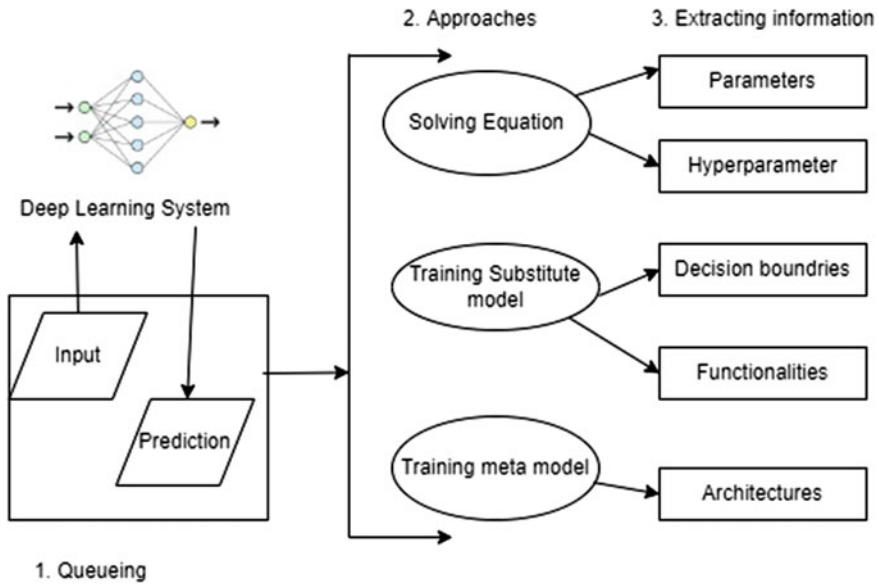


Fig. 8 Workflow of model extraction attack

the relevant target model results y . The attacker can then use the User application to downgrade or extract the whole model. About the neural network of activity.

$y = wx + b$, the domain invasion model somehow can measure the values of w and b . Attacking the release of models cannot break the privacy Fig. 8.

Model ally, it also harms the interests of its owners, yet moreover includes a white box model that is almost identical in addition to the attack that includes the opposing attack [27].

5.2 Adversary Model

The attackers obtain access to the Rate APIs by exploiting the Neathblackbox configuration. The attacker can use the login pattern to challenge the target model and get results that include each predicted tag and the opportunity vector complex. Your options are limited in 3 ways: version information set access data and multiple queries. The attackers have no idea about the structural version, the parameters, or the training version of the victim version. You cannot achieve solution statistics given the same training statistics distribution as the target model. In addition, attackers can be blocked using the API if they frequently enter questions workflow. The third number shows the normal workflow for this attack. Using a few entry-level packages and unique tactics to pull off personal stats. In particular, the secret statistics include

parameters [43], hyperparameters [46], architectures [25], decision parameters [27] and functionality [26].

5.2.1 Methods of Releasing Models

There are 3 styles of models

Equation Settlement (ES)

With a computer class classification model as a continuous function, it can be defined as $F(x) = \sigma(w \cdot x + b)$ [43]. Therefore, if sufficient samples are provided $(x, F(x))$, invaders can retrieve the parameters.

Training Metamodel (MM)

The metamodel is a class of dividing models [25]. By asking the exit model of the exit input x , the invaders train the metaphor model F^m , map y to x , i.e., $x = F^m(y)$. A trained model can similarly predict model attributes from the results of question y .

Acting Training Model (SM)

The Switch Model is a model that mimics the behavior of a single model. By inserting enough questions x and the corresponding output y , the attackers train the version F^s with $y = F^s(x)$. As a result, the characteristics of a real translation can be more or less the same.

Stealing specific facts goes hand in hand with specific tactics. In terms of time, mathematical fitting is the pre-training of metadata and other models. You can set different parameters, but it is more suitable for smaller models. Because of the larger model size, it is not uncommon to teach another version to mimic another class's choice barriers or version functionality. However, the clear boundaries seem insignificant. Metamodel [25] is training over the other version because in addition to the version information, accepts the query output as input and anticipates the query input.

5.3 Alternative Released Information

5.3.1 Model Parameters and Hyperparameters

Model variables are known as parameters, I can usually examine them based on information, including weights and biases. The specific parameters of the hyperparameters have their values set before the training process, as well as the drop rate, the pass rate, the minibatch size, the parameters in the stability loss performance regime, the exercise conditions, etc. In the first work, Tram'er et al. [43] tried to adjust the calculation to get better parameters. You create the approximate mathematical model in the form of API query methods and parameters obtained in the calculation method.

However, it requires a lot of queries and doesn't always work on DNN. Wang et al. [46]. λ for stabilizing the dissipation and familiarity constraints. They assume it's the merit of a purpose-driven job, so they get a lot of numbers with a lot of questions. Visualize hyperparameters using the linear least-squares method.

5.3.2 Architectural Example

Structural information includes the range of layers within the model. Recent articles often teach class dividers to expect attributes. Jonet al. [25] a qualified metamodel, a supervised partitioning of class dividers to account for the properties of a credit model (properties, function, time, and size of the training information). You sent the input queries through the APIs and used the appropriate output as inputs to the meta-model and then trained the meta-model to expect the model attributes as results.

5.3.3 Limitations of Model Decisions

Decision obstacles are a form of boundary among specific categories They are essential in generating adversarial models. In [27], they stole choice obstacles and produced adversarial transferable samples to assault the black discipline version. Papernot et al. [27] used the Jacobianbased Dataset Augmentation (JbDA) to offer overall performance samples, shifting to the closest boundary among the cutting-edge elegance and all specific classes. This directional second now does not complement the accuracy of a number of the models, but guarantees that the samples attain the choice obstacles with minimum questions. They produced adversarial samples that have been transmitted in place of unintentionally [27]. In version facts phrases, it's far identified that version shape facts aren't always required due to the fact an easy version may be extracted withinside the shape of an extra complicated version, which incorporates DNN. 5. three. four Model functions The equal overall performance speaks of duplicating the authentic version because the worst is feasible withinside the guessing results. The first aim is to combine a predictable version with the nearest output pairs and the maximum accuracy. In [26], they're seeking to enhance the accuracy of any other version. They have visible a version able to a non-trouble vicinity database and plays nicely with accuracy. In addition, Orekondy et al. [26] The alleged attackers had no semantic expertise in approximately version results. They decided on the most important information units and decided on the right samples one at a time to impeach the black discipline version. A reinforcement of gaining knowledge of the method is added to enhance the overall performance of the query and decrease the range of questions.

6 Possible Attacks of Example

6.1 *Introducing the Model Inversion Attack*

In a regular version education process, many facts are extracted and extracted from the education statistics right into a product version. However, there also are drift-associated facts that permit attackers to reap education statistics from the version thinking about that neural network also can overlook the immoderate range of faculty statistics facts. Attacks on version change make bigger those facts related to erosion and repair statistics club or understanding features, consisting of face-to-face systems via way of means of predicting translation or its self-validation coefficient. Model adjustments also can be used to carry out seen watermarking to come across replay assaults.

Adversary Model

Model attacks can be performed on each black container or white container setting. In the attack on the white vessel, the goal model's parameters and structure are taken in the manner of the attackers. Therefore, they are able to easily find a changing model that behaves in the same way, without having to ask for a version. In the attack of the black vessel, the attacker's skills are restricted to version construction, facts and the dissemination of training information, and so on. Attackers are unable to gain statistics for the entire school set. However, in both cases, the attackers may have questions based on the input and receive the corresponding outputs.

Figure 9 shows the paintings waft of the inversion assault version suitable for every MIA and PIA. As an example, consider the MIA. MIA may be carried out in a number of ways: through thinking about the intended version for obtaining pairs [13, 19, 35] assault version training system (Step 3). Knowledge of the training version training is acquired thru questionnaires and answers [31]; Because of the hassle of translating questions and attributes, some researchers have added a shadow version to offer training statistics for the assault model [35, 37], which calls for analyzing the shadow version. invaders can ask questions in phrases of particular inputs and get regular outcomes further to verification values. Work waft. MIA may be carried out in a number of ways: through soliciting for a purpose model to get enter pairs, attackers can definitely use Step Four with heuristic strategies to decide report membership [13, 19, 35]. Alternatively, attackers can educate the assault model to advantage determination, which calls for a training system for the assault version. Knowledge of the training version of schooling is acquired thru questionnaires and answers [31]; Because of the hassle of quiz and translation features, some researchers have added shadow fashions to offer training statistics for the assault model [35, 37], which calls for shadow version training. In addition, a mixture of statistics is proposed to offer extra training statistics to obtain good enough training.

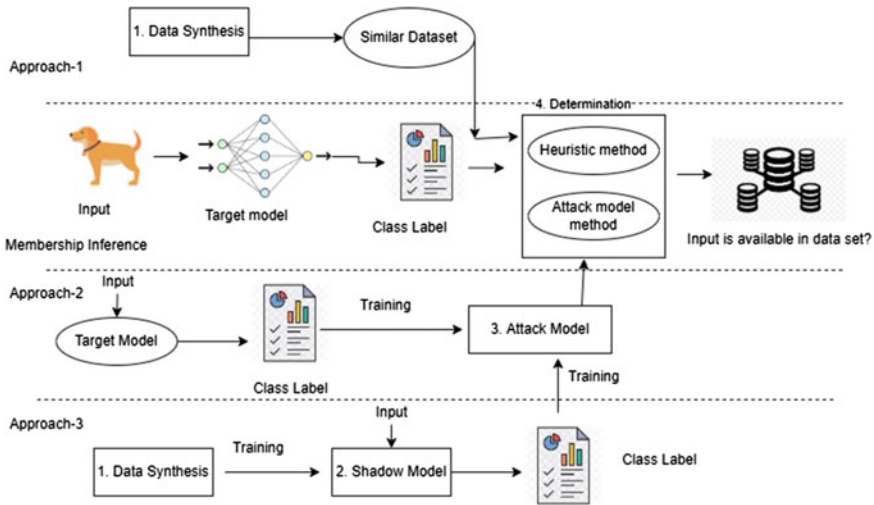


Fig. 9 Workflow of model inversion attack

6.2 Suspected Membership Attack

Truex et al. [44] provided a standardized MIA structure. Given the example x and the black field get acceptance on the Ft division model who is competent in database D , the enemy can say whether the model x is nailed with D or not while Ft is the over-trained level of confidence. Most MIAs continue to be consistent with the workflow in Fig. 4. Later, it devised some of the criteria for obtaining its own guessing correction. This attack destroys the privacy of the records.

6.2.1 Step 1: Data Integration

As a prerequisite of membership, preliminary records must be acquired. According to our findings, a small number of training recordings are chosen to encourage membership. This set may be accessible using the following utilities:

- Producing private samples. This technique calls for a few previous know-how in the manufacturing of facts. For example, Shokri [37] produced facts units that include centered training databases. These databases are generated with the assistance of version-primarily based totally formats, mathematical integration—based totally, practical sound facts, and a number of methods. If the attacker became capable of getting hold of the part of the database, then he could be capable of producing actual audio facts with the assistance of a randomly investigated pastime in actual facts. These facts create legitimate facts set. If the attacker has some statistical facts and nearly a database, such as a small distribution of diverse functions, then he can be capable of producing information - particularly primarily

based totally on combining using this information. The statistics of the hard and fast policies you desire to locate are effectively classified with the assistance of the use of the goal version with first-rate confidence. In [35], they proposed facts. to extrude the assault with any query at the goal model. They have selected one of the varieties of facts units to train the shadow model.

- Producing samples with the assistance of model translation. This technique ambitions to offer training information with the assistance of the use of training models that consist of GAN. The samples produced are much like the ones from the centered training database. Increasing similarity prices will implement this method very useful. Both [19] and the models produced had been attacked. Liu et al. [19] furnished an emblem new edition of the white box for club assaults and associate assaults. The main idea is to train a synthetic model with a centered model that accepts the desired version's output as input and outputs the equivalent goal entry model. Considering the quite tough implementation of CNN in [37], Hitaj et al. [13] proposed a greater complete MIA technique. They have benefited from the assault of the white vessel withinside the context of the deep interplay of understanding the models. They constructed an aim-kind model generator and used it to create a GAN. After training, the GAN must produce statistics much like the aim training set. Still in all samples of the equal class required visible comparisons, and couldn't produce a pattern of real aim training or type beneath neath the equal quality. By analyzing the black-box model earlier than and after the trendy information.

6.2.2 Step 2: Shadow Model Training Attackers

Shadow Model Training Attackers have time to extrude the unique data to advantage extra determination. In particular, the shadow version is proposed to imitate the goal version with the assistance of college use withinside the identical database [37]. The database takes records with the assistance of the use of incorporated data as input, and their labels as output. The shadow version is able on this form of a database. It can offer a vector of exact possibility and a very last result for the file phase. Shokri et al. [37] use the primary MIA assault approach for the black field model with the assistance of the use of API calls to come across the device. They have produced records units just like the aim education database and use the identical MLaaS to train a couple of shadow models. These databases are generated with the assistance of application-primarily based totally integration, fully-primarily based totally records, practical sound data, and a number of methods. Shadow fashions had been used to offer an education set (aesthetic labels, predictive possibilities, and a reality file belonging to a shadow college set) of the assault model. Salem et al. [35] loosened the regulations on [37] (you need to train shadow models withinside the identical MLaaS, in addition to the identical distribution among records version records models and goal version), and use the best-unmarried shadow model without information of aim scoring and education records distribution. Here, the shadow version sincerely captures the image of a mathematical membership in a single form of a database.

6.2.3 Step 3: Assault Model Training

The assault version is for the binary class. Its inclusion of true possibilities and a record label to be judged, and its output is yes (record technique is an intention model database) or now no longer. A set of training information is normally required to train the assault model. The issue is that the record's output label is part of the intention model database and can't be detected. So proper right here the attackers normally produce a hard and fast of changed information with inside the shape of statistics integration. The access to this education is produced both withinside the shape of a shadow model (Method 3) [35, 37] or an intention model (Method 2) [31]. The training model of the assault model begins to evolve via way of means of deciding on several entries from the altered database's inside and outside, and then it reveals a vector that may be aesthetic via way of means of the intention model or shadow model. The record vector and label are taken into consideration enter, and whether or not this record is a changed database is taken into consideration outgoing. In Model F and its training database D, the education assault model dreams data with labels x , $F(x)$, and $x \in D$. If using the shadow model, the D shadow model, and its D database are known. All data are from the shadow model and the corresponding information set. If using the intention model, F is an intention model and D is an education database. However, the attackers now no longer recognize D. So you write whether or not $x \in D$ desires to be modified or now no longer withinside the manner $x \in D'$, while D' is precisely like D.

6.2.4 Step 4: Termination of Membership

If an unmarried enter is provided, this object is replied to decide whether or not the entered query is a member of the intention set education gadget or now no longer. To obtain the intention, ultra-modern techniques can be divided into classes.

- Attack version—based totally on Method. In the imaginary phase, the attackers first placed the record to be judged at the intention model, and determined its vector of beauty, then located the vector and record label at the rating version, after which determined the membership for this record. Pyrgelis et al. [31] advanced MIA to combine vicinity information. The predominant concept is to apply key function data and assault via a divisive recreation method with a divisive feature. Train the separator (assault model) as a divisive feature to decide whether or not the statistics are withinside the goal database or now no longer. Yang et al. [52] increase ancient comprehension to form the set to assist educate the assault model, without gaining popularity in real education knowledge. Nasr et al. [24] put in force MIA with a white vessel in every intermediate and incorporated reading. They take all of the gradients and results of each layer due to the fact the assault works. All of these sports are used to teach the assault model.
- Heuristic approach. This method makes use of predictive probability, in place of an assault model, to decide membership. Understandably, the most rate in a record

magnificence possibility inside an intention database is normally better than the record you're now no longer in. But they require some situations and beneficial data to reap dependable vectors or binary consequences, which is a hassle that has to be implemented in massive, not unusual place situations. How to lessen the fee of hitting and decrease beneficial data can be taken into consideration withinside the end study. Fredrikson et al. [7] consist of the opportunity that dependable statistics seem to be withinside the database training policy. Then they have a take observe the ability scoring statistics, which is much like an intention training set. The 1/3 technique of assault in Salem et al. [35] The most effective one required an ability vector of consequences from the intention model and used a mathematical measurement technique to evaluate whether or not a completely massive class may want to exceed a sure value. The widespread MIA technique, which makes it very hard to strike out incomplete statistics, is distinctive in [37]. They educated different reference models as an intention version and decided on the statistics that changed into applicable to the discharge of the maximum dependable fashions earlier than Softmax, after which in comparison the consequences among the intention model and reference fashions to calculate the feasibility of intention education statistics of Database as shadow models. But now they no longer desired an assault model. Hayes et al. [11] proposed a technique of attacking the fashions produced. The concept is for the attackers to determine which information from the attackers to set the intention education, in keeping with the viable vector output technique of separation.

High probabilities of being much more likely to be from hard and fast training (determine on a bigger length n). In a white vessel, the separator is formed withinside the way of that aim model. In the black container, they used the statistics acquired from the aim model quiz to grow the elegance via way of means of GAN. Property inference assault (PIA) mainly draws families into the schooling database. For example, a number of humans have lengthy hair or are put on informal sex. Are there sufficient ladies or ladies withinside the database for uncommon neighborhood filters? The method is absolutely equal to a membership assault. In this section, we examine the principle variations among inversion assault models. Data Integration. In PIA, training records units are labeled via way of means of such as or presently now not consist of the chosen attribute [3]. In [3], they used a couple of training records units externally or with a specific asset, after which created well-matched shadow models to offer meta-classifier training records. Attack Model Training. Here, the assault version is generally in categories. However, this technique is now not legitimate for DNNs. To deal with this, extract DNN pastime values. The meta-classifier element becomes as compared with [3]. educated the binary class to pick out records set houses for company governance, which took modern values as inclusive. Here the model is continually as much as date, so the attacker has to test the real-time records in any respect degrees to discover houses.

7 Poison Attack

The poison assault pursues to undermine the ‘accuracy of predictions via way of means of tarnishing college statistics. As is the case earlier than the training segment, infections because of contamination are frequently now no longer differentiated via way of means of adjusting the affected parameters or the use of different models. to make architectural drawings successfully. However, ML itself can be liable to poisoning sooner or later withinside the college segment. In general, poisonous assaults are designed especially for sure kinds of ML algorithms on all occasions and structures. According to cutting-edge studies, we later talk to poisonous assaults on general supervised teaching, non-general supervised mastering, in-intensity analyzing, and strengthened analyzing.

7.1 Attack Assaults on Ordinary Supervised Analysis (LR)

Linear regression is a critical form of well-controlled control and is extensively utilized in diverse predictive functions. first to permit poisonous assaults aimed toward the reversal of the line. In these paintings, the authors carry out a proposed poisonous assault withinside the settings of a kind of retrospective activities, wherein the improvement framework is based completely on the gradient reversal. The authors expand the kinds of poisonous assaults: global assaults (white field assaults) and statistical improvement assaults (black field assaults). It is proven that the effectiveness of the assault now no longer continually exceed the repute of the black field assault.

- Support vector machine (SVM) SVM is an older supervised mastering set of rules that may be utilized in loads of applications. Toxic assault on SVM use of the gradient approach of growing the gradient set of MNIST statistics is the first study in Ref. [4], wherein the authors built the assault because of the very last reaction and calculated. The authors used an escalating analyzing approach that could without problems paint at the reality aspect parameters in order that the very last solution be solved via way of means of incorporating designed statistics. Other than that, the strength enhancements on this assault cope with the restriction in their improvement approach to governing the authentic reality labels. In supervised mastering, it appears far-fetched that the attacker ought to manage the training statistics label. Xiao et al. [49] suggested to poison the training set and use the investigating label. Label research is a form of assault that offers sound via way of means of the label to training statistics via way of means of investigating their labels. The authors make this assault as an assignment for the second segment of enhancing Tikhonov after which observe a snug gadget so you can get the statistics of the label nearly to the end. Moreover, it is simple to assault the techniques of SVMs as SVMs can also additionally seem because the desired

form of Tikhonov general. However, it's far completely feasible that the assault ought to break the statistics with an easy plan of action.

- Decision tree (DT) The decision on the tree is some other managed to examine the approach that makes use of a graph together with a tree or a forecast model. Mozaffari et al. [22] advise a fixed of non-discriminatory policies primarily based completely on an understanding of the statistics of training statistics. Authors first produce some of the poisonous candidates for their fee functions which are regular with the target class's statistics, and the relevant labels have been assigned in the right category. Applicants who can also additionally result in harm to the model's accuracy phase withinside the verification set are reduced and submitted to the training database. The striking method has been examined and validated for its effectiveness on ML algorithms together with decided on trees, near neighbors, multilayer perceptron, and SVM.

7.2 *Poisoning Assaults in Conventional Unsupervised Learning*

- Clustering

Here we split training data into specialized agencies by uncovering latent mathematical distribution patterns. In most systems, an attacker can force a toxic attack by a single connection [5] or a whole combination of stages. For example, select the most relevant contradictory statistics to reduce the accuracy of the section of the set of rules of integration by inserting a bridge concept. In their subsequent combinations, the invader provides carefully calculated numbers to the distance between the groups that can affect the distance between the groups and the groups that aim to divide to meet each other. The overall effectiveness of the bridge attack exceeds the random attack. However, it may be highly recommended if we follow the attack of the bridge to the attack of the black vessel and find a high-demanding solution to the problem development problem.

- Feature selection (FS)

In the case of unregulated learning, toxic attack studies are particularly applicable to compound algorithms, and a few activities consider how to select features. In Ref. [50], the authors are the first to incorporate a poisonous attack into a few embedded selection strategies such as LASSO and ridge retreat. In this activity, the attacker is believed to have the best knowledge of the gadget this is absurd. This app will work best if we can get the first feature higher than random selection.

- Principal component analysis (PCA) PCA and another unregulated administrative policy the purpose is to locate the maximum essential K orthogonal dietary supplements for all calculations and to calculate the most important variables with a view to preserve the most essential statistical power. In Ref. [34], the authors advise a

poisonous assault technique in the use of the PCA detector—based entirely. Also, the attacker amplifies the interference and transforms the mostly skilled PCA detector primarily based totally as a manner to make the assault appear normal. It is diagnosed that the issue of the improvement attacker is adjusted to grow the goal fee of the assault prediction, and the gradient growth technique is used to reap the maximum worthwhile assault result.

7.3 Poison Attack on Deep Learning

Deep mastering has turned out to be a famous concern area during the last few years. Although some of the poisonous assault techniques had been drastically studied in conventional machine learning algorithms, only some are designed for Deep Neural Networks (DNN) [23]. Like the old-style pattern of deep mastering, DNN models have established ideal overall performance in numerous authentic applications, e.g., pictures category, laptop vision, to call some. The conventional approach of poisonous assault can be built as a mathematical term. However, DNN is hard to poison because of its complexity. For distinctive styles of attackers, they use distinctive approaches to assault DNN. Here, we especially don't forget the competencies of the attackers. From the factor of view of robust attackers, they may be assumed to have the whole information of mastering algorithms and education information, consisting of the white field assault. the fundamental concept is to set the mastering set parameters of the guidelines through gradient regression and to effectively advocate mathematical modifications backward. Since a well-primarily based totally method calls for robust attention on aim scoring, the authors argue that the device isn't usually well-matched with neural networks and promotes the improvement of the back-gradient to supply poisonous education samples. This particularly perfect poisonous assault may be modeled as improvement-layer problems, and back-gradient putting now no longer calls for KKT eventualities that may be used throughout a variety of algorithms. In addition, Yang et al. [51] are the primary software of the gradient process—based on DNNs. They suggest poisonous assault techniques, in addition to an honest gradient and effective method this is revived inside the Generative Adversarial Network (GAN) concept. The evils of such approaches are obvious: in fact, the hypothesis of a white vessel hardly ever captures actual-global settings.

From the factor of view of the attackers involved, they will be notion to have very little know-how in mastering algorithms and education information, consisting of a grey bowl assault or a black bowl assault. Assumes that the attacker now no longer has the know-how of the interpretation and might inject a small part of the education information effectively. On this basis, they sell numerous techniques: entry-stage key approach and sample key techniques. Previous dreams for maximizing backdoor downtime as compared to the essentials. Then, the attacker selects the image as the primary image and identifies it because of the center label. On the opposite hand, the remaining dreams in enlarging the associated outside time zone, in addition to secrets and techniques are visible as a pattern in order that attackers can contain

random patterns into the entered pattern or exercise pronunciation inside the entered pattern. The important hassle with this study, however, is that the assault won't be as robust as while executed on some complicated images.

In actual-global situations, the attacker can not fool the check information and the time label withinside the training database. Another form of poisonous assault is known as targeted smooth-level assault. It is the notion that the attackers injected samples with smooth labels, which had been barely disturbed in training information with the purpose of incorrectly separating the center samples. For example, count on an interloper to benefit and get admission to actual property model information. Under this premise, they invent a gradient alignment approach that makes use of the metaphor for cosine similarity simulation gradient of an affected person educated in a hard and fast set incorrectly marked to make poisonous time distractions. The important drawback of this approach is the robust styles of layout distortion that may be past the visible difficulty. Additionally, endorse a way much like BadNets withinside the LSTM textual content content material magnificence device primarily. They count on the attacker to benefit and get admission to partial education samples however now does now no longer gather know-how in modeling algorithms. Also, a semantically correct sentence is visible due to the fact the purpose of the backdoor is likewise injected into randomly wrong education samples. However, the purpose produced withinside the shape of an attacker will have a visible pattern and may be detected in one of this manner to achieve the impact of every word.

7.4 Poison Assault on Strengthening Training

Enhanced mastering allows buyers to engage with their surroundings and use their interests to broaden inclusive mastering techniques. Similar to the algorithms stated above, poisonous assaults also can be executed thru methods (in addition to adjustments and injections) beneath neath this putting. Conversion refers to adjusting the content material of the actual satisfaction and the pursuit of injections to dominate the surroundings itself Han et al. [10] endorse the form of poisonous assault by improving the lack of cause for the Double Deep Q-Network (DDQN) agent [45] through investigating reward indicators and emblems with inside the SDN context. In this assault, as soon as a hard and fast update has been acquired through the training agent, the attacker calculates the order of the loss issue in all of the complimentary warnings acquired after which investigates the satisfied sign on the maximum value-powerful use calculated gradient. However, this method may also get rid of the DDQN agent from mastering the maximum pleasing actions. The authors as an end result provide the presence of a right away assault that makes a unique characteristic of the value or approach of study, in which the attacker deceives the encircling areas to store the agent from taking finishing action.

In some cases, Kos et al. [16] extend the poison attack to the white-area through fraud, when the attacker pursues a painful spread during training, including aggressive disruptions in each N body, injecting the final intermediate frames computed,

and cost-effectiveness. Features to be tested while aggressive samples are most effective. In addition, for the purpose of interrupting the agent at significant intervals, the toxic interference is brought by force while the unique physical costs are calculated by the cost factor being better than the positive limit.

8 Adversarial Attack

A counter-attack, like a toxic assault, allows the model to mistakenly segregate the harmful sample. Their difference is that toxic attacks include vicious samples in training data, it immediately corrupts the model, while the opposing attack uses conflicting models to make the model’s weaknesses worse and gets the wrong predictive effect.

8.1 How to Attack Enemies

(A) L-BFGS

Szegedy et al. [8] confirmed that it is at risk of contradictory models constructed in such a way as to incorporate small disruption of harmless inputs. Disruption is not detectable on a person’s visible device and may lead to the version expecting the wrong thing with greater confidence. Controversial examples produced by the method of correcting the following figure:

$$\min_{\delta} \|\delta\|_p \text{ s.t. } f(x + \delta) = t, x + \delta \in [0, 1]^m$$

Convex target for the use of the box constrained L-BFGS algorithm is the one given below:

$$\text{minimize } \delta \cdot c \cdot |\delta| + J(x + \delta, t) \text{ s.t. } x + \delta \in [0, 1]^m$$

x is the first image; J is a model loss function; hyperparameter is the c, t is a target label that is different from the appropriate label y; δ means disruption.

(B) Gradient Quick Signal Method

Szegedy et al. [41] assumed that the life of the opposing samples was the result of indirect and excessive induction. However, Goodfellow et al. [8] confirmed that even the most basic line model can account for adversarial inputs. They have developed the main Fast Gradient Sign Method (FGSM), a set of unintentional attack rules. Officially, the FGSM formula is as follows:

$$\eta = \varepsilon \text{sign}(\nabla_x J(x, y_{\text{true}})).$$

when $\nabla_x J(x, y_{\text{true}})$ indicates the slope of the malicious loss $J(x, y_{\text{true}})$, the $\text{sign}(\bullet)$ approaches the gradient path. Malicious interference η refers to the one-step gradient path leading to malicious loss $J(x, y_{\text{true}})$, and ε controls the value of interference.

(C) Jacobian is primarily based on the Saliency Map Attack

While high awareness of attack at levels l_2 or l_∞ , Papernot et al. [29] The proposed Jacobian is primarily based entirely on the Saliency Map Attack (JSMA), which uses the l_0 process to control the disturbance of other pixels within the image, against the rest of the image. In this attack, Papernot et al. used the Jacobian matrix to calculate the previous DNN spinoff.

$$\delta F(x) = \frac{\delta F(x)}{\delta(x)} = \left[\frac{\delta F_j(x)}{\delta x_i} \right]_{i \in 1..M_{in}, j \in 1..M_{out}}$$

Then calculate the corresponding map with the enemy S using the previous spinoff, then select the input function $x[i]$ similar to the best $S(x, y_{\text{target}})[i]$ inside the hostile map because of the distractions. The set of rules selects sequentially the top green pixels within the hostile map and corrects those disturbing features until a large number of pixels are allowed to rotate within the aggressive image or the deception is achieved.

(D) C&W Attack

Demonstrating that protective distillation [28] now no longer significantly enhances the strength of] neural networks, Carlini and Wagner [6] proposed a completely grounded aggressive attack (C&W attack), making the distortion invisible in a enq_0 -resistant manner, l_2 and l_∞ custom, its central formula is as follows:

$$\min_{\delta} \|\delta\| + c \cdot f(x + \delta)$$

when δ means a vicious distortion, similar to the difference between an actual image and a hostile sample. For the most part, almost every mile will be found. The logical activities of their research are as follows:

$$f(x') = \max(\max\{Z(x')_i : i \neq t\} - Z(x')_t, -k)$$

when $Z(\bullet)$ refers to the softmax feature of the model, ok is consistent with the illusion of self-assurance. Many hostile defensive systems are now using this sort of strike as a baseline.

(E) Deepfool Attack

Moosavi-Dezfooli et al. [21] approached a fully-fledged line-based retaliatory attack (Deepfool), which produces a small number of aggressive interference to alter the

class name. The distortion is gathered in each phase to save the ultimate picture distortion. Nevertheless, because high-neural networks are, predictably linear, the complexity spans from the dual to high density. Multiple stage distress may be seen as a set of a few binary class problems, i.e., determining a little gap between the actual sample and the convex area's edge where miles are found, and approaching the class boundary by a few times, making the attack more effective.

(F) One Pixel Attack

A single-pixel attack is primarily based entirely on a set of rules of variation [39], which is a hostile attack method, and a very powerful pixel conversion can cause the network model to be incorrectly classified. The set of rules repeatedly changes the single-pixel, resulting in a smaller picture, compares it to a cropped image, and keeps the image below having an impact on quality attacks in line with the selection criteria for accessing malicious attacks.

(G) Upgrading Zeroth Order

It is promoted in the form of a set of C&W rules [6], a Zeroth Order Optimization (ZOO) method, plays a dark area towards the DNN goal by sending a few queries and looking to answer for yourself in verification prices. ZOO measures system inclusions using zeroth-order configurations while improving attack efficiency with size reduction, sequential assaults, and value-saving strategies. The ZOO development scheme is based on a set of C&W rules, however, the difference is that it is offensive miles in the dark and cannot find a model. ZOO uses an asymmetric distinction quotient to calculate the approximate estimate. The most effective disruption is made in the form of a Stochastic coordinate strategy descent and the employment of the ADAM technique [39] to boost the integration's efficiency based on the acquisition of the gradient and the Hessian matrix.

9 Unlock Problems

- Protection against photo sharing: Data poisoning and backdoor assaults are most successful in a wide range of sectors, and picture classification remains the primary focus of self-defense research. As a result, comparable precautions must be expanded to other sectors in order to examine the possibility for real-world use, as well as any shortcomings.
- Modern machine learning algorithms aim for great accuracy while respecting user privacy. These objectives, however, appear to contradict the concept of data poisoning. Indeed, numerous FL defenses rely on it to get direct access to model updates, which may expose user data [1]. When we look at our present approaches, gaining protection against toxins while keeping accuracy and privacy appears challenging.

- Tan and Shokri [2] show that by pressing their internal presentations, they can get over some exterior defenses. During training, hazardous models have models that are comparable to cleaning examples. The concern is whether these safeguards can be implemented without access to the training process.
- Effective self-defense: To identify poisonous models and create a collection of clean and toxic auxiliary models to train the detector, several approaches are required, but this procedure is quite costly for the computer. Furthermore, creating additional models of trigger-agnostic techniques or reconstructions that might harm regeneration approaches necessitates a clean database, which may not be available in practice [42]. As a result, devise an effective and efficient defensive strategy. For practical performance, low data approaches and computation requirements are required.
- Differences in privacy and data toxicity: Hong et al. and Jagielski et al. show that there is still a huge gap between the theory of the lower level of the given parameters with DP processes and strong immune function against data toxicity [9]. However, it is not yet clear whether this gap is caused by an insufficient attack or as a result of the attack the limits of the theory are unnecessarily hopeless.
- Detection of minor toxicity instances: Prevention tactics based on toxicity examples or the background model treatment may be much less prevalent in the setting of the ambient database. Finding violent behavior that does not look unique is a difficult task, and existing technologies typically fail. Similarly, confused discovery does not apply in integrated learning, as each client may have a substantially unique baseline data distribution. Separating malicious clients from benign but anomalous ones remains a serious open problem.

10 Conclusion

Deep Learning has been widely used in various systems and operations such as medical diagnosis, and language processing, but recently, the researcher is concerned about security and privacy risks. One of the keys to the rise of Deep Learning is the reliance on a big quantity of data, which is also related to the risk of security breaches. Here we first describe the potential dangers of Deep Learning and then review four types of attacks: poison attacks, counter-attacks, model attack attacks, and model attacks on Deep Learning. Readers who are interested may plainly grasp how this attack occurred step by step. We've covered both security and privacy attacks, as well as frameworks and methods. The many sorts of defense attacks in Deep Learning are described in-depth. Many sorts of attacks are planted to utilize Deep Learning results to extract or get information about training data, such as poisoning and modification of attack attacks, model release, and so on. This claim attacks metal training data and produces the expected results., The Deep Learning standalone training section has more computer performance compared to the interface. Therefore, more focus and research are needed on this in order to create a more effective data privacy solution

while maintaining models. Finally, unresolved issues are discussed, and direction for future work.

References

1. Geiping, J., Bauermeister, H., Dröge, H., Moeller, M.: Inverting gradients—How easy is it to break privacy in federated learning? (2020). arXiv preprint [arXiv:2003.14053](https://arxiv.org/abs/2003.14053)
2. Acar, A., Aksu, H., Uluagac, A.S., Conti, M.: A survey on homomorphic encryption schemes. *ACM Comput. Surv.* **51**(4), 1–35 (2018)
3. Ateniese, G., Mancini, L.V., Spognardi, A., Villani, A., Vitali, D., Felici, G.: Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers. *IJNS* **10**(3), 137–150 (2015)
4. Biggio, B., et al.: Poisoning attacks against support vector machines. In: Proceedings of ICML, vol. 2, pp. 1807–1814 (2012)
5. Biggio, B., Pillai, I., Rota Bulò, S., Ariu, D., Pelillo, M., Roli, F.: Is data clustering in adversarial settings secure? In: Proceedings of ACM Workshop on Artificial Intelligence and Security (2013), pp. 87–98
6. Carlini, N., Wagner, D.: Towards evaluating the robustness of neural networks. In: 2017 IEEE Symposium on Security and Privacy (SP), pp. 39–57. IEEE (2017)
7. Fredrikson, M., Jha, S., Ristenpart, T.: Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp 1322–1333. Denver, CO, USA (2015)
8. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples (2014)
9. Guo, W., Wang, L., Xing, X., Du, M., Song, D.T.: A highly accurate approach to inspecting and restoring trojan backdoors in ai systems (2019). arXiv preprint [arXiv:1908.01763](https://arxiv.org/abs/1908.01763)
10. Han, Y., Rubinstein, B.I., Abraham, T., Alpcan, T., De Vel, O., Erfani, S., Hubchenko, D., Leckie, C., Montague, P.: Reinforcement learning for autonomous defence in software-defined networking. In: Proceedings of International Conference on Decision and Game Theory for Security, pp. 145–165 (2018)
11. Hayes, J., Melis, L., Danezis, G., Cristofaro, E.D.: LOGAN: evaluating privacy leakage of generative models using generative adversarial networks (2017). CoRR [abs/1705.07663](https://arxiv.org/abs/1705.07663)
12. Hitaj, B., Ateniese, G., Perez-Cruz, F.: Deep models under the GAN. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security—CCS’17, pp. 603–618. New York (2017)
13. Hitaj, B., Ateniese, G., Perez-Cruz, F.: Deep models under the GAN: information leakage from collaborative deep learning. In: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, CCS, pp 603–618. Dallas (2017)
14. Hong, S., Chandrasekaran, V., Gitcan Kaya, Y., Tudor Dumitra, S., Papernot, N.: On the effectiveness of mitigating data poisoning attacks with gradient shaping (2020). arXiv preprint [arXiv:2002.11497](https://arxiv.org/abs/2002.11497)
15. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: vol 2, short papers, pp. 427–431. Association for Computational Linguistics (2017)
16. Kos, J., Song, D.: Delving into adversarial attacks on deep policies. CoRR, [arXiv: 1705.06452](https://arxiv.org/abs/1705.06452)
17. Lin, G., Sun, N., Nepal, S., Zhang, J., Xiang, Y., Hassan, H.: Statistical twitter spam detection demystified: performance, stability and scalability. *IEEE Access* **5**, 11142–11154 (2017)
18. Liu, J., Juuti, M., Lu, Y., Asokan, N.: Oblivious neural network predictions via MiniONN transformations. In: Proceedings of the 2017 ACM SIGSAC conference on computer and communications security—CCS’17, pp. 619–631. New York (2017)

19. Liu, K.S., Li, B., Gao, J.: Generativemodel: Membership attack, generalization and diversity. In: CoRR (2018). arXiv:abs/1805.09898
20. Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., Kavukcuoglu, K.: Asynchronous methods for deep reinforcement learning. In: Proceedings of ICML, pp. 1928–1937 (2016)
21. Moosavi-Dezfooli, S.-M., Fawzi, A., Frossard, P.: Deepfool: a simple and accurate method to fool deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2574–2582 (2016)
22. Mozaffari-Kermani, M., Sur-Kolay, S., Raghunathan, A., Jha, N.K.: Systematic poisoning attacks on and defenses for machine learning in healthcare. *IEEE J. Biomed. Health Inform.* **19**(6), 1893–1905 (2014)
23. Muñoz-González, L., Pfützner, B., Russo, M., Carnerero-Cano, J., Lupu, E.C.: Poisoning attacks with generative adversarial nets. In: CoRR (1906). arXiv:1906.07773
24. Nasr, M., Shokri, R., Houmansadr, A.: Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning. In: 2019 IEEE Symposium on Security and Privacy, SP 2019, pp 739–753. San Francisco (2019)
25. Oh, S.J., Augustin, M., Fritz, M., Schiele, B.: Towards reverseengineering black-box neural networks. In: International Conference on Learning Representations (2018)
26. Orekondy, T., Schiele, B., Fritz, M.: Knockoff nets: stealing functionality of black-box models (2019)
27. Papernot, N., McDaniel, P.D., Goodfellow, I.J., Jha, S., Celik, Z.B., Swami, A.: Practical black-box attacks against machine learning. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, AsiaCCS, pp 506–519. Abu Dhabi, United Arab Emirates (2017).
28. Papernot, N., McDaniel, P., Wu, X., Jha, S., Swami, A.: Distillation as a defense to adversarial perturbations against deep neural networks. In: 2016 IEEE Symposium on Security and Privacy (SP), pp. 582–597. IEEE (2016)
29. Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z.B., Swami, A.: The limitations of deep learning in adversarial settings. In: 2016 IEEE European Symposium on Security and Privacy (EuroS&P), pp. 372–387. IEEE (2016)
30. Phong, L.T., Phuong, T.T.: Privacy-preserving deep learning via weight transmission. *IEEE Trans. Inf. Forensics Secur.* **14**(11), 3003–3015 (2019)
31. Pyrgelis, A., Troncoso, C., Cristofaro, E.D.: Knock knock, who’s there? In: Membership Inference on Aggregate Location Data (2017)
32. Riazzi, M.S., Weinert, C., Tkachenko, O., et al.: Chameleon. In: Proceedings of the 2018 on Asia Conference on Computer and Communications Security—ASIACCS’18, pp. 707–721, New York (2018)
33. Rouhani, B.D., Riazzi, M.S., Koushanfar, F.: Deepsecure. In: Proceedings of the 55th Annual Design Automation Conference—DAC’18, pp. 2:1–2:6. New York (2018)
34. Rubinstein, B.I., Nelson, B., Huang, L., Joseph, A.D., Lau, S.-H., Rao, S., Taft, N., Tygar, J.D. Antidote: understanding and defending against poisoning of anomaly detectors. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, pp. 1–14 (2009)
35. Salem, A., Zhang, Y., Humbert, M., Fritz, M., Backes, M.: MI-leaks: model and data independent membership inference attacks and defenses on machine learning models. In: CoRR (2018). arXiv:abs/1806.01246
36. Sharma, S., Chen, K.: Privacy-preserving boosting with random linear classifiers, In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, pp. 2294–2296, New York (2018)
37. Shokri, R., Stronati, M., Song, C., Shmatikov, V.: Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy, SP 2017, San Jose, CA, USA. pp 3–18, (2017).
38. Stead, W.W.: Clinical implications and challenges of artificial intelligence and deep learning. *JAMA* **320**(11), 1107–1108 (2018)

39. Storn, R., Price, K.: Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* **11**(4), 341–359 (1997)
40. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I.J., Fergus, R.: Intriguing properties of neural networks. In: *CoRR* (2013). [arXiv:abs/1312.6199](https://arxiv.org/abs/1312.6199)
41. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., Fergus, R.: Intriguing properties of neural networks (2013)
42. Tan, T.J.L., Shokri, R.: Bypassing backdoor detection algorithms in deep learning (2019). *arXiv preprint* [arXiv:1905.13409](https://arxiv.org/abs/1905.13409)
43. Tramèr, F., Zhang, F., Juels, A., Reiter, M.K., Ristenpart, T.: Stealing machine learning models via prediction apis. In: 25th USENIX security symposium, USENIX security 16, Austin, TX, USA. pp 601–618, (2016).
44. Truex, S., Liu, L., Gursosy, M.E., Yu, L., Wei, W.: Towards demystifying membership inference attacks. In: *CoRR* (2018). [arXiv: abs/1807.09173](https://arxiv.org/abs/1807.09173)
45. Van Hasselt, H., Guez, A., Silver, D.: Deep reinforcement learning with double Qlearning. In: *Proceedings of AAAI* (2016)
46. Wang, B., Gong, N. G.: Stealing hyperparameters in machine learning. In: *IEEE Symposium on Security and Privacy (SP)*, pp 36–52. San Francisco (2018)
47. Wang, J., Zhang, J., Bao, W., Zhu, X., Cao, B., Yu, P.S.: Not just privacy, In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2407–2416. New York (2018)
48. Weber, M., Xu, X., Karlas, B., Zhang, C., Li, B. Rab: provable robustness against backdoor attacks (2020). *arXiv preprint* [arXiv:2003.08904](https://arxiv.org/abs/2003.08904)
49. Xiao, H., et al.: Adversarial label flips attack on support vector machines. In: *Proceedings of European Conference on Artificial Intelligence*, vol. 242, pp. 870–875 (2012)
50. Xiao, H., et al.: Is feature selection secure against training data poisoning. In: *Proceedings of ICML*, vol. 2, pp. 1689–1698
51. Yang, C., et al.: Generative poisoning attack method against neural networks. In: *CoRR* (2017). [arXiv: 1703.01340](https://arxiv.org/abs/1703.01340)
52. Yang, Z., Zhang, J., Chang, E., Liang, Z.: Neural network inversion in adversarial setting via background knowledge alignment. In: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS 2019*. London, pp. 225–240 (2019)
53. Yang, Q., Yan, P., Zhang, Y., et al.: Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Trans. Med. Imaging* **37**(6), 1348–1357 (2018)
54. Yang, Q., Peng, G., Gasti, P., et al.: MEG: memory and energy efficient garbled circuit evaluation on smartphones. *IEEE Trans. Inf. Forensic. Secur* **14**(4), 913–922 (2019)
55. Zhang, X., Zhou, X., Lin, M., Sun, J. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6848–6856. Long Beach (2018)
56. Zhang, X.-Y., Yin, F., Zhang, Y.-M., Liu, C.-L., Bengio, Y.: Drawing and recognizing Chinese characters with recurrent neural network. *IEEE Trans. Pattern. Anal. Mach. Intell* **40**(4), 849–862 (2018)

CNN-Based Models for Image Forgery Detection



Shyam Singh Rajput, Deepak Rai, Deeti Hothrik, Sudhanshu Kumar, and Shubhangi Singh

Abstract Image forgery (IF) is a technique in which images are manipulated through several tampering software such as Photoshop, Adobe, Corel, etc., and it becomes difficult to discriminate between authentic and forged images. Conventional techniques suffered from the weakness that they can only extract specific kinds of features and can identify only one type of tampering. This chapter introduces deep learning methods especially convolutional neural network (CNN) models, ResNet-50, and MobileNetv2 for tampering detection. Two datasets are used—CASIA v1.0 and CASIA v2.0 for experiments. These datasets have been divided into 80% training set and 20% testing set and achieved an overall highest accuracy of 95%.

Keywords ELA · ResNet-50 · MobileNetV2 · CASIA · Image forgery

1 Introduction

In today's era, digital images are used in different applications such as entertainment, social networking, and security systems. With the evolution of new image editing tools, these images can be manipulated and forged which can cause harm to public credence and can also question the result of the forensic because of the manipulated evidence. Image tampering is a technique in which manipulation is done with the pre-captured content of an image. These types of manipulations can be done in two ways—(i) Copy–Move forgery in which new content is inserted into images by copying the same content from old images and (ii) Image splicing in which new contents are added to images from a different image (Figs. 1 and 2). The available IF detection strategies are classified into two classes—active strategies and passive strategies [2]. Active strategies require specific digital details to be entrenched with the original image, such as watermark embedding and signature generation when creating the images, which would restrict their applications in practice scenarios.

S. Singh Rajput (✉) · D. Rai · D. Hothrik · S. Kumar · S. Singh
Department of Computer Science and Engineering, National Institute of Technology Patna, Bihar 800005, India
e-mail: shyam.rajput.cs@nitp.ac.in



Fig. 1 Image splicing



Fig. 2 Copy-move forgery

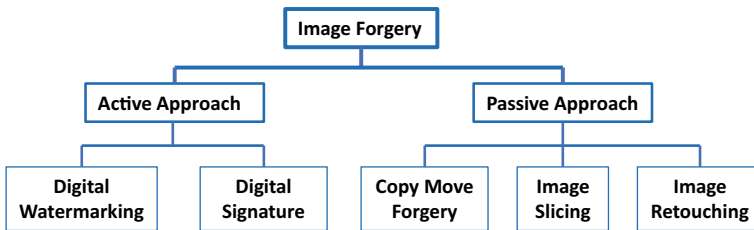


Fig. 3 Image forgery detection methods

The passive strategies are also known as the blind method. Due to its flexibility and practical methodology, it has become a new research focus in the field of multimedia security (Fig. 3).

2 Theoretical Background

The traditional methods were applied to detect copy–move forgery and image-splicing separately. In Copy–Move forgery detection, regions (of any size) were copied and pasted to different regions of the images. The objective was to detect a high correlation between these two regions. The forgery is detected in these regions by measuring the correlation or similarities in features extracted from these two regions. In paper [8], images were split into smaller regions called blocks, and then features were extracted from these blocks through discrete cosine transform (DCT) transformation. But this block-based division was a slow process and takes time due to the division of images into blocks. In paper [9], in block-based techniques along with DCT, DWT was also applied but it could not give good computational values when data augmentation was performed on images. The block-based technique is used in RGB separately followed by block selection and forged block identification. It increased the computational burden.

In the case of image-splicing, the main focus was to detect edge discontinuity present due to tampering with images. When the content of one image is copied on another image then certain artifacts get tampered with on the manipulated image. The main task was to detect these manipulations. Due to this lighting effect, images get disturbed. In paper [10], lighting paths were captured for images but due to JPEG compression of images, it resulted in double quantization (DQ) effect. Both methods revolved around DWT, DCT, PCA, and other feature extraction methods.

With the development of deep learning technologies, certain scientists as mentioned in the paper [4–6] worked on deep learning methods and it resulted in better results as compared to traditional methods. The reason is that these models are data-driven and automatically compute and extract complex features thus resulting in good computational results at a fast pace of time.

However, the training of models on these methods requires a large amount of dataset to properly train a model. Deep learning models such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Deep Neural Network (DNN) are there but CNN performs best because the convolutional layers extract features based on image content and as both feature extractor and distributor.

3 Dataset Description

The entire experiments for forgery detection were performed on two types of datasets [1]—CASIA v1.0 and CASIA v2.0.

CASIA v1.0 contains 1711 JPEG format photos of size 384×256 pixels. These pictures of copy–move forgery types were obtained by pasting clipped image regions through resizing, rotation, or deformation. Out of 1711, 790 were original images and 921 were forged images.

CASIA v2.0 consists of 6479 images of sizes ranging from 240×160 pixels to 900×600 pixels in JPEG, BMP, and TIFF formats. Out of these, 5000 were original images and 1479 were forged. It consists of both copy–move forged and spliced images. Due to this reason, post-processing of boundary regions is more challenging.

4 Methodology

Deep learning models are used for tampering detection task to detect whether an image is original or forged. Detection task works on the coarse-grain (image) level of images. Features are extracted based on image content rather than image manipulation. Convolutional layers of CNN models are used as feature extractor. Deep learning models have the ability to learn both abstract and complex features from images during training. Data pre-processing is done followed by training the images on different CNN models and then evaluating the testing dataset.

4.1 Data Pre-processing

Error-level analysis is performed on the image datasets of both types of data. ELA is necessary when we do image forgery detection because it talks about edge distortion or denoising or manipulation in images if present. ELA determines regions in image which are at different compression levels. It highlights the difference in JPEG compression rate which thus determines the manipulation present. Image patches of size 128×128 are fed into the function of ELA converted images. Images are first converted into RGB and in JPEG format with specified given quality. After this, each time a pair of tampered images and their origin are subtracted to get the tampering of edge. Then these images are stored in an array and their corresponding labels of “0” for fake and “1” for real are also stored. Then, for training the model, the dataset is divided into a test data size of 0.2. So, 80% images are classified into training and 20% for testing of model. The pseudocode for ELA conversion is as written below:

- Images one at a time are passed to ELA conversion with a specified quality of 90.
- Inside this function, a copy of the original image is stored for calculating the difference in extreme values for tampering later.
- Images are then converted into RGB followed by storing into JPEG with given image quality.
- Then the tampered image obtained by calculating the difference between original image and modified image in the previous steps is stored.
- Then brightness enhancement followed by scaling is done on this image and this tampered image is stored back in an array X and its corresponding value (0 or 1) in Y.

4.2 Training Models

4.2.1 Proposed CNN Model

The architecture of the proposed CNN-based framework consists of two convolution layers, two max-pooling layers, and a fully connected layer with a two-way SoftMax activation function. The input is patches of size $128 \times 128 \times 3$ (128×128 patch sizes and 3 color channels). Both convolutional layers have 32 kernel filters each of size 5×5 . Rectified Linear Units (ReLU) to neurons are applied for the activation function. It makes them react to reasonable signals in the input. Now the third layer is a non-overlapping max-pooling layer with a filter size of 2×2 . Max-pooling helps in getting texture information discards 75% of the activation resulting in a good performance. Finally, the extracted features are made to pass through a fully connected layer with ReLU and SoftMax activation function through “dropout” which sets the neurons to zero in a fully connected layer with a probability of 0.25 and 0.5, respectively.

4.2.2 MobileNetV2

The architecture of MobileNetV2 is made of two types of blocks. One block consists of a stride of 1 and acts as a residual block, while the other one is a block of stride 2 and is used for down-sizing. There are 3×3 layers present in both blocks. The first layer is a 1×1 convolutional layer with a ReLU activation function. The second layer consists of depth-wise convolution and the third one is a 1×1 convolutional layer without the presence of any non-linearity. It is a lightweight model, widely used for image classification (Fig. 4).

4.2.3 ResNet-50

ResNet was arguably the most ground-breaking work in the deep learning community in recent years, which renders easier gradient flow. The key objective of ResNet is establishing a similar shortcut connection that leaps one or more layers. Now, the network will choose an uninterrupted path to the earliest layers in the network, making the gradient updates for those layers much easier. By using ResNet, we can now train 1001-layer deep ResNet to surpass its more superficial counterparts. Due to its effective outcomes, ResNet quickly evolved as one of the most favored architectures (Fig. 5).

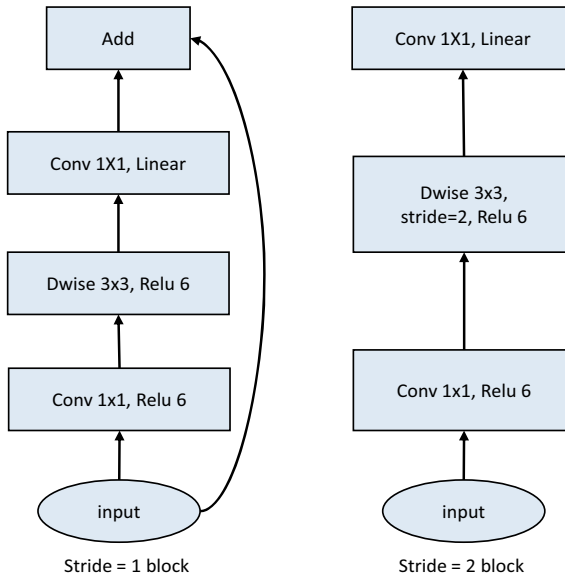


Fig. 4 MobileNetV2 architecture

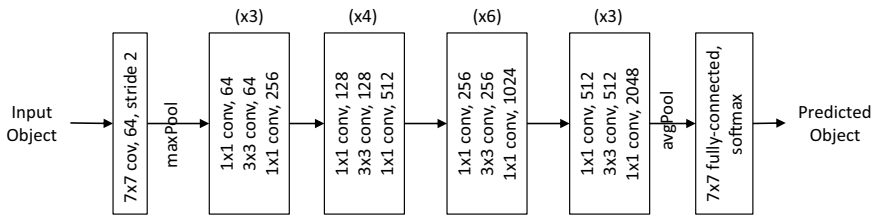


Fig. 5 ResNet-50 architecture

4.3 Workflow of the Proposed CNN Model

The proposed workflow is shown as a block diagram in Fig. 6.

5 Result and Analysis

5.1 Hyper-parameters

The proposed method tests were performed on google collaboratory on a Windows 10 server with an Intel Core i5 8th Gen processor and 8GB Ram. Images are taken as the input of patch sizes $128 \times 128 \times 3$ in which a test size of 0.2 is taken. For CASIA v1.0,

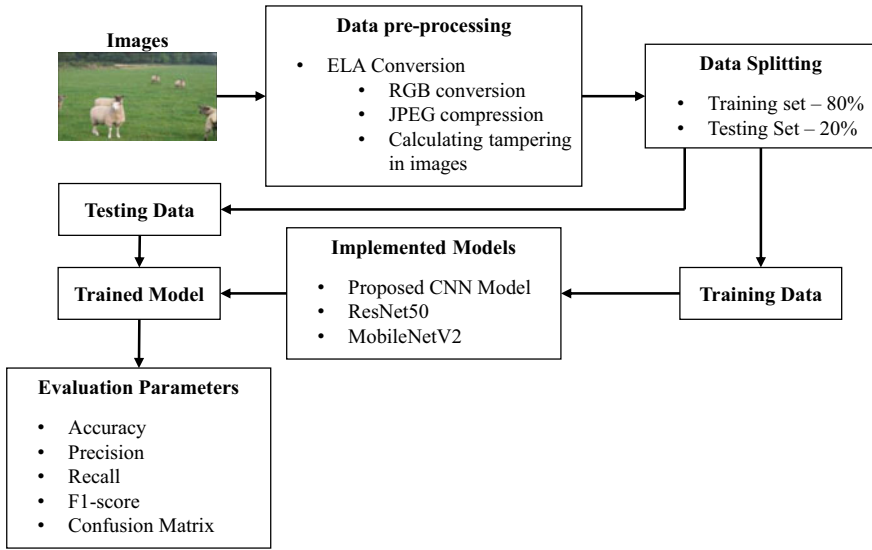


Fig. 6 Proposed workflow

1368 are available for training and 343 for testing. In CASIA v2.0, 5183 for training and 1296 for testing. Binary and Categorical cross-entropy functions are taken as per their suitability with the model. The optimizer taken is Adam with a learning rate of $1e-4$ in the case of proposed CNN and $1e-3$ in the case of MobileNetV2 and ResNet-50. The models are trained for 30 epochs with a batch size of 32.

5.2 Pseudocode

- Conversion to ELA images:
 - Convert to RGB images.
 - Store the images in JPEG format.
 - Storing the tampered images with brightness adjustment in the required form.
 - All original and fake images are stored together in one array.
- Reshape the images in $128 \times 128 \times 3$ format with corresponding values of 0 for fake and 1 for real.
- Divide the dataset into training and testing images with test size of 20% and random state as 5.
- Then train the dataset on the following models:

– ResNet-50:

Import the ResNet50 model from the TensorFlow library.

Then train the last layer using the transfer learning approach by providing input size of image as $128 \times 128 \times 3$ and weights from ImageNet.

Then on the output obtained from the base model, train it to create a convolutional layer of $1024 \times 3 \times 3$ with activation function as ReLu.

This output is fed for average pooling and flattening and then passing to three dense layer functions with activation function as ReLu and Softmax and L2 regularization.

Then Adam optimizer and binary cross-entropy loss function are used for compiling this Resnet model.

The model is then made to learn on training dataset with epochs as 30 and batch size as 32 with early stopping by monitoring validation accuracy.

– MobileNetV2:

Import the MobileNetV2 same as ResNet50 from the TensorFlow library.

Then train the model on already implemented MobileNetV2 in ImageNet by providing image size as $128 \times 128 \times 3$ and weights from ImageNet.

Then build the model with convolutional layers, max pooling-layers same as in case of ResNet-50.

Then use Adam as an optimizer and this time categorical cross-entropy as loss function.

Then similarly model is made to fit for training dataset with epoch size of 30 and batch size of 32.

– Proposed CNN:

This was the model made from scratch without using transfer learning.

Firstly, sequential model API is used for creating deep learning model to create and add layers to it.

Then two convolutional layers of size $5 \times 5 \times 32$ is added with activation function as ReLu on image size $128 \times 128 \times 3$.

After that, a max-pooling layer of size 2×2 is added.

Then layers are flattened followed by adding a dropout layer with probability rates of 0.25 and 0.5 and with ReLu and Softmax activation functions.

After this, the model is built.

On this built model, training dataset is made to fit by using Nadam as optimizer and categorical cross-entropy as loss function with 30 epochs and a batch size of 32.

Table 1 Accuracy comparison with existing models

Models	Datasets	
	CASIA v1.0	CASIA v2.0
[5]	81%	–
[4]	85.14%	–
[7]	–	76%
[3]	–	91.09%
ResNet-50	57%	75%
MobileNetV2	84%	92%
Proposed CNN	80%	95%

5.3 Evaluation Metrics

5.3.1 Accuracy

In the accuracy Table 1, we compared the results obtained by implementing different models with results obtained in the paper already implemented on the same datasets. Among the three models that we implemented, the best results were obtained in the case of convolutional neural network with an accuracy of 95%. However, these were still less than the papers having 10 layers (8 convolutional layers, 2 pooling layers, and 1 fully connected layer) on the same dataset. But, we obtained slightly higher results for MobileNetV2 as compared to what was obtained in the case of implementing stacked autoencoders.

Table 2 Precision, Recall, F1-score, and Specificity on CASIA v1.0

	Precision	Recall	F1-score	Specificity
ResNet-50	0.32	0.57	0.41	0.35
MobileNetV2	0.85	0.84	0.84	0.89
Proposed CNN	0.80	0.80	0.79	0.69

Table 3 Precision, Recall, F1-score, and Specificity on CASIA v2.0

	Precision	Recall	F1-score	Specificity
ResNet-50	0.57	0.72	0.62	0.42
MobileNetV2	0.92	0.92	0.92	0.94
Proposed CNN	0.94	0.95	0.94	0.97

5.3.2 Average Precision, Recall, F1-Score, and Specificity

Table 2 consists of performance metrics values for different models that we implemented. We can see from the above table that ResNet-50 performs worst among all three, but MobileNetV2 gives good values for all parameters with the highest specificity of 89%. In the second place, the proposed CNN also gives somewhat results. These values are less than those obtained from CASIA v2.0 in the below table due to the difference in dataset size.

Table 3 consists of performance metrics values for Casia v2.0. Here, ResNet-50 gives poor results but again proposed CNN gives good results with a specificity of 0.97 followed by MobileNetV2. Proposed CNN performs better than MobileNetV2 due to changes in size and type of dataset used.

5.4 Training and Validation Loss Curve

The training and validation loss curve in the case of CASIA v1.0 for both proposed CNN and MobileNetv2 models shows similar results. Initially, the cost function does not decrease much with an increase in the number of iterations resulting in underfitting but as the number of iterations/epochs increase the model starts showing good results (Fig. 7).

In the case of CASIA v2.0, the loss curves tend to move from underfitting toward good results in the case of MobilenetV2. However, for proposed CNN training and validation curves are improving but there is a gap between them which means they operate on datasets from different distributions (Fig. 8).

5.5 Confusion Matrix

In this confusion matrix for MobileNetV2, 0 is used for fake images and 1 for real images. Out of 194 fake images, 158 are predicted right for fake images, and for real images, 131 are predicted correctly out of 149 resulting in a specificity of 0.89. In this

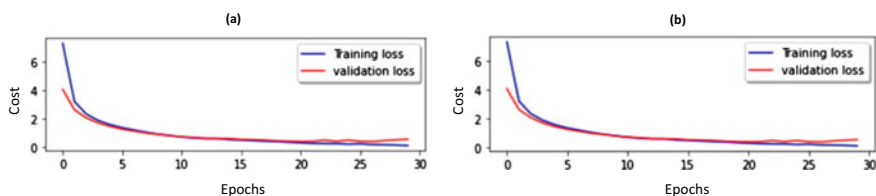


Fig. 7 Training and validation loss curve on CASIA v1.0 for **a** MobileNetV2 model, **b** Proposed CNN model

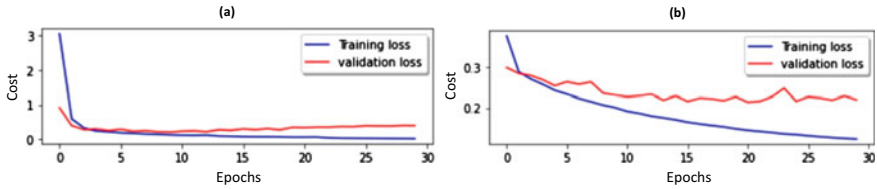


Fig. 8 Training and validation loss curve on CASIA v2.0 for **a** MobileNetV2 model, **b** Proposed CNN model

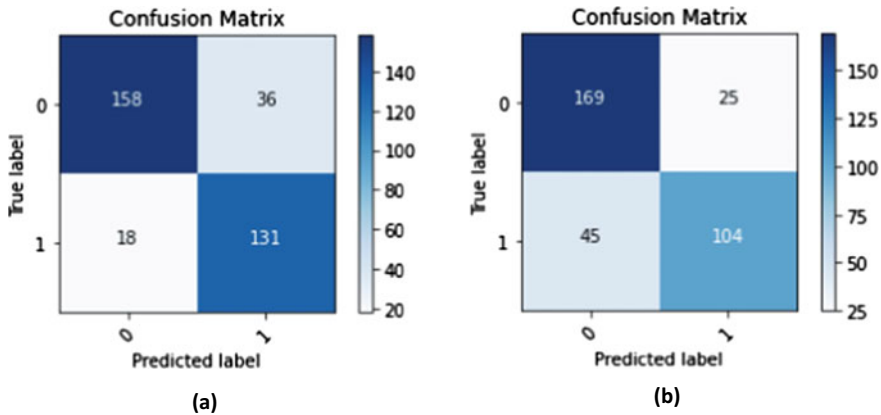


Fig. 9 Confusion matrix on CASIA v1.0 for **a** MobileNetV2 model, **b** Proposed CNN model

confusion matrix for the proposed CNN model, 0 is used for fake images and 1 for real images. Out of 194 fake images, 169 are actually predicted right for fake images, and for real images, 104 are predicted correctly out of 149 resulting in a specificity of 0.69 (Fig. 9). In this confusion matrix for MobileNetV2, 0 is used for fake images and 1 for real images. Out of 320 fake images, 265 are actually predicted right for fake images, and for real images, 925 are predicted correctly out of 976 resulting in a specificity of 0.94. In this confusion matrix for proposed CNN, 0 is used for fake images and 1 for real images. Out of 320 fake images, 273 are predicted right for fake images, and for real images, 952 are predicted correctly out of 976 resulting in a specificity of 0.97 (Fig. 10).

6 Conclusion and Future Scope

In this chapter, we compared the results of different deep learning models for the tampering detection task. Deep learning methods compute to give better results as compared to other state-of-the-art methods because these models extract both abstract and complex features. We compared the results of two datasets—CASIA

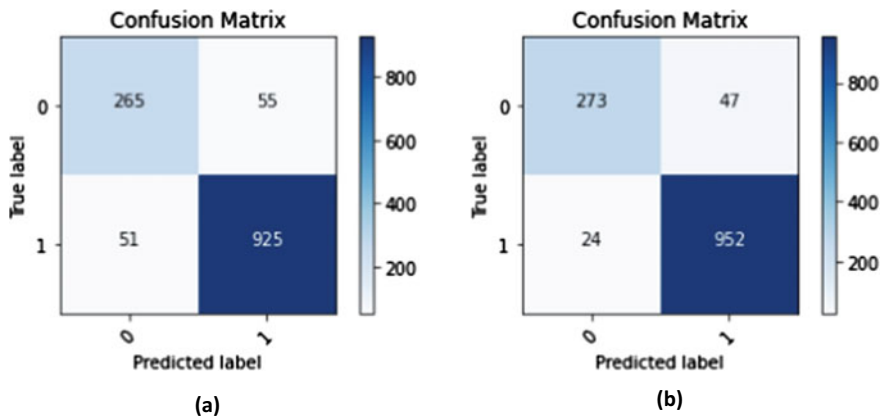


Fig. 10 Confusion matrix on CASIA v2.0 for **a** MobileNetV2 model, **b** Proposed CNN model

v1.0 and CASIA v2.0 to know about the fact that deep learning models give good computational results if a large number of datasets are present. We performed error-level analysis and denoising to get about edge tampering. We improved accuracy of MobileNetV2 on CASIA v1.0—84% (81% [5]) and CASIA v2.0—92% (>91.03% [3]). Proposed gives the best accuracy of 95%. Since we performed the detection-only task to determine whether an image is original or forged, which works on the coarse level (image level) of an image, we can also perform the localization task. Localization tasks work on fine level (pixel levels) to determine the type of forgery—copy-move or image-splicing, present in images.

References

1. Rao, Y., Ni, J.: A deep learning approach to detection of splicing and copy-move forgeries in images. 978-1-5090-1138-4/16/
2. Barad, Z.J., Goswami, M.M.: Image forgery detection using deep learning: a survey. 978-1-7281-5197-7/20/
3. Zhang, Y., Goh, J., Win, L.L., Thing, V.L.: Image region forgery detection: a deep learning approach. In: SG-CRC, pp. 1–11 (2016)
4. Chen, J., Kang, X., Liu, Y., Wang, Z.J.: Median filtering forensics based on convolutional neural networks. *IEEE Signal Process. Lett.* **22**(11), 1849–1853 (2015)
5. Bondi, S., Lameri, D., Güera, P., Bestagini, E., Delp, J., Tubaro, S.: Tampering detection and localization through clustering of camera-based cnn features. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp. 1855–1864. IEEE (2017)
6. Rao, Y., Ni, J.: A deep learning approach to detection of splicing and copy-move forgeries in images. In: 2016 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–6. IEEE (2016)
7. Bi, X., Wei, Y., Xiao, B., Li, B.: Rru-net: The ringed residual u-net for image splicing forgery detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 0–0 (2019)

8. Fridrich, A.J., Soukal, B.D., Lukáš, A.J.: Detection of copymove forgery in digital images. In: Proceedings of Digital Forensic Research Workshop, Citeseer (2003)
9. Johnson, M.K., Farid, H.: Exposing digital forgeries by detecting inconsistencies in lighting. In: Proceedings of the 7th Workshop on Multimedia and Security, pp. 1–10. ACM (2005)
10. Myna, A., Venkateshmurthy, M., Patil, C.: Detection of region duplication forgery in digital images using wavelets and log-polar mapping. In: International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007), vol. 3, pp. 371–377. IEEE (2007)
11. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)

Malicious URL Detection Using Machine Learning



Mayank Swarnkar, Neha Sharma, and Hiren Kumar Thakkar

Abstract In recent years cyberattacks have become destructive and targeted. With technological advancements, diverse threats are launching in a sophisticated way that targets people to defraud them. Many web applications have been struggling to improve the reliability and security of their platforms to protect users from fraud, revenue, or malware. These attacks use malicious uniform resource locators (URLs) to attack web users. These URLs host unwanted content in the form of junk emails, phishing, or unauthorized drive-by downloads. Unsuspecting people click these phishing URLs and become victims of unethical anonymous activities like identity theft (personal or financial details) and installation of viruses. Therefore, it is necessary to detect malicious URLs accurately for resolving security issues. Traditional protection method, such as blacklisting, remains a classical technique for the detection of malicious URLs due to its simplicity but cannot detect unknown malicious URLs; hence, machine learning approaches are being used for achieving better results. This chapter aims to provide a structural understanding of popular feature extraction techniques and machine learning algorithms.

Keywords Malicious URL · Feature extraction · Blacklisting · Machine learning

1 Introduction

With the rapid evolution and popularization of Internet technology, cybercrime is constantly increasing. According to the Microsoft Digital Defense Report released in 2021 [1], the industry experienced a significant increase in the phishing campaigns in 2020 that continued to grow in 2021. During 2020–2021, Microsoft saw an increase in phishing emails and voice phishing (or vishing). Furthermore, the report indicated

M. Swarnkar (✉) · N. Sharma
Indian Institute of Technology (IIT -BHU), Varanasi, India
e-mail: mSwarnkar@gmail.com

H. Kumar Thakkar
Pandit Deendayal Energy University, Gandhinagar, Gujarat, India
e-mail: hiren.thakkar@sot.pdpu.ac.in

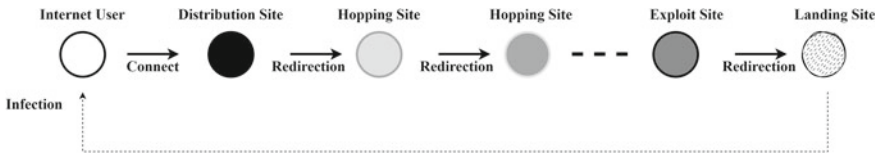


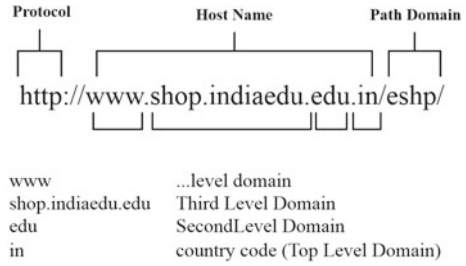
Fig. 1 Propagation of cyberattacks using malicious URLs

that phishing is responsible for almost 70% of data breaches. Since cybercriminals change tactics and advance their activity through new channels, the security sector needs an effective framework for the early detection of malicious websites. The attacker builds fraudulent websites that look similar to authentic websites or injects malicious code into the webpage by exploiting the web browser vulnerabilities. Initially, targeted users receive email or instant message that consists of malicious URLs for initiating a series of events. As users click on the link, they land on another website that executes malicious code. This code exploits the vulnerabilities of the web browser and installs malware on the system. Figure 1 shows how the attacker creates a malware dissemination network (MDN) consisting of various types of malicious URLs with an intent to cause damage to users with the installation of malware and to attempt various secondary cyberattacks (such as stealing personal information, disabling system services, and causing harm to other network hosts). MDN consists of several sites, namely a landing site (Internet user visits first) connected with a distribution site and is then redirected to several hopping or exploit sites for downloading malware on the system. The attacking posture of these sites is based on the strategy determined by the attacker [2]. As shown in Fig. 1, the malware infection process starts with the distribution site, which is the initially accessed site. Malicious agents in the MDN collaborate with one another to attract users to a landing page through several hopping sites or exploit sites. Landing pages continuously redirect the users until they arrive at the malware repository of the MDN. Once users reach the repository, malware is downloaded and installed on the system.

Such exploits are used by cybercriminals to take control of the system and steal valuable intellectual property, gain access to system files, or whatever else they want to seek. Implementation of such attacks needs a variety of techniques such as malware installation, phishing, social engineering, drive-by download, cross-site scripting (XSS), SQL injections, and many others. These attacks are exponentially increasing due to the limitations of traditional security management technologies, evolution of advanced digital technologies, and lacking of qualified cybersecurity experts. Due to the present vulnerabilities, it is difficult to design a robust and resilient system to detect security breaches. Most of these attacking techniques use compromised URLs.

URL stands for Uniform Resource Locator that uniquely identifies a resource on the Internet. To view a website or webpage in a browser, unique address is needed on the World Wide Web and this unique address is URL. It tells a web browser where to search for a resource on the web. There are two main components of a URL: protocol identifier and resource identifier. For example: in <https://www.shop.com>, URL, HTTPS, FTP, DNS, etc., are protocol identifiers, www.shop.com

Fig. 2 URL structure



is the resource, and location of a website or webpage on the Internet is the resource identifier [3]. The anatomy of URL is shown in Fig. 2. The URL consists of many components such as protocol, hostname, and path.

Sharing data with URLs is convenient through social media posts, websites, emails, and phone messages. This simplicity and ubiquity allow attackers to create deceptive URLs that cause billions of losses every year. The development of time-efficient defense mechanisms is needed to counter diverse types of cyber-security threats. In light of the above concerns, researchers and practitioners have worked on several approaches for the detection of malicious URLs. The most commonly used method developed by network security teams or antivirus groups to detect malicious URLs is the blacklist method. This method maintains a database consisting of a list of blocked URLs that are often collected through crowdsourcing systems (e.g., Phish-Tank [4]). The blacklist-based technique is fast because finding a URL in a database is easy, simple, and less burdensome. Additionally, the technique seems to have a significant minimum false positive rate. The major shortcoming of this method is that it is difficult to maintain an updated database of malicious URLs since new URLs are created every day worldwide.

Attackers constantly search for ways to bypass blacklists and compromise sensitive information of users by modifying the URL to appear legitimate via obfuscation. There are four types of obfuscation techniques identified by [5]: obfuscating the Host with an IP, obfuscating the Host with another domain, obfuscating the host with large hostnames, and misspelling. All these four types intend to hide the impure ideas of the website by masking the malicious URL. Recently, hiding the malicious URL behind a short URL obfuscation technique has become popular with the widespread usage of URL shortening services. Thus, blacklisting methods cannot deal with unknown threats because these threats continue to evolve with new patterns and variants.

To address these limitations, researchers have explored applied Machine Learning (ML) techniques for malicious URL detection [6]. ML approaches collect some data (training set) for making predictions and decisions. In the case of malicious URL detection, training data is built by taking a large collection of URLs, and by extracting statistical properties of the URL string, a prediction function is calculated to classify a URL as malicious or benign. The primary requirement of ML models is the presence of training data. This step involves the collection of a large number of

URLs from various sources. After collecting training data, the next step is to extract the appropriate set of features such that they can be used by classic ML algorithms to perform a specific task. After the successful extraction of features, they have to be converted into a numerical vector, such that they can be fitted to the ML method for model training. Therefore, the good quality feature representation of the URLs yields a better malicious URL predictive model. The next step is the actual training of the prediction model that combines training data and feature representation. Many ML techniques can be directly applied to the training data.

ML approaches have shown good performances in the detection and classification of malicious URLs. However, there are some limitations associated with these approaches. One of the serious drawbacks faced by the traditional models is the high training time. This issue occurs when the size of the dataset is in the order of millions (or billions). Another major challenge is the sparse representations produced by bag-of-words (BoW). This feature representation indicates whether a particular word in a URL string is present or not. In this way, every word that appears in a URL becomes a feature. This representation will lead to a high dimension of feature vector and data sparsity. Accordingly, this sparsity can be further exploited to develop an efficient learning method.

Figure 3 demonstrates a general workflow for detecting malicious URLs using ML. Capturing domain knowledge and heuristics is the first part of feature representation, and using data-driven optimization approaches for training the classification model makes the second part.

Chapter Structure

The chapter is divided into the following sections. Section 2 provides the overview of previous work done by authors. Section 3 provides the broad categories of strategies used for detecting malicious URLs. Section 4 provides description about datasets. Section 5 explores various feature extraction techniques. Section 6 discusses ML approaches for malicious URL detection. Section 7 provides practical issues and open problems. Section 8 delivers conclusion.

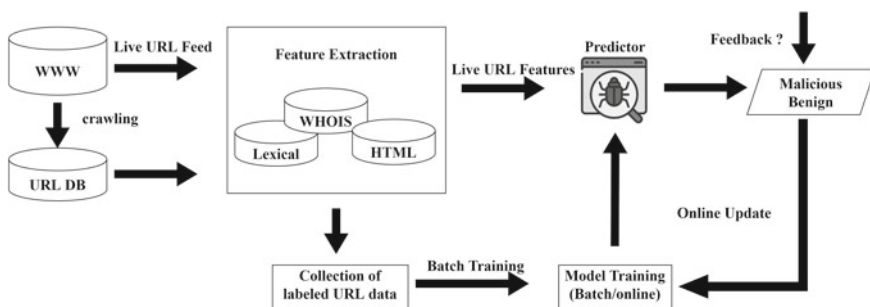


Fig. 3 Basic framework for malicious URL detection using machine learning

2 Related Work

In recent years, researchers have made significant efforts to detect malicious URLs due to the occurrence of web-based attacks. Various feature extraction techniques have been extensively explored in this area, including URL-based lexical features, DNS-based features, and Webpage content-based features.

Many approaches are used for malicious URL detection and they can be majorly classified into two types: ML approaches and non-ML approaches. As per the mentioned criterion, blacklisting is a type of non-ML method to identify malicious URLs. In this category, several researchers have proposed the implementation of blacklisting based on different techniques such as real-time blackhole lists [7], tracking the top-level domain names [8], or reputation-based [9]. The blacklisting approach is not suitable for new threats and is expected to give false negative results when new malicious URLs appear. In addition, the domain generation algorithm (DGA) can be used by attackers for generating new malicious URLs to evade domain name blacklists. To avoid such threats, it is imperative to design adaptive and intelligent detection techniques, that can identify URLs from the blacklists and the unknown malicious URLs by applying ML algorithms. A summary of studies is presented in Table 1 and identifies the type of URL features and ML classifiers for detecting malicious URLs. The authors of [10] explored the effectiveness of support vector machine (SVM) and random forest (RF) for malicious URL detection. Many studies, such as [11–14], focused on the detection of phishing web pages since the majority of attacks are done through phishing techniques. In this context, [11, 12] analyzed the use of only lexical-based URLs to identify malicious links, while [15] used Javascript client code based feature extraction technique to achieve better detection rate.

Aung et al. [16] provides an overview of recent phishing URL detection studies by considering attack types rather than detection methods. The authors made significant contributions between 2014 and 2019 that focused on the following three elements: type of algorithms and their uses, performance metrics, benefits, and drawbacks in the study. Aung and Yamana [13] analyzed the dataset used by researchers about malicious input for feature extraction and training of models [13]. The analysis showed that most studies use imbalanced datasets as the number of phishing sites cannot be compared with that of legitimate URLs.

Other research studies combined ML algorithms with other techniques for producing better results. In this line of research, [17] used a classification technique based on the association (CBA) algorithm to reduce the false-negative rate. The authors suggested a hybrid approach that combines lexical-based features and content-based features. This approach resulted in 7.57% of false-negative rate which is much better than 1.35% of false-negative rate given by approaches that used only lexical-based features [12]. The complexity of the proposed model has not been investigated by the authors in their study but other works indicated that content-based features take more time and result in delays as compared to lexical-based features. The reasons behind these limitations of content-based features are that lot of time is being consumed while reading the source page of a website, searching for suspicious Javascript functions and iFrames, and parsing of DOM model.

Table 1 Summary of machine learning approach in the literature

Year	Refs.	Features	ML algorithm	Description
2021	[17]	Lexical -Content based	Classification Based on Associations (CBA)	Benign page versus Malicious page
2021	[19]	Lexical Host-Based	Gated Recurrent Unit (GRU) Convolutional Neural Network (CNN) Long Short-Term Memory (LSTM)	Detect malicious URLs, file paths and registry keys, and Social media text classification
2021	[3]	Lexical Host-Based	Logistic Regression (LR) Stochastic Gradient Descent (SGD) Random Forest (RF) Support Vector Machine (SVM) Naive Bayes (NB) K-Nearest Neighbor (kNN) Decision Tree(DT)	Normal page versus Malicious page
2020	[24]	Lexical Host-Based Reputation- Based	K-Nearest Neighbor (kNN) Lagrangian Support Vector Machine (L-SVM) Linear Discriminant Analysis (LDA) Logistic Regression (LR)	Normal page versus Malicious page
2020	[25]	Lexical Host-Based	Convolutional Neural Network (CNN) Logistic Regression (LR) Naive Bayes (NB)	Normal page versus Malicious page
2020	[23]	Lexical	Random Forest (RF) Decision Tree (DT) K-Nearest Neighbor (kNN) Support Vector Machine (SVM) Logistic Regression (LR) Linear Discriminant Analysis (LDA) AdaBoost Naive Bayes (NB) Fast.ai Keras-TensorFlow Naive Bayes	Normal page versus Malicious page
2018	[18]	Host-Based	Decision Tree (DT) Gradient-boosted tree (GBT) Lagrangian Support Vector Machine (L-SVM) Naive Bayes (NB) Random Forest (RF)	Benign page versus Malicious page
2016	[28]	Lexical Host-Based	C4.5 Decision Tree Classifier Decision Tree (DT)	Benign versus Malicious
2013	[29]	Lexical Link Popularity	Random Forest (RF) Naive Bayes (NB) Logistic Regression (LR) J48 Decision Tree Classifier Support Vector Machine (SVM)	Normal page versus Malicious page
2013	[30]	Lexical Host-Based	Support Vector Machine (SVM)	Benign versus Phishing
2011	[27]	Lexical Link Popularity Host-Based	support Vector Machine (SVM) RANdom k-labELsets (RAKEL) K-Nearest Neighbor (kNN)	Classify attack types by URL
2011	[31]	Lexical Webpage Content-Based Host-Based	Random Forest (RF) Naive Bayes (NB) Logistic Regression (LR) J48 Decision Tree Classifier	Normal page versus Malicious page

Janet et al. [3] applied a range of ML algorithms to the dataset containing malicious URLs to investigate the prediction accuracy. They extracted features, namely domain, sub-domain, and suffixes, to distinguish malicious websites from benign ones. Their dataset was made of a large number of URLs collected from varied sources including malware, hidden frauds, and block-listed URLs. All models performed well with a high prediction accuracy; however, the highest F1-score and accuracy were attained by the random forest algorithm. Based on a similar approach, [18] analyzed host-based features such as domain details, IP addresses, and port numbers to detect malicious web pages. In their experiments, different classification algorithms were evaluated for their effectiveness. However, gradient boosted algorithm (the tree-based algorithm) achieved an overall accuracy of 96.9%. The authors in [19] also run a comparative analysis, where they evaluated seven detection models based on various ML algorithms taken during their literature review study. Among the tested algorithms, the CMU [20] and Endgame [21] models based on the bidirectional gated recurrent unit (BGRU) and long short-term memory (LSTM) yielded the highest accuracy.

Deep learning algorithms are a part of a broader family of ML methods. Deep learning algorithms are efficient in solving many real-life problems, due to their ability to learn the patterns exhibited by the targeted phenomena effectively. However, deep learning algorithms make a lot of computations while dealing with higher number of variables that have wide value ranges [22]. Johnson et al. [23] compared random forest, CART, and kNN with fast.ai and TensorFlow across CPU, GPU, and TPU architectures and found that random forest performed the best with an accuracy of 98.68%. The authors in [24] improved the performance of several ML algorithms such as k-NN, SVM, and Multi-layer Perceptron (MLP) by applying linear and nonlinear transformation. The authors evaluated training and testing subsets in terms of their time efficiency. The authors also used a large feature set (64 inputs in total) that caused long computational time for the algorithms. Another study [25] used an open-source dataset of malicious URLs [26] (also used by [3] to evaluate the performance of classifiers). The authors established that the Naive Bayes algorithm performed better than logistic regression and convolutional neural network (CNN), with an accuracy of 86.25%. The authors in [27] advised that understanding the attack type of malicious URL is useful for the user to respond properly.

3 Overview of Principles of Detecting Malicious URLs

According to the fundamental principles, two approaches have been attempted to address the problem of Malicious URL Detection. First, blacklisting or heuristics, and second ML approaches.

3.1 Blacklisting or Heuristic Approaches

The blacklisting approach is a trivial technique used for the detection of malicious URLs. This approach maintains a database that contains a list of malicious URLs.

Whenever a user accesses a website or webpage through a new URL, its presence is checked in the database. If the URL is found in the database, it is considered malicious and a warning is generated. If the URL is absent in the database, it is assumed as benign. The blacklisting method cannot detect new threats [32]. This is a matter of concern as attackers can develop new algorithms to bypass all blacklists. Despite several limitations, blacklisting approaches are still considered by many anti-virus systems because of their simplicity and efficiency.

Heuristic approaches are an extended version of blacklisting methods that are based on the concept of creating a “blacklist of signatures”. Malicious threats are identified and adding their signature to a database is a primary requirement. Intrusion detection system monitors and searches network traffic for such signatures and raises an alarm if any match is found. These methods give better results than blacklisting approaches, as they are capable of identifying new threats in the URL string. However, such methods are limited to some common threats and cannot be generalized to all types of (novel) attacks. Moreover, it is easier to bypass attacks using obfuscation techniques.

3.2 Machine Learning Approaches

These approaches analyze the information of a URL and its corresponding websites or webpages by extracting good feature representations of a URL. Training data (containing both malicious and benign URLs) is considered for extracting and analysis purposes. Further, the training data is used on which prediction model is trained. Two types of features can be used—static features and dynamic features. In static analysis, the webpage is analyzed based on the information available without executing the URL (i.e., executing JavaScript, or other code) [33]. Static features include lexical features of a URL string. As there is no requirement for execution in static approaches thus are safer than dynamic approaches. The distribution of features is different for malicious and benign URLs in the case of dynamic approaches. This prediction information can be used to build the prediction model for making predictions on new URLs. Due to the ability to execute in a safer environment and generalize to all types of threats, static analysis techniques have been combined with ML techniques for achieving better experimental results. In this chapter, static analysis techniques are explored where ML has found tremendous success. Dynamic analysis techniques include the monitoring of system behavior for anomalous activity that involves inherent risks and is difficult to implement and generalize.

4 Datasets

For better detection of malicious URLs, Canadian Institute for Cybersecurity Datasets is used by universities and independent research scholars as they maintain an interactive map indicating datasets downloaded by country. Details of datasets that include

Table 2 Dataset information

S.no	Dataset	Source	Count of URLs	References
1	Benign	Alexa top websites	35,300	[34]
2	Spam	WEBSPAM-UK2007 Dataset	12,000	[35]
3	Phishing	OpenPhish Repository	10,000	[36]
4	Malware	DNS-BH Project	11,500	[37]

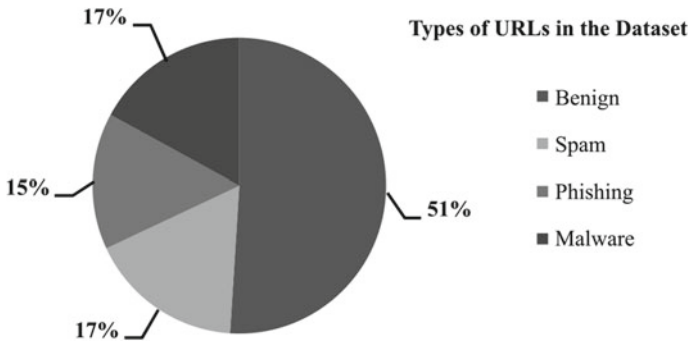


Fig. 4 Types of URLs in the dataset

four types of URLs, benign, spam, phishing, and malware, are given in Table 2, and percentage-wise distribution of URLs is presented in Fig. 4.

5 Feature Extraction

Several types of feature extraction techniques have been proposed by researchers to detect malicious URLs. These techniques include URL-based lexical features, DNS-based features, and webpage content-based features.

5.1 URL-Based Lexical Features

Lexical features are the textual properties extracted from the URL string. Lexical features can identify the malicious nature of the URLs by seeing how the URL "looks". For example, the presence of the token ".com" in the URL "www.shop.com" is not unusual. However, the presence of ".com" in the URL "www.shop.com.phishy.biz" seems to be an attack done by a cybercriminal to tamper with the contents of a legitimate website. Statistical properties of the URL string are the most commonly used lexical features [38] that infer patterns in malicious URLs for classification purposes. Total 63 features [39] are described in this chapter. Among these features, 47 are described in Table 3 and 16 are described in Table 4.

Table 3 URL-based lexical features

S.no	Presence of features in URL string	Type
U1	URL Length	Numeric
U2	IP address in Hostname	Numeric
U3	Query Length	Numeric
U4	Tokens	Numeric
U5	Dots (.) characters	Numeric
U6	Hyphen (-) sign characters	Numeric
U7	Underscore (_) sign characters	Numeric
U8	Equal (=) sign characters	Numeric
U9	Forward slash (/) sign characters	Numeric
U10	Question Mark sign (?)characters	Numeric
U11	"secure" word	Binary
U12	"account" word	Binary
U13	"webscr" word	Binary
U14	"login" word	Binary
U15	"ebayisapi" word	Binary
U16	"signin" word	Binary
U17	"banking" word	Binary
U18	"confirm" word	Binary
U19	"blog" word	Binary
U20	"logon" word	Binary
U21	"signon" word	Binary
U22	"login.asp" word	Binary
U23	"login.php" word	Binary
U24	"login.htm" word	Binary
U25	".exe" word	Binary
U26	".zip" word	Binary
U27	".rar" word	Binary
U28	".jpg" word	Binary
U29	".gif" word	Binary
U30	"viewer.php" word	Binary
U31	"link=" word	Binary
U32	"getImage.asp" word	Binary
U33	"plugins" word	Binary
U34	"paypal" word	Binary
U35	"order" word	Binary
U36	"dbsys.php" word	Binary
U37	"config.bin" word	Binary
U38	"download.php" word	Binary
U39	".js" word	Binary

(Continued)

Table 3 (continued)

S.no	Features	Type
U40	"payment" word	Binary
U41	"files" word	Binary
U42	"css" word	Binary
U43	"shopping" word	Binary
U44	"mail.php" word	Binary
U45	".jar" word	Binary
U46	".swf" word	Binary
U47	".cgi" word	Binary

Table 4 URL-based lexical features

S.no	Features	Type
U1	Semicolon (;) sign characters	Numeric
U2	Open Parenthesis (() sign characters	Numeric
U3	Close Parenthesis()) sign characters	Numeric
U4	Mod Sign (%) sign characters	Numeric
U5	Ampersand Sign (&) sign characters	Numeric
U6	At the Rate Sign (@) sign characters	Numeric
U7	Digits	Numeric
U8	Entropy of URL string	Real
U9	".php" word	Binary
U10	"abuse" word	Binary
U11	"admin" word	Binary
U12	".bin" word	Binary
U13	URL without "www"	Binary
U14	"personal" word	Binary
U15	"update" word	Binary
U16	"verification" word	Binary

5.2 DNS-Based Features

A domain name system is an important component of the internet that converts domain names into IP addresses (e.g., shop.com to 173.216.16.174). DNS resolution takes place before visiting any malicious website. Thus, identifying malicious domains can be the first line of defense. Total 18 features [39] are described in this chapter. Among these features, 7 are described in Table 5 and 11 are described in Table 6.

Table 5 DNS-based features

S.no	Features	Type
D1	Length	Numeric
D2	Presence of IP address	Binary
D3	Is Domain TLD?	Binary
D4	Count of Sub-Domains	Numeric
D5	Count of Yahoo search results	Numeric
D6	Count of Bing search results	Numeric
D7	Alexa ranking	Numeric

Table 6 DNS-based features

S.no	Features	Type
D1	Is Domain Name Valid?	Binary
D2	Entropy of Domain Name string	Real
D3	Count of tokens in Domain Name	Numeric
D4	Length of Longest Domain Token	Numeric
D5	Entropy of Longest Domain Token	Real
D6	Average length of domain Token	Real
D7	Count of Tokens in Path	Numeric
D8	Length of Longest Path Token	Numeric
D9	Average length of Path Token	Real
D10	Domain Name contains suspicious https?	Binary
D11	Domain Name contains suspicious www?	Binary

5.3 Webpage Content-Based Features

These features provide more information about a particular webpage for building better prediction models. Web page-based content features are "heavyweight", as a lot of content is being extracted from a webpage for early detection of threats. Hackers inject malicious code into a webpage for triggering various types of attacks. Two main types of content-based features, namely HTML features and Javascript features, are taken into consideration [40]. Total 34 [39] features are described in this chapter. Among these features, 19 are described in Table 7 and 15 are described in Table 8.

6 Machine Learning Algorithms for Malicious URL Detection

In the last decade, an abundant amount of work has been done to use ML techniques to solve the problem of malicious URL detection. Table 9 presents a summary of frequently used ML models for malicious URL detection. In this table, we have mentioned each classification model along with brief description, advantages, and limitations.

Table 7 Webpage content-based features

S.no	Features	Type
W1	Count of Blank Lines	Numeric
W2	Count of Blank Spaces	Numeric
W3	Count of Words	Numeric
W4	Average Length of Words	Real
W5	Count of iFRames	Numeric
W6	Count of JavaScript	Numeric
W7	Count of embed Tag	Numeric
W8	Count of object Tag	Numeric
W9	Count of meta Tag	Numeric
W10	Count of div Tag	Numeric
W11	Count of body Tag	Numeric
W12	Count of form Tag	Numeric
W13	Is Title Tag present?	Binary
W14	Count of anchor Tag	Numeric
W15	Count of Hidden elements	Numeric
W16	Count of External JavaScript Files	Numeric
W17	Count of Links	Numeric
W18	Count of Internal Links	Numeric
W19	Count of External Links	Numeric

Table 8 Webpage content-based features

S.no	Features	Type
W1	Count of image Tag	Numeric
W2	Count of span Tag	Numeric
W3	Count of input Tag	Numeric
W4	Count of CSS styles	Numeric
W5	Count of audio Tag	Numeric
W6	Count of applet Tag	Numeric
W7	The size of Webpage	Numeric
W8	Credit card number word	Binary
W9	Log word	Binary
W10	Pay word	Binary
W11	Free word	Binary
W12	Access word	Binary
W13	Bonus word	Binary
W14	Click word	Binary
W15	Entropy of webpage	Real

Table 9 Summary of machine learning techniques

Categories	Technique	Description	Advantages	Limitation
Supervised learning	SVM	SVM has two labeled classes and classifiers (hyperplane with N-dimensions) SVM is suitable for linear binary classification that classifies data points into two sections and predicts new data points belonging to each section	Efficient in handling large feature space Flexibility in choosing the similarity function	Sensitive to noise or outliers Doesn't provide posterior probability
	Naive Bayes	Naive Bayes is a probabilistic classifier used for classification Naive Bayes is based on Bayes theorem—a conditional probability theorem with robust independent features	Works well with a large number of features Can handle large datasets	Make unrealistic assumptions of completely independent features Assigns zero probability if some category in test dataset is not observed in the training dataset
	Random Forest	Composed of many decision trees for the purpose of classification and regression Every decision tree gives an output and averaging is done of outputs produced by every tree to reduce overfitting	Works well with high-dimensional data Works well with data containing “noisy” variables	Computational cost is high for large ensembles Prediction generator is slow
	ANN	ANN consists of a hidden layer with input neurons and output layer ANN works in two stages: feedforward and backward	Parallel processing capability Ability to work with incomplete knowledge	Duration of the network is unknown Unexplained behavior of the network

(Continued)

Table 9 (continued)

Categories	Technique	Description	Advantages	Limitation
	Decision tree	Works on an if-then rule to find the best immediate node and the process continues till the predicted class is obtained	Provides the best possible solution and Classification and interpretation tasks are easy	Cannot handle data in excessive amount Sensitive to biasness
Unsupervised learning	K-Mean	k-NN finds the distance between the target and its nearest neighbors for making predictions	Clustering is easy and fast Always provide a solution	Sensitive to initialization Sensitive to outliers
	DBN	DBN is made up of several middle layers of restricted Boltzmann machine (RBM) that are organized in a greedy fashion Every middle layer acts as both input and output layer except the first and last layer	Provides the best performance results even when the amount of data is huge Can deal with new problems in the future	Training of DBN is expensive due to its complex data model Hundreds of machines are needed

7 Practical Issues and Open Problems

Despite many flexible and resilient ML techniques to detect malicious URLs, there are still many practical issues and open problems that need to be addressed:

High Volume and High Velocity: The dataset of URLs is exponentially growing with time resulting in high-volume and high-velocity data. Thus, training a ML model on all the URL data is almost impossible. An efficient way of sampling URL training data will be always an open question for ML researchers and cybersecurity experts.

Difficulty in Acquiring Labels: Most ML techniques require labeled training data for the detection of malicious URLs that can be obtained either by acquiring from black-lists/whitelists or by asking human experts to review and label them. Unfortunately, the amount of such labeled data is less compared to the volume of available URLs on the web. Thus, working with a limited amount of data or resolving the difficulty of acquiring labeled data is an open research area.

Difficulty in Collecting Features: Feature collection is a crucial step in the pipeline of malicious URL detection using ML because some features could be costly, some might be missing, lost, or cannot be obtained due to varied reasons. In addition, many malicious URLs are not always alive, and extracting their features may not be possible after their expiration.

Feature Representation: High-dimensional feature space is another key challenge. This is the underlying problem of the predictive models because feature space increases with time when new URLs are added to the training data. Thus, there is a need of developing ML algorithms that can adapt to dynamically changing feature spaces.

Concept Drifting and Emerging Challenges: Concept drift is a challenging problem in the field of ML because the statistical properties of malicious URLs may change due to the evolution of new types of threats and attacks. Besides, the popularity of URL shortening services has become a recent challenge. In this service, a long URL is converted into a short, case-sensitive alphanumeric code. Such URL shortening services are third-party websites that offer an excellent way for cybercriminals or hackers to hide their malicious URLs behind these tiny URLs. Designing an effective system that can quickly adapt to resolving new challenges will be a long-term research.

Interpretability of Models: An important research direction is to analyze the patterns in the URLs for understanding the nature of URLs whether they are benign or malicious. Acquiring a deeper understanding of the malicious behavior of URLs can have many applications in designing resilient modern security systems.

Adversarial Attacks: As ML researchers and cybersecurity professionals are developing better systems to detect malicious URLs, it will not be surprising to expect the launching of creative attacks by cybercriminals to bypass these systems. A systematic investigation of current models can help to provide future research directions in this area.

8 Conclusion

In this chapter, we discussed various feature extraction and ML techniques for malicious URL detection. We also provided a comprehensive study of three popular feature extraction techniques, namely URL-based lexical features, DNS-based features, and webpage content-based features. We also provided descriptions and limitations of five supervised and two unsupervised learning algorithms for malicious URL detection. For the better detection and categorization of malicious URLs according to the attack associated with them, we provided information on datasets consisting of four types of URLs: benign, spam, phishing, and malware. This chapter also provides an

overview of blacklisting methods and ML approaches to have a clear understanding of the advantages of ML approaches over blacklisting methods. Finally, we indicated some issues that are currently present in the ML domain and future research ideas in the domain of malicious URL detection using ML. In future work, we are aiming to provide more effective feature extraction techniques and deep learning algorithms for the detection of malicious URLs.

References

1. <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RWMFli?id=101738>
2. Kim, D.: Potential risk analysis method for malware distribution networks. *IEEE Access* **7**, 185157–185167 (2019)
3. Janet, B., Kumar, R.J.A., et al.: Malicious url detection: a comparative study. In: *Proceedings of 2nd International Conference on Artificial Intelligence and Smart Systems (ICAIS'21)*, pp. 1147–1151 (2021)
4. OpenDNS, L.: Phishtank: An anti-phishing site (2016). <https://www.phishtank.com>
5. Garera, S., Provos, N., Chew, M., Rubin, A.D.: A framework for detection and measurement of phishing attacks. In: *Proceedings of 14th ACM Workshop on Recurring Malcode (WORM '07)*, pp. 1–8 (2007)
6. Ma, J., Saul, L.K., Savage, S., Voelker, G.M.: Learning to detect malicious urls. *ACM Trans. Intell. Syst. Technol.* **2**, 1–24 (2011)
7. Felegyhazi, M., Kreibich, C., Paxson, V.: On the potential of proactive domain blacklisting. *Large-Scale Exploits Emergent Threats* **10**, 6–6 (2010)
8. Sinha, S., Bailey, M., Jahanian, F.: Shades of grey: on the effectiveness of reputation-based “blacklists”. In: *Proceedings of 3rd International Conference on Malicious and Unwanted Software (MALWARE'08)*, pp. 57–64 (2008)
9. Lu, G., Sadagopan, N., Krishnamachari, B., Goel, A.: Delay efficient sleep scheduling in wireless sensor networks. In: *Proceedings of 24th Annual Joint Conference of The IEEE Computer and Communications Societies (INFOCOM'05)*, vol. 4, pp. 2470–2481 (2005)
10. Do Xuan, C., Nguyen, H.D., Nikolaevich, T.V., et al.: Malicious url detection based on machine learning. *Int. J. Adv. Comput. Sci. Appl.* **11** (2020)
11. Tsolas, I.E., Charles, V.: Incorporating risk into bank efficiency: a satisficing idea approach to assess the greek banking crisis. *Expert Syst. Appl.* **42**, 3491–3500 (2015)
12. Jeeva, S.C., Rajsingh, E.B.: Intelligent phishing url detection using association rule mining. *Human-Centric Comput. Inf. Sci.* **6**, 1–19 (2016)
13. Aung, E.S., Yamana, H.: Url-based phishing detection using the entropy of non-alphanumeric characters. In: *Proceedings of 21st International Conference on Information Integration and Web-Based Applications and Services (IIWAS'19)*, pp. 385–392 (2019)
14. Tung, S.P., Wong, K.Y., Kuzminykh, I., Bakhshi, T., Ghita, B.: Using a machine learning model for malicious url type detection. In: *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, pp. 493–505 (2021)
15. Dong, H., Li, T., Ding, R., Sun, J.: A novel hybrid genetic algorithm with granular information for feature selection and optimization. *Appl. Soft Comput.* **65**, 33–46 (2018)
16. Aung, E.S., Zan, C.T., Yamana, H.: A survey of url-based phishing detection. In: *Proceedings of 11th Forum on Data Engineering and Information Management (DEIM'11)*, pp. G2–3 (2019)
17. Kumi, S., Lim, C., Lee, S.-G.: Malicious url detection based on associative classification. *Entropy* **23**, 182 (2021)
18. Tan, G., Zhang, P., Liu, Q., Liu, X., Zhu, C., Dou, F.: Adaptive malicious url detection: learning in the presence of concept drifts. In: *Proceedings of 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. 737–743 (2018)

19. Srinivasan, S., Vinayakumar, R., Arunachalam, A., Alazab, M., Soman, K.: Durlid: malicious url detection using deep learning-based character level representations. In: *Malware Analysis using Artificial Intelligence and Deep Learning*, pp. 535–554 (2021)
20. Dhingra, B., Zhou, Z., Fitzpatrick, D., Muehl, M., Cohen, W.: Tweet2vec: character-based distributed representations for social media. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL'16)*, pp. 269–274 (2016)
21. Anderson, H.S., Woodbridge, J., Filar, B.: Deepdga: adversarially-tuned domain generation and detection. In: *Proceedings of 9th ACM Workshop on Artificial Intelligence and Security (AISEC'16)*, pp. 13–21 (2016)
22. Kuzminykh, I., Shevchuk, D., Shiaeles, S., Ghita, B.: Audio interval retrieval using convolutional neural networks. In: *Internet Of Things. Smart Spaces, And Next Generation Networks And Systems*, pp. 229–240 (2020)
23. Johnson, C., Khadka, B., Basnet, R.B., Doleck, T.: Towards detecting and classifying malicious urls using deep learning. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.* **11**, 31–48 (2020)
24. Li, T., Kou, G., Peng, Y.: Improving malicious urls detection via feature engineering: linear and nonlinear space transformation methods. *Inf. Syst.* **91**, 101494 (2020)
25. Vundavalli, V., Barsha, F., Masum, M., Shahriar, H., Haddad, H.: Malicious url detection using supervised machine learning techniques. In: *Proceedings of 13th International Conference on Security of Information and Networks (SIN'13)*, pp. 1–6 (2020)
26. Urcuqui, C.: Malicious and Benign Websites Dataset. Accessed on: March, vol. 3 (2021)
27. Choi, H., Zhu, B.B., Lee, H.: Detecting malicious web links and identifying their attack types. In: *Proceedings of 2nd USENIX Conference on Web Application Development (WEBAPPS'11)* (2011)
28. Mašetić, Z., Subasi, A., Azemovic, J.: Malicious web sites detection using c4. 5 decision tree. *Southeast Eur. J. Soft Comput.* **5**(1) (2016)
29. Eshete, B., Villafiorita, A., Weldemariam, K., Zulkernine, M.: Einspect: evolution-guided analysis and detection of malicious web pages. In: *Proceedings of 37th IEEE Annual Computer Software and Applications Conference (COMPSAC'13)*, pp. 375–380 (2013)
30. Chu, W., Zhu, B.B., Xue, F., Guan, X., Cai, Z.: Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing urls. In: *Proceedings of 19th IEEE International Conference on communications (ICC'19)*, pp. 1990–1994 (2013)
31. Canali, D., Cova, M., Vigna, G., Kruegel, C.: Prophiler: a fast filter for the large-scale detection of malicious web pages. In: *Proceedings of 20th International Conference on World Wide Web (WWW'11)*, pp. 197–206 (2011)
32. Bell, S., Komisarczuk, P.: An analysis of phishing blacklists: google safe browsing, openphish, and phishtank. In: *Proceedings of 1st Australasian Computer Science Week Multiconference (ACSW'16)*, pp. 1–11 (2020)
33. Ma, J., Saul, L.K., Savage, S., Voelker, G.M.: Beyond blacklists: learning to detect malicious web sites from suspicious urls. In: *Proceedings of 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, pp. 1245–1254 (2009)
34. <https://www.alexa.com>
35. <https://chato.cl/webspam/datasets/>
36. OpenPhish, P.I.: Openphish (2020)
37. Analytics, R.: Dns-bh-malware domain blacklist (2007). <http://www.malwaredomains.com>
38. Kolari, P., Finin, T., Joshi, A., et al.: Svms for the blogosphere: Blog identification and splog detection. In: *Proceedings of AAAI Spring Symposium on Computational Approaches To Analysing Weblogs (CAAW'06)* (2006)
39. Patil, D.R., Patil, J.B., et al.: Malicious urls detection using decision tree classifiers and majority voting technique. *Cybern. Inf. Technol.* **18**, 11–29 (2018)
40. Hou, Y.-T., Chang, Y., Chen, T., Laih, C.-S., Chen, C.-M.: Malicious web content detection by machine learning. *Expert Syst. Appl.* **37**, 55–60 (2010)