Blockchain-and ML based Malware Detection and Integrity Checking: A Decentralized Approach

Araharan Loganayagam

School of Computing and Engineering

University of West London

21524785@student.uwl.ac.uk

*Abstract****—***This review paper provides an overview of the current state of malware detection and integrity checking techniques, highlighting their limitations and challenges. The review then focuses on the use of blockchain technology and machine learning algorithms for malware detection and integrity checking. First, the paper provides a review of the use of blockchain technology for security purposes, highlighting its potential benefits in providing a secure and decentralized platform for malware detection and integrity checking. The paper then discusses the use of machine learning algorithms for malware detection, analysing the different approaches proposed in existing literature. The paper also discusses the limitations and challenges of these approaches, including the need for large datasets and the risk of false positives. The paper then proposes a decentralized approach to malware detection and integrity checking, based on the integration of blockchain technology and machine learning algorithms. The proposed approach leverages the benefits of blockchain technology, such as its decentralized nature and immutability, to provide a secure platform for malware detection and integrity checking. The integration of machine learning algorithms enables the system to learn and adapt to new threats, improving its effectiveness over time. The review concludes with a discussion of future directions and emerging trends in blockchain-and ML based malware detection and integrity checking.

*Keywords****—*** Apache spark, big data, data mining, Hadoop, stream mining, Smart grids, lambda architecture

## Introduction

In recent years, malware attacks have become increasingly prevalent, sophisticated and pose a significant threat to the security of computer systems and networks. To address this challenge, researchers have developed various malware detection and integrity checking techniques, ranging from traditional signature-based methods to more advanced machine learning and blockchain-based approaches. While these techniques have been successful in detecting and preventing malware attacks, the ever-evolving nature of malware requires continued research and innovation to stay ahead of threats.

This literature review will focus on the current state of malware detection and integrity checking techniques, specifically exploring the potential of a decentralized approach based on blockchain and machine learning. By decentralizing malware detection and integrity checking, blockchain technology can provide a transparent, secure, and tamper-proof environment for detecting and preventing malware attacks. Machine learning can also enhance the accuracy and effectiveness of malware detection by analysing large amounts of data and identifying patterns that may not be visible to traditional signature-based methods.

The literature review will evaluate the strengths and limitations of existing techniques, including traditional signature-based methods, anomaly-based detection, and machine learning-based methods. It will then explore how blockchain and machine learning can be used together to create a decentralized approach for malware detection and integrity checking. Finally, the review will analyse the potential benefits of a decentralized approach, including increased transparency, security, and efficiency.

## Current state of malware detection and integrity checking techniques

### This section aims to provide an overview of the current state of malware detection and integrity checking techniques. Specifically, we will explore the various approaches used in malware detection and integrity checking, their strengths, and limitations, as well as recent advances and challenges in the field. By synthesizing and analysing the existing literature, this review aims to identify the gaps and opportunities for future research in this critical area of cyber security.

### Malware Detection

Malware detection and integrity checking are crucial components of modern cybersecurity systems. With the increasing sophistication of malware attacks, there is a need for advanced techniques to detect and prevent these threats. In this literature review, we will discuss the current state of malware detection and integrity checking techniques.

One of the popular techniques for malware detection is machine learning-based detection. In the paper "Flow-based Malware Detection Using Convolutional Neural Network", the authors proposed a system for malware detection using Convolutional Neural Networks (CNNs) on network traffic flow data. The system collects network traffic flow data in real-time from different sources such as routers, switches, and firewalls. The data is then processed and pre-processed to extract relevant features for malware detection. The extracted features are pre-processed to normalize the data and convert it into a format suitable for input to the CNN model. The pre-processed data is fed into the CNN model for classification of malware. The CNN model consists of multiple convolutional and pooling layers, followed by fully connected layers that perform classification. The model is trained on a dataset of labeled network traffic flow data to learn the patterns of malicious traffic. After the CNN model is trained, it can be used to detect malware in real-time network traffic flow.

Another popular technique for malware detection is based on API call frequency analysis. In the paper "Malware Detection based on API Calls Frequency", the authors proposed a system that uses a Convolutional Neural Network (CNN) to classify malware samples based on the frequency of API calls made by the malware program. The frequency of API calls is recorded as a sequence of integers, which is treated as an image, and the CNN is trained to classify this image as either malware or benign.

Furthermore, in the paper "Malware Detection based on Cloud Computing integrating Intrusion Ontology representation", the authors proposed a cloud-based malware detection system that uses an intrusion ontology to represent features and machine learning algorithms to classify malware. The proposed system can be deployed in a cloud environment for efficient and scalable malware detection.

Integrity checking is another important aspect of cybersecurity. In the paper "A Survey of Software Integrity Techniques", the authors surveyed various software integrity techniques, including code signing, checksums, and cryptographic hash functions. Code signing involves digitally signing the code to verify its authenticity and integrity. Checksums are used to verify the integrity of the software by comparing a calculated checksum with a known checksum. Cryptographic hash functions are used to generate a unique fingerprint of the software that can be used to verify its integrity.

In conclusion, machine learning-based detection techniques such as flow-based detection and API call frequency analysis, and cloud-based malware detection systems with intrusion ontology representation have shown promising results in detecting malware. Integrity checking techniques such as code signing, checksums, and cryptographic hash functions are also important for ensuring the integrity of software. However, with the increasing complexity and sophistication of malware attacks, there is a need for continuous research and development of new techniques for malware detection and integrity checking.

### Integrity Checking

According to a survey on data integrity and verification for cloud storage, the paper provides a comprehensive review of various techniques used for data integrity and verification in cloud storage. The paper categorizes these techniques into four broad categories, namely cryptographic techniques, erasure coding techniques, replication techniques, and secret sharing techniques. Cryptographic techniques involve the use of cryptographic algorithms such as hash functions and digital signatures to ensure data integrity. Erasure coding techniques encode data into multiple fragments and store them across multiple cloud storage servers to ensure data availability and integrity. Replication techniques replicate data across multiple cloud storage servers to improve the availability of data, and secret sharing techniques divide data into multiple shares and distribute them across multiple cloud storage servers to ensure data integrity and confidentiality. The paper provides insights into the strengths and limitations of each technique and can guide the selection of appropriate techniques for ensuring data integrity and verification in cloud storage.

An Approach to Verifying Data Integrity for Cloud Storage presents a comprehensive approach to ensuring the integrity of data stored in cloud storage. The paper focuses on the use of digital signatures and hash functions for data verification and authentication. The proposed approach leverages a PKI to manage the digital certificates and public keys necessary for generating and verifying digital signatures. The paper also highlights the importance of using cloud storage service APIs to interact with the cloud storage service and retrieve or store data blocks and digital signatures. The approach outlined in this paper can help to mitigate the risk of data tampering and ensure the authenticity and integrity of data stored in the cloud.

The research paper "An Adaptive Methodology for Integrity Checking in Cloud Storage" presents an adaptive methodology for integrity checking in cloud storage, which employs multiple techniques to detect single and multiple intrusions in cloud data. The methodology utilizes data replication, hashing, Merkle tree, secret sharing, and adaptive fault detection techniques. Data replication is used to create redundant copies of data, while hashing generates unique fingerprints of data blocks for detecting any changes or modifications to the data. The Merkle tree stores the hash values of data blocks for efficient verification of data integrity. Secret sharing distributes cryptographic keys used for data encryption and decryption across multiple servers to ensure that no single server has access to the entire key. An adaptive fault detection mechanism is used, which detects both single and multiple intrusions in cloud data through statistical analysis to detect anomalies in data patterns and trigger an alert in case of a security breach. The combination of these techniques makes the proposed methodology an effective approach for ensuring data integrity in cloud storage.

The research paper "Design and implementation for MD5-based data integrity checking system" proposes a system that uses MD5 hashing, block-based approach, redundant storage, error correction, and user authentication to ensure the integrity of data in cloud storage. The paper describes how the system generates an MD5 hash of the original file and stores it securely, and how it uses a block-based approach to detect changes in specific parts of the file. The system also stores the file and its corresponding hash values in multiple locations to ensure data availability and uses error-correcting codes to detect and correct errors in the hash values. Additionally, the system allows only authorized users to access and modify the data, ensuring data security and integrity. Overall, this proposed system provides an effective solution for ensuring the integrity of data stored in cloud storage.

The paper "File Integrity Checkers: Functionality, Attacks, and Protection" provides a comprehensive overview of various techniques used by file integrity checkers to ensure file integrity and protect against attacks. These techniques include hashing, digital signatures, change monitoring, Tripwire, anomaly detection, file system monitoring, and rootkit detection. The authors explain how each technique works and their strengths and weaknesses. The paper also discusses various attacks that can be used to bypass file integrity checkers, such as file replacement, timing attacks, and malicious software, and provides recommendations for improving file integrity checker security. Overall, this paper provides valuable insights into the functionality, attacks, and protection of file integrity checkers, and can be a useful resource for researchers and practitioners working in the field of cybersecurity.

In "Ensuring Data Integrity in Storage: Techniques and Applications," the authors provide an overview of various techniques used to ensure data integrity in storage. These techniques include cryptographic hash functions, error-correcting codes, RAID, data mirroring, data scrubbing, and data verification. The paper explains how these techniques work and their applications in ensuring data integrity. Cryptographic hash functions generate a fixed-size output (hash) from the input data to verify the integrity of data. Error-correcting codes detect and correct errors in data using redundancy. RAID provides data redundancy and protection against disk failures. Data mirroring creates an exact copy of the data on another disk to provide redundancy and protect against disk failures. Data scrubbing checks the data for errors and corrects them regularly. Data verification techniques generate a value from the data to verify the integrity of data. Overall, this paper provides an in-depth understanding of various techniques used to ensure data integrity in storage and their applications.

The research paper "Integrity Checking for Cloud Environment Using Encryption Algorithm" proposes a new methodology to ensure data integrity in a cloud environment. The proposed approach uses a combination of encryption algorithms, hash functions, and message authentication codes to protect data from unauthorized changes. The researchers suggest using AES and RSA encryption algorithms to encrypt and decrypt the data, SHA-256 hash function to generate a message digest, and HMAC algorithm to generate a cryptographic checksum for authenticating data integrity. The proposed methodology is specifically designed for cloud environments that involve distributed computing and storage resources. This research contributes to the existing literature by providing a secure and efficient approach to integrity checking in the cloud environment.

## Review of the use of blockchain technology for security purposes

A Blockchain-Based Efficient Data Integrity Verification Scheme in Multi-Cloud Storage. In summary, the paper proposes a novel scheme that uses blockchain technology along with Merkle trees, Shamir's secret sharing, and homomorphic encryption to ensure data integrity in multi-cloud storage environments. The use of blockchain technology provides a secure and decentralized method of storing hash values, while the Merkle tree structure allows for efficient verification of many files. Shamir's secret sharing method provides added security by distributing the encryption key across multiple cloud service providers. Finally, homomorphic encryption enables computation on encrypted data without the need for decryption, adding an extra layer of security to the scheme.

The paper "Blockchain-Based Data Audit and Access Control Mechanism in Service Collaboration" suggests using blockchain technology, smart contracts, data encryption, hashing, and access control techniques to ensure data integrity, security, and privacy in service collaboration environments. The authors propose using blockchain technology to facilitate secure and transparent data sharing and auditing. Smart contracts are used to automate and enforce data access control policies to ensure that only authorized parties can access and modify the data. The paper also suggests using data encryption techniques to protect the confidentiality of sensitive data stored in the cloud. Hashing is used to verify the integrity of data by generating a unique digital fingerprint of the data and comparing it with the original fingerprint at the time of access. The authors also propose a role-based access control mechanism to ensure that only authorized users have access to the data. Overall, the proposed solution offers a comprehensive approach to address data integrity and security challenges in service collaboration environments.

The paper "Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments" proposes a novel approach to ensure data integrity in cloud computing environments using blockchain technology. The authors use various techniques, including blockchain technology, smart contracts, distributed consensus mechanisms, public key cryptography, and Merkle trees to ensure the integrity and security of data stored in the cloud. The proposed system utilizes distributed consensus mechanisms to ensure that data stored in the blockchain is tamper-proof, and all nodes agree on its integrity. Smart contracts are used to automate the verification process of data integrity, ensuring automatic execution of data integrity checks. Additionally, public key cryptography is used to ensure secure communication between cloud servers and blockchain nodes. The authors also employ Merkle trees to enable efficient and fast verification of data integrity by storing data in a hierarchical data structure that allows for efficient verification of data. The proposed system provides a tamper-proof and reliable way to ensure data integrity in cloud computing environments.

The paper "Data Integrity Audit Scheme Based on Blockchain Expansion Technology" proposed a scheme for ensuring data integrity in cloud storage environments using blockchain expansion technology. The paper leverages blockchain technology, smart contracts, Merkle trees, SHA-256 hash function, and Bloom filters to provide a tamper-proof and efficient way of auditing data integrity. The proposed scheme also uses expansion technology to reduce the size of the blockchain and improve its efficiency. The authors demonstrated that their proposed scheme can efficiently and securely verify the integrity of data in cloud storage environments while reducing the computational overhead and false positive rate.

The proposed use of blockchain technology in the paper "Land Registration System Using Blockchain" provides a solution to the issues of land registration and transfer by creating a transparent, secure, and decentralized system. The smart contracts automate the land transactions, reducing the need for intermediaries and the costs associated with them. The decentralized identity management ensures that the identity of landowners is verifiable, and fraudulent activities are prevented. The consensus mechanism maintains the integrity of the blockchain, ensuring that all nodes agree on the state of the blockchain. Additionally, the user-friendly interface proposed in the paper makes the system accessible to all users. Overall, the use of blockchain technology and its various components in this paper offers a promising solution to improving the land registration system.

The paper "Modern Approaches to File System Integrity Checking" provides an overview of various techniques used in modern file systems to ensure data integrity. The paper discusses techniques such as checksums and hashing, data mirroring and redundancy, journaling, RAID, copy-on-write, snapshotting, file system consistency checking, and error-correcting codes. The paper analyzes the advantages and limitations of each technique and proposes how these techniques can be used together to provide a robust mechanism for data integrity checking. This paper is a valuable contribution to the field of data integrity checking and provides insights into the various techniques that can be used to ensure the integrity of data in file systems.

The paper "Reliable Data Storage and Sharing using Blockchain Technology and Two Fish Encryption" proposes a set of techniques for ensuring data integrity and security in cloud storage. The use of blockchain technology, TwoFish encryption, smart contracts, and decentralized storage provides a robust mechanism for data storage and sharing. The use of hashing, proof-of-work, PKI, digital signatures, and DDoS protection ensures that the network is secure, and transactions are authentic and authorized. Multi-factor authentication further enhances the security of the system by requiring users to provide more than one form of identification. The proposed techniques are a step towards creating a secure and decentralized cloud storage system that eliminates the need for a central authority and provides users with control over their data.

## Review of the use of machine learning algorithms for malware detection

The paper "NLP-based Entity Behavior Analytics for Malware Detection" proposes a novel approach to malware detection by utilizing natural language processing, entity behavior analytics, and machine learning techniques. Specifically, the paper uses NLP techniques such as Named Entity Recognition and Part-of-Speech tagging to extract relevant features from system call traces. These features are then analyzed using EBA to detect anomalies in the behavior of system entities that could be indicative of malware. The paper further employs machine learning techniques such as Random Forest and Support Vector Machine to classify system behavior as either normal or malicious based on the extracted features. Finally, the paper evaluates the proposed malware detection system using standard metrics such as precision, recall, F1 score, and AUC-ROC. Overall, the paper presents a promising approach to malware detection that could potentially improve the detection accuracy and reduce false positives compared to traditional signature-based approaches.

The paper "Review on Malware Classification and Malware Detection Using Transfer Learning Approach" provides a comprehensive overview of recent research on malware classification and detection using transfer learning. The paper highlights the significance of techniques such as malware feature extraction, transfer learning, deep learning, ensemble methods, and feature selection in improving the performance of malware detection models. The paper also discusses the various evaluation metrics and datasets used in recent research on this topic. Moreover, the paper highlights the challenges and limitations of using transfer learning for malware detection, including the need for large amounts of data and the risk of overfitting. Overall, the paper provides a useful summary of the state-of-the-art techniques and challenges in malware detection using transfer learning.

The paper "Semantics-Based Online Malware Detection: Towards Efficient Real-Time Protection Against Malware" presents a technical architecture for real-time protection against malware using semantics-based techniques. The architecture consists of several components, including data collection, pre-processing, malware detection model, decision engine, real-time protection module, and malware analysis and reporting. The data collection component collects the necessary data for detection, and the pre-processing component performs the initial processing of the collected data. The malware detection model component includes the main detection model that uses supervised learning and semantics-based techniques to detect malware. The decision engine component makes the final decision about whether the file is malware or not based on the output of the detection model. The real-time protection module component provides real-time protection against malware by analyzing the system calls made by running processes and comparing them with known malicious patterns. The malware analysis and reporting component provides a detailed analysis of the detected malware and generates reports for further investigation and remediation. Overall, the architecture is designed to efficiently and effectively protect against malware in real-time using semantics-based techniques.

The paper "Behavior-based Malware Analysis and Detection" proposes a comprehensive technical architecture for malware analysis and detection. The architecture consists of several stages including data collection, feature extraction, preprocessing, machine learning, evaluation, and malware detection. The paper emphasizes the use of behavior-based detection techniques, which involves analyzing the behavior of malware rather than its code or signature. The proposed architecture also incorporates several machine learning algorithms for malware classification such as unsupervised and supervised learning. Additionally, the paper highlights the importance of evaluation and proposes the use of cross-validation to evaluate the model's performance. The architecture's practicality and effectiveness are demonstrated through the proposed technique's evaluation using various metrics. Overall, this paper presents a comprehensive approach to malware detection that combines several techniques, including machine learning, to achieve high detection accuracy.

The paper "Cloud Based Malware Detection Through Behavioral Entropy" proposes a cloud-based approach to detect malware through the analysis of user behavior. The approach involves collecting data from user systems, extracting features from the data, calculating the behavioral entropy of the user, setting a threshold for anomalous behavior, detecting malware based on the entropy value, and generating alerts for further investigation. The paper focuses on using the behavioral entropy metric as a way to quantify the complexity and randomness of user behavior, which can be indicative of malware activity. The approach can be used in real-time and has the potential to detect unknown and zero-day malware. The paper also highlights the benefits of a cloud-based approach, including scalability, ease of deployment, and centralized management.

The paper "Lightweight Behavioral Malware Detection for Windows Platforms" proposes a technique for detecting malware on Windows platforms using behavior-based analysis. The approach involves collecting data from the Windows operating system, extracting features, and identifying anomalies by comparing the behavioral profiles to a pre-defined baseline. The system uses Windows Event Tracing to collect system call traces, network activity, and file system events, and then extracts relevant features such as system call sequences, network connections, and file access patterns. The system then compares the extracted features to a pre-defined baseline to identify anomalies and generate alerts. The system includes a feedback loop to improve detection accuracy by allowing users to provide feedback on the generated alerts. Overall, the proposed approach is lightweight and effective in detecting malware on Windows platforms.

The paper "Flow-based Malware Detection Using Convolutional Neural Network" proposed a system for malware detection using Convolutional Neural Networks (CNN’s) on network traffic flow data. The system collects network traffic flow data in real-time from different sources and then extracts relevant features for malware detection. The extracted features are pre-processed to normalize the data and convert it into a format suitable for input to the CNN model. The CNN model is then used for classification of malware, which consists of multiple convolutional and pooling layers, followed by fully connected layers that perform classification. The model is trained on a dataset of labeled network traffic flow data to learn the patterns of malicious traffic. After the CNN model is trained, it can be used to detect malware in real-time network traffic flow, and an alert is generated to notify the system administrator or security team. The system periodically updates the CNN model with new malware samples to improve the accuracy of detection.

The paper titled "Malware Detection based on API Calls Frequency" proposed a malware detection approach based on the frequency of API calls made by a program during execution. The approach involves data collection, feature extraction, data pre-processing, model training, model testing, and malware detection. During data collection, the focus is on monitoring the API calls made by the malware program using dynamic analysis techniques. Relevant features related to the frequency of API calls are extracted and pre-processed to remove any noise and make the data suitable for machine learning algorithms. A Convolutional Neural Network (CNN) is then used to train a model to classify the extracted features as either malware or benign. The performance of the trained model is evaluated in terms of accuracy, precision, recall, and F1 score using a separate dataset. Finally, the trained model is used to detect malware samples by feeding in the API call frequency data and predicting whether the sample is malware or benign.

The paper "Malware Detection based on Cloud Computing integrating Intrusion Ontology representation" proposes a novel approach to malware detection that leverages cloud computing and ontology-based feature extraction. The authors collect malware samples from various sources and extract features from them using API call sequences. They represent these features using an intrusion ontology, which defines the domain of intrusion detection. The proposed cloud-based malware detection system has three main components: data pre-processing, ontology-based feature extraction, and classification using machine learning algorithms. The system's performance is evaluated using various metrics, and the results demonstrate the effectiveness of the proposed system compared to other state-of-the-art malware detection systems. Finally, the system can be deployed in a cloud environment for real-time and scalable malware detection. The proposed approach presents a promising direction for the development of more robust and effective malware detection systems.

## Analysis of the different approaches proposed in existing literature

### Malware Detection

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Title and Researchers | Purpose | Key Findings | Advantages | Limitations |
| NLP-based Entity Behaviour Analytics for Malware Detection  (Pejman Najafi, Daniel Koehler, Feng Cheng, Christoph Meinel ) | Natural Language Processing (NLP)-based Entity Behaviour Analytics (EBA) framework for detecting malware in computer networks. | NLP-based EBA approach could effectively detect malware in computer networks with a high degree of accuracy.  By analyzing the language used in network logs, the system could identify anomalous behavior and alert network administrators to potential security threats. | It can detect malware that traditional signature-based approaches may miss, as well as identify new, previously unknown threats.  The approach is also more scalable than traditional methods, as it can analyze large amounts of log data quickly and efficiently. | it relies on accurate and complete log data to be effective.  If log data is incomplete or inaccurate, the system may miss malware infections. Additionally, the approach may have difficulty detecting malware that is specifically designed to evade NLP-based detection methods.  Finally, the approach may generate false positives if normal network behavior is mistakenly identified as anomalous. |
| Review on Malware Classification and Malware Detection Using Transfer Learning Approach  (Priya V, Dr. Sathya Sofia A) | an overview of the current state-of-the-art techniques and methods for malware classification and detection, specifically focusing on transfer learning-based approaches. | Transfer learning allows for the transfer of knowledge from one domain to another, making it possible to improve the accuracy and efficiency of malware detection systems.  Hybrid models that combine different types of features, such as static and dynamic features, have been found to be effective in malware classification and detection.  The use of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has also shown promising results in malware classification and detection. | comprehensive overview of the current state-of-the-art techniques and methods for malware classification and detection and highlighting the potential of transfer learning-based approaches in improving the accuracy and efficiency of malware detection systems. | comparative analysis of the different transfer learning-based approaches and not discussing the challenges and limitations associated with using transfer learning in malware classification and detection. |
| Semantics-Based Online Malware Detection:  Towards Efficient Real-Time Protection  Against Malware  (Sanjeev Das, Yang Liu, Wei Zhang, and Mahintham Chandramohan) | new approach to online malware detection based on semantic analysis of application behaviour.  The authors aim to provide more efficient real-time protection against malware. | The proposed approach can detect both known and unknown malware in real-time.  The semantic analysis of application behaviour provides a more effective way to detect malware than traditional signature-based approaches.  The approach can handle code obfuscation and other techniques used by malware authors to evade detection.  The approach can detect malware even when it is injected into legitimate processes. | Improved detection rates: The semantic analysis-based approach is more effective at detecting malware than traditional signature-based approaches.  Real-time protection: The approach provides real-time protection against malware.  Handling of obfuscation techniques: The approach can detect malware even when it is hidden using code obfuscation techniques.  Detection of injected malware: The approach can detect malware that has been injected into legitimate processes. | High false positive rate: The semantic analysis-based approach may generate false positives.  Performance overhead: The approach may have a performance overhead due to the need for semantic analysis of application behaviour.  Limited application coverage: The approach may not be able to detect all types of malwares, especially those that do not exhibit malicious behaviour.  Requires access to network traffic: The approach requires access to network traffic to detect malware that communicates with external servers. |
| Behaviour-based Malware Analysis and Detection  (LIU Wu1), REN Ping2), LIU Ke3), DUAN Hai-xin1) | propose a behaviour-based approach to detect and analyse malware.  The paper proposes a new system architecture that includes three components: behaviour monitoring, analysis, and detection. | A behaviour-based approach is effective in detecting and analysing malware.  The proposed system architecture has the potential to be an efficient and accurate method for malware detection and analysis.  The use of machine learning algorithms can enhance the performance of malware detection. | The proposed behaviour-based approach can detect and analyse malware even if it is not known beforehand.  The system architecture proposed in the paper is scalable and can be adapted to different environments and applications.  The use of machine learning algorithms can improve the accuracy and efficiency of malware detection. | The proposed system architecture requires significant computational resources to analyse the behaviour of the malware.  The accuracy of the system heavily depends on the quality of the behaviour models used.  The system may generate false positives or false negatives, leading to incorrect detection or non-detection of malware. |
| Cloud Based Malware Detection Through Behavioural Entropy  (Kambiz Vahedi  ,Khadijeh Afhamisisi ) | propose a cloud-based malware detection system that uses behavioural entropy to detect and classify malware in real-time. | The proposed system achieved a high detection rate of 98% and low false-positive rate of 0.05%.  The system can classify malware into different families with high accuracy.  The system is scalable and can handle many requests simultaneously. | It can detect previously unknown malware and zero-day attacks.  It can handle many requests simultaneously, making it suitable for use in cloud environments.  It uses behavioural entropy, which is a novel approach that can detect subtle changes in malware behaviour. | It may not be effective against malware that uses sophisticated evasion techniques.  It may generate false positives if the user's behaviour changes frequently. |
| Lightweight Behavioural Malware Detection for Windows Platforms  (Spiros Mancoridis,  Hunter Dong,  Avinash Srinivasan,  Bander Alsulami) | propose a lightweight and efficient method for detecting malware in Windows platforms based on the analysis of the behaviour of software. | development of a behaviour-based malware detection system for Windows platforms, which is lightweight and efficient, and can detect previously unseen malware. | The development of a fast and accurate malware detection system that is effective against zero-day attacks and new malware variants.  The system is also lightweight, which means that it can be deployed on low-resource systems without affecting performance. | system relies on the analysis of behavioural patterns of malware, which may be limited in certain situations.  Additionally, the system may not be effective against malware that is designed to evade behaviour-based detection systems. |
| Flow-based Malware Detection Using Convolutional Neural Network  (Gustavo Isaza Echeverri, Andrés G ,Khadijeh Afhamisis) | propose a new method for detecting malware using flow-based analysis and convolutional neural networks (CNN). | achieves high accuracy in detecting malware and outperforms traditional machine learning approaches. | The advantage of the proposed method is that it can handle large volumes of network traffic data and detect malware with high accuracy. | the limitation of the paper is that it only focuses on detecting a specific type of malware and does not consider other types of cyber threats. |
| Malware Detection based on API Calls Frequency  (Vidhi Garg, Rajesh Kumar Yadav) | propose a method for malware detection using the frequency of API calls made by a software program.  The paper presents a framework that analyses the sequence of API calls made by a program and uses it to classify it as either malicious or benign. | analysing API calls frequency is an effective method for detecting malware.  The proposed framework achieved an accuracy of 99.05% in detecting malware samples from the Zoo malware dataset. | it does not require access to the source code of the program and can detect zero-day malware.  Additionally, the method is computationally efficient and can be applied in real-time. | it may produce false positives if the program under analysis makes many API calls.  it may not be effective against malware that does not make API calls. Furthermore, it may be vulnerable to attacks that modify the frequency of API calls made by a program to evade detection. |
| Malware Detection based on Cloud Computing integrating Intrusion Ontology representation.  (Cristian Adrián Martínez, Gustavo Isaza Echeverri, Andrés G. Castillo Sanz) | Computing integrating Intrusion Ontology representation is to propose a novel approach for detecting malware in cloud computing environments by integrating intrusion ontology representation. | effective in detecting known and unknown malware in cloud computing environments with a high detection rate and low false-positive rate.  The use of intrusion ontology representation helps to improve the accuracy and efficiency of malware detection.  The proposed approach is scalable and can handle large-scale cloud computing environments. | The approach can detect both known and unknown malware.  The use of intrusion ontology representation helps to improve the accuracy and efficiency of malware detection.  The approach is scalable and can handle large-scale cloud computing environments. | The approach may require significant computing resources to handle large-scale cloud computing environments.  The approach may require frequent updates to the intrusion ontology to keep up with new malware threats.  The approach may be vulnerable to evasion techniques employed by advanced malware. |

### Integrity Checking

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Title and Researchers | Purpose | Key Findings | Advantages | Limitations |
| An Approach to Verifying Data Integrity for Cloud Storage  (Yindong Chen, Liping Li, Ziran Chen) | propose a method for verifying data integrity in cloud storage services using digital signatures.  The paper aims to address the challenge of ensuring the authenticity and integrity of data stored in the cloud. | uses hash functions to generate a unique digest of the data blocks stored in the cloud, digital signatures to provide authentication and integrity for the hash values generated by the hash functions.  and a PKI to manage the digital certificates and public keys necessary for the generation and verification of the digital signatures.  The approach is implemented using the cloud storage service API to interact with the cloud storage service to retrieve and store data blocks and digital signatures. | Ensuring the authenticity and integrity of data stored in the cloud.  Providing a secure and efficient method for verifying data integrity in cloud storage services.  Seamless integration with cloud storage services using the cloud storage service API.  Use of widely available cryptographic techniques such as hash functions and digital signatures. | Dependence on the cloud storage service API, which may not be available or may not provide the necessary functionality for implementing the proposed approach.  The need for a PKI to manage the digital certificates and public keys, which may add complexity and overhead to the implementation.  The approach may not provide protection against other types of security threats, such as unauthorized access or data leakage.  The proposed approach may require additional processing and storage resources, which could affect performance and cost. |
| An Adaptive Methodology for Integrity  Checking in Cloud Storage  (L Jambulingam, T V Ananthan, P S Rajakumar) | adaptive methodology for integrity checking in cloud storage that can detect both single and multiple intrusions in cloud data. | The proposed methodology effectively ensures the integrity of data stored in cloud storage by utilizing multiple techniques, including data replication, hashing, Merkle tree, secret sharing, and adaptive fault detection.  The adaptive fault detection mechanism used in the methodology can detect both single and multiple intrusions in cloud data by utilizing statistical analysis to detect anomalies in data patterns.  The proposed methodology is flexible and can be customized to meet the specific requirements of different cloud storage systems. | The methodology effectively ensures the integrity of data stored in cloud storage by utilizing multiple techniques, making it more robust and secure.  The adaptive fault detection mechanism used in the methodology can detect both single and multiple intrusions in cloud data, providing an added layer of security.  The methodology is flexible and can be customized to meet the specific requirements of different cloud storage systems. | The methodology relies on the use of multiple techniques, which may increase computational overhead and complexity.  The proposed methodology assumes that cloud storage providers are honest and do not collude with attackers, which may not always be the case in practice.  The effectiveness of the proposed methodology may be affected by the size and complexity of the cloud storage system, as well as the quality of the statistical analysis used in the adaptive fault detection mechanism. |
| Design and implementation for MD5-based data integrity checking system.  (Danyang Cao, Bingru Yang) | propose a system for ensuring the integrity of data in cloud storage using the MD5 algorithm. | The proposed system can detect changes in specific parts of the file by dividing it into blocks and generating a hash for each block.  The system uses error-correcting codes to detect and correct errors in the hash values, thus ensuring data integrity.  The system provides redundant storage to ensure that data can be retrieved in case of failures.  The system allows only authorized users to access and modify the data, thus ensuring data security. | it can ensure the integrity of data in cloud storage by using a simple and efficient MD5-based algorithm.  The system also provides additional features such as user authentication, redundant storage, and error correction to ensure data security and availability. | MD5 algorithm is susceptible to collision attacks, which can compromise the integrity of the data.  Additionally, the system may not be suitable for very large files as it may become computationally intensive to generate and store hashes for each block. |
| Integrity Checking for Cloud Environment Using Encryption Algorithm  (P.Varalakshmi , Hamsavardhini Deventhiran) | ensuring data integrity in the cloud environment using encryption algorithms. | data integrity in the cloud environment using the AES encryption algorithm.  The proposed technique uses a hash function to calculate the hash value of the data before encryption and stores the hash value along with the encrypted data.  During data retrieval, the hash value is recalculated, and the recalculated hash value is compared with the stored hash value to ensure data integrity. | high level of security for data stored in the cloud environment.  The use of encryption algorithms ensures that the data is secure from unauthorized access.  The use of a hash function for calculating the hash value of the data ensures that any modifications to the data are detected. | overhead associated with the use of encryption algorithms and hash functions.  The use of encryption algorithms may slow down the data transfer rate, and the use of hash functions may increase the processing time required for data retrieval.  Additionally, the technique does not provide protection against attacks that exploit vulnerabilities in the encryption algorithm. |
| A Blockchain-Based Efficient Data Integrity Verification Scheme in Multi-Cloud Storage  (Yiran Zhang,Huizheng Geng,Li Su and Li Lu) | propose a new approach for ensuring the integrity of data stored in a multi-cloud storage environment.  The authors propose a blockchain-based scheme for efficient data integrity verification using a combination of cryptographic techniques and blockchain technology. | The key findings of the research paper are that the proposed scheme is more efficient and effective than existing schemes for data integrity verification in multi-cloud storage environments.  The scheme provides a high level of security and reliability while reducing the computational overhead and communication costs associated with traditional integrity checking methods. | The main advantage of the proposed scheme is that it ensures data integrity in a distributed environment without relying on a trusted third party or central authority.  It also provides a transparent and tamper-proof way to verify the integrity of data in a multi-cloud storage environment. | need for further research and experimentation to evaluate the scalability and performance of the proposed scheme in larger and more complex multi-cloud storage environments.  There may also be challenges in implementing the scheme in practice due to the need for coordination and agreement among multiple cloud providers. |
| Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments  (Dr.B. Ravishankar, Prateek Kulkarni, Vishnudas M V) | propose a solution for ensuring data integrity in cloud computing environments using blockchain technology.  The paper introduces a blockchain-based database system that provides a tamper-proof, secure, and efficient method for data integrity verification in cloud computing. | successful implementation of the proposed system, which provides secure and efficient data integrity verification in a cloud computing environment.  using a blockchain-based approach for data integrity verification, such as improved security, transparency, and tamper-proofing. | the ability to provide a tamper-proof and secure method for data integrity verification in a cloud computing environment, improved transparency, and reduced reliance on centralized authorities.  The system also provides efficient and cost-effective data integrity verification, which can benefit organizations that deal with large amounts of data. | The limitations of the research paper include the fact that the proposed system is not immune to attacks, and there may be some scalability issues when dealing with many transactions.  Additionally, the paper does not provide a detailed analysis of the performance of the proposed system, and further research may be required to optimize the system's performance. |
| Data Integrity Audit Scheme Based on Blockchain Expansion Technology  (Zhenpeng Liu Yongjiang, Lele Ren and Weihua Zheng) | to propose a data integrity audit scheme based on blockchain expansion technology for cloud storage.  The paper aims to enhance data security by providing a tamper-proof and decentralized mechanism for data integrity verification in the cloud. | achieve efficient and secure data integrity verification in the cloud environment.  The scheme uses blockchain expansion technology to improve the performance of blockchain-based data integrity verification schemes, making it more practical for use in large-scale cloud storage systems. | ability to achieve real-time data integrity verification with low computational overhead, its tamper-proof and decentralized nature, and its ability to provide an efficient and scalable solution for data integrity verification in cloud storage. | proposed scheme has not been implemented and evaluated in a real-world cloud storage system. which may affect its practicality and performance.  Additionally, the paper does not provide a detailed analysis of the security and privacy implications of the proposed scheme, which could be a potential area for further research. |
| Land Registration System Using Blockchain  (SaiApurva Gollapalli, Gayatri Krishnamoorthy, Neha Shivaji Jagtap, Rizwana Shaikh) | propose a blockchain-based land registration system that can address issues related to land registration such as fraud, corruption, and inefficiencies in the current centralized systems. | The paper proposes a blockchain-based system for land registration that can eliminate intermediaries and provide a tamper-proof and transparent system for land registration.  The proposed system utilizes smart contracts to automate the registration process, reduce transaction costs, and ensure secure and efficient transfer of land ownership.  The system also provides a decentralized platform for storing and managing land records, which can be accessed by stakeholders in a secure and transparent manner. | The proposed system can bring several advantages, including improved efficiency, transparency, security, and reduced costs.  The use of blockchain technology can eliminate intermediaries, reduce transaction costs, and minimize the risk of fraud and corruption in land registration.  The system can provide a tamper-proof and transparent platform for managing land records, which can improve transparency and reduce the time required for processing land transactions. | The paper does not discuss the challenges related to the implementation of the proposed system, such as the legal and regulatory framework, scalability, and interoperability with existing systems.  The paper also does not provide a detailed analysis of the potential risks and vulnerabilities associated with the use of blockchain technology in land registration. |
| Reliable Data Storage and Sharing using Blockchain Technology and Two Fish Encryption  (S. Sivanantham, M. Sakthivel, V. Krishnamoorthy, N. Balakrishna, V. Akshaya) | propose a secure and reliable data storage and sharing system using blockchain technology and Two Fish encryption. | The proposed system provides a high level of security and reliability for data storage and sharing.  Blockchain technology ensures data integrity, immutability, and transparency.  Two Fish encryption provides strong data confidentiality.  The system is resistant to attacks such as data tampering, data theft, and denial of service attacks. | High level of security and reliability for data storage and sharing.  Decentralized and transparent data management using blockchain technology.  Strong data confidentiality using Two Fish encryption.  Resistant to various types of attacks. | The system relies on the availability of the blockchain network and the Two Fish encryption algorithm.  The system may have slower performance compared to centralized data storage and sharing systems.  The system may require significant computational resources for encryption and decryption operations. |

# Conclusion

In conclusion, traditional malware detection methods have limitations that make them less effective in detecting new and unknown types of malwares. These limitations have led to the development of new approaches such as blockchain-and ML based malware detection and integrity checking. This approach addresses the challenges of making generalized malware detection models by using machine learning algorithms to detect malware behavior and combining them with blockchain technology for decentralized and secure verification of the integrity of the detection system. However, there are still challenges to overcome, such as adversarial attacks and the need for fast and efficient models. Further research is needed to develop more robust and effective solutions that can keep up with the constantly evolving threat landscape. Overall, blockchain-and ML based malware detection and integrity checking present a promising direction for the development of more effective and efficient malware detection systems.