An Analysis of Blockchain-and ML based Malware Detection and Integrity Checking Systems

Name: Araharan Loganayagam
Student Id: 21524785
School of Computing and Engineering
University of West London
21524785@student.uwl.ac.uk

Supervisor: Dr Waqar Asif
Senior Lecturer in Computing Science and Cyber Security
School of Computing and Engineering
waqar.asif@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 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— NLP, Malware, Blockchain, CNN, Cloud, Integrity, Cryptography, RAID

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 (Cristian, 2010).

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.

The 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 by Kang et al. (2019), 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 architecture of the CNN model includes several convolutional and pooling layers, as well as fully connected layers which are responsible for carrying out the classification process. The model is trained on a dataset of labelled 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 (Kang, 2019).

Another popular technique for malware detection is based on API call frequency analysis. In the paper by Vidhi et al. (2019), 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 CNN is trained to classify this image as either malware or benign (Vidhi, 2019).

Furthermore, Cristian et al. (2010) 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 (Cristian, 2010).

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

The literature review covers several papers that propose different techniques to ensure data integrity in various systems. Suchetha et al. (2020) conducted a survey on data integrity and verification techniques for cloud storage, categorizing them into cryptographic, erasure coding, replication, and secret sharing techniques. Cryptographic techniques use cryptographic algorithms to ensure data integrity and authentication. They include techniques such as digital signatures, message authentication codes (MACs), and hash functions. Digital signatures provide non-repudiation by verifying the identity of the sender of the data, while MACs provide message integrity by generating a code based on the data that can only be verified by a key holder. Hash functions generate a fixed-size output, or hash, of the data that can be used to verify the data's integrity. Erasure coding techniques divide data into small blocks and generate additional redundant blocks that can be used to recover the original data if some blocks are lost or corrupted and high fault tolerance and availability for data stored in cloud storage. Replication techniques involve making multiple copies of the data and storing them in different locations, this can provide high availability and fault tolerance for data stored in cloud storage. Secret sharing techniques divide data into multiple shares, each of which is stored separately. The data can only be recovered if enough shares are combined. Secret sharing techniques can provide high security for data stored in cloud storage by ensuring that no single party can access the data without the required number of shares. 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 a similar vein, Yindong Chen et al. (2017) proposed an approach that leverages digital signatures and hash functions for data verification and authentication. 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. These two papers can provide a comprehensive overview of data integrity and verification techniques in cloud storage and guide the selection of appropriate techniques for ensuring data integrity and verification. (Suchetha R, 2020; Yindong Chen, 2017)

The research papers by Jambulingam et al. (2019) and Danyang et al. (2010) propose methodologies for ensuring data integrity in cloud storage. Jambulingam et al. (2019) presents an adaptive methodology that employs multiple techniques, such as data replication, hashing, Merkle tree, secret sharing, and adaptive fault detection, to detect single and multiple intrusions in cloud data. Data replication involves storing multiple copies of the same data in different locations, which can help to detect and recover from data loss or corruption. Hashing involves generating a unique, fixed-length digital signature or hash value from the data, which can be used for verifying the integrity of the data. The Merkle tree is a hash-based data structure that can be used to efficiently verify the integrity of large datasets by dividing the data into smaller blocks and constructing a tree structure where each node represents the hash value of its child nodes. Secret sharing involves dividing the data into multiple shares, which are distributed across different locations or devices, and requiring a certain number of shares to be combined to reconstruct the original data. Finally, adaptive fault detection involves dynamically monitoring the system for unusual or abnormal behaviour and adjusting the detection thresholds and parameters accordingly to improve the accuracy and reliability of intrusion detection. The combination of these techniques makes the proposed methodology an effective approach for ensuring data integrity in cloud storage.

On the other hand, Danyang et al. (2010) 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 system generates an MD5 hash of the original file and stores it securely, uses a block-based approach to detect changes in specific parts of the file, and stores the file and its corresponding hash values in multiple locations to ensure data availability. Additionally, the system allows only authorized users to access and modify the data, ensuring data security and integrity. These proposed methodologies provide effective solutions for ensuring the integrity of data stored in cloud storage and can guide the selection of appropriate techniques for data integrity and verification in cloud storage (Jambulingam, 2019; Danyang, 2010).

Ahmed et al. (2022) provides an overview of various techniques used by file integrity checkers, including hashing, digital signatures, change monitoring, and rootkit detection, and discuss attacks that can be used to bypass them. Gopalan et al. (2005) discuss techniques such as cryptographic hash functions, error-correcting codes, RAID, data mirroring, data scrubbing, and data verification used to ensure data integrity in storage. Cryptographic hash functions generate a unique fixed-size digest of a message or file, and any change in the message or file will result in a different digest value. By storing and comparing the hash values of the original and received data, the integrity of the data can be verified. Error-correcting codes are used to detect and correct errors that occur during data transmission or storage. These codes add extra bits to the data that can detect and correct errors in the original data. Redundant Array of Independent Disks (RAID) is a data storage technology that uses multiple disks to create a single logical unit. RAID can provide data redundancy and increase the performance and reliability of storage systems. Data mirroring is a method of replicating data to multiple disks or servers. The data is stored on two or more disks in real-time, and any changes made to one copy of the data are automatically updated on the other copies. Data scrubbing is a process that detects and corrects errors in data stored on disks or other storage media. This process involves scanning the data for errors and correcting them to ensure data integrity. Data verification involves checking the integrity of data stored on disks or other storage media. This can be done by comparing the stored data with a checksum or hash value to ensure that the data has not been modified or corrupted. Varalakshmi et al. (2012) propose a new methodology that uses encryption algorithms, hash functions, and message authentication codes to protect data integrity in cloud environments. These papers contribute to the existing literature by providing effective solutions for ensuring data and file integrity in cloud storage and can be valuable resources for researchers and practitioners working in the field of cybersecurity.

The Review of the use of blockchain technology for security purposes

In recent years, ensuring data integrity has become increasingly important due to the proliferation of data breaches and cyber-attacks. Several research papers propose different techniques to ensure data integrity, each with its own strengths and weaknesses. Ahmed et al. (2022) provides a comprehensive overview of various techniques used by file integrity checkers, while Gopalan et al. (2005) describe techniques used to ensure data integrity in storage. Yiran et al. (2022) proposes a novel scheme that uses blockchain technology for data integrity in multi-cloud storage environments. The authors suggest that the use of blockchain technology can improve data integrity and security in multi-cloud storage environments by reducing the risk of single points of failure and eliminating the need for centralized authorities. The proposed scheme employs a combination of techniques, including data replication, Merkle trees, and smart contracts to provide efficient and scalable solutions for data integrity verification. The authors also present a detailed evaluation of the proposed scheme's performance and scalability, showing that it is capable of handling large-scale data sets efficiently while maintaining high levels of security and data integrity. Overall, the research paper presents a promising approach to addressing the challenges associated with data integrity in multi-cloud storage environments. Finally, Chao et al. (2019) proposed a comprehensive approach to ensure data integrity, security, and privacy in service collaboration environments. This approach utilizes a combination of various techniques, A decentralized and distributed ledger system that maintains a tamper-proof and transparent record of transactions. Smart contracts Self-executing contracts with the terms of the agreement between the parties being directly written into code. Data encryption of transforming data into an unreadable format to prevent unauthorized access. Hashing The process of generating a unique fixed-length string of characters (hash) from a given input data, which can be used for data integrity verification. Access control techniques used to restrict or grant access to data and resources based on user authentication and authorization. By combining these techniques, the proposed approach aims to provide a robust and secure framework for data sharing and collaboration in service environments.

Several research papers propose different schemes for ensuring data integrity in cloud computing environments using blockchain technology. Ravishankar et al. (2010) proposed a novel approach that uses 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. Zhenpent et al. (2022) have proposed a scheme for auditing data integrity that makes use of various technologies such as blockchain expansion, smart contracts, Merkle trees, SHA-256 hash function, and Bloom filters. The proposed scheme aims to provide an efficient and tamper-proof way of auditing data integrity. The use of blockchain expansion technology and smart contracts helps to ensure the immutability and transparency of the audited data. The Merkle trees provide a way to efficiently verify the integrity of large amounts of data while the SHA-256 hash function helps to ensure the integrity of the data. Bloom filters, on the other hand, are used to improve the efficiency of the scheme by reducing the need to access the data stored in the blockchain. Overall, the proposed scheme aims to provide a robust and efficient way of auditing data integrity that can be useful in various applications such as cloud storage and service collaboration environments. Their proposed scheme uses expansion technology to reduce the size of the blockchain and improve its efficiency. Both schemes provide efficient and secure ways to verify the integrity of data in cloud storage environments, while reducing the computational overhead and false positive rate (Ravishankar, 2020; Zhenpeng, 2022).

Review of the use of machine learning algorithms for malware detection

The literature review covers several papers that propose technical architecture and approaches for malware detection and classification using different techniques. Pejman et al. (2021) presents a novel approach to malware detection using natural language processing, entity behaviour analytics, and machine learning techniques. The approach involves extracting features from system call traces using NLP techniques and analysing them using EBA to detect anomalies in system entity behaviour that could be indicative of malware. Machine learning techniques such as Random Forest and Support Vector Machine are then used to classify the system behaviour as normal or malicious based on the extracted features. On the other hand, Priya et al. (2023) 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. Sanjeev et al. (2016) presents a technical architecture for real-time protection against malware using semantics-based techniques that involves 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 gathers various types of data, such as system call traces, network traffic, and file system activities. The pre-processing component then pre-processes this data to extract relevant features for the malware detection model. The malware detection model uses a combination of static and dynamic analysis techniques to identify malicious code. The decision engine combines the results of the malware detection model and other contextual information to decide about the maliciousness of the code. If the code is deemed malicious, the real-time protection module takes action to prevent it from causing harm to the system. Finally, the malware analysis and reporting component provides detailed information about the malware to aid in future analysis and prevention. Overall, the proposed architecture provides a comprehensive approach to real-time malware protection that incorporates multiple techniques and components to ensure effective detection and prevention. Finally, the paper by Wu et al. (2011) proposes a technical architecture for malware analysis and detection, which involves several stages to achieve accurate and efficient detection. The first stage is data collection, where data samples containing malware are collected from various sources. The next stage is featuring extraction, where the relevant features from the collected data are extracted. The extracted features are then pre-processed in the third stage, which involves techniques such as normalization, filtering, and scaling. In the fourth stage, machine learning algorithms are applied to the pre-processed features to train a model for malware detection. The fifth stage involves evaluating the trained model's performance using various metrics such as accuracy, precision, recall, and F1-score. Finally, in the last stage, the trained model is used to detect malware in real-time. The paper emphasizes the use of behaviour-based detection techniques over signature-based techniques due to their higher accuracy and effectiveness against new and unknown malware. Additionally, the paper highlights the importance of cross-validation techniques to ensure that the model's performance is accurate and reliable. Overall, the proposed architecture provides a comprehensive approach to malware detection and analysis that could improve the accuracy and efficiency of existing systems.

The research paper by Kambiz et al. (2021) presents a novel cloud-based approach for detecting malware using a behavioural entropy metric. The proposed approach is designed to detect unknown and zero-day malware by analysing user behaviour. The authors argue that traditional signature-based approaches are limited in their ability to detect unknown and zero-day malware, and that behavioural analysis can provide a more effective solution. The proposed approach involves collecting data on user behaviour and analysing it using a variety of techniques, including entropy-based analysis. The authors demonstrate the effectiveness of their approach through experiments and evaluations using real-world datasets. The results show that the proposed approach can effectively detect unknown and zero-day malware with a high level of accuracy. The paper provides valuable insights into the use of behavioural analysis techniques for malware detection and highlights the potential of cloud-based approaches for improving the effectiveness and efficiency of existing malware detection systems. The research paper by Bander et al. (2017) presents a lightweight and efficient approach to malware detection on Windows platforms using behaviour-based analysis. The proposed approach involves comparing behavioural profiles of running processes to a pre-defined baseline to detect any deviations from normal behaviour. The authors highlight the limitations of traditional signature-based approaches to malware detection and emphasize the need for a more dynamic and adaptive approach that can detect new and unknown malware. The paper describes the various stages involved in the proposed approach, including data collection, feature extraction, behavioural profiling, and classification. The authors also conducted experiments to evaluate the effectiveness of the proposed approach, which demonstrated high detection rates and low false positives. The research paper by Cristian et al. (2010) proposes a new approach to malware detection that utilizes cloud computing and ontology-based feature extraction using machine learning algorithms. The paper highlights the effectiveness of the proposed system compared to other existing malware detection systems. The authors explain that the proposed system's effectiveness is due to its ability to detect malware using a combination of techniques, including cloud computing, ontology-based feature extraction, and machine learning. The paper demonstrates the importance of using multiple techniques to improve malware detection accuracy and scalability. The results of the proposed system are encouraging, suggesting that combining these techniques can significantly improve the accuracy of malware detection. The study demonstrates the potential benefits of using cloud computing and ontology-based feature extraction techniques for improving malware detection.

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)	A framework that utilizes Entity Behaviour Analytics (EBA) based on Natural Language Processing (NLP) to detect 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.	For the method to work efficiently, it requires precise and comprehensive log data. 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)	This paper provides a summary of the current advanced methods and techniques for detecting and categorizing malware, with a specific emphasis on approaches that employ transfer learning.	Transfer learning allows for the transfer of knowledge from one domain to another, this enables the enhancement of precision and effectiveness 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. Deep learning methods such as CNNs and RNNs have demonstrated encouraging outcomes in the field of malware classification and detection.	A thorough examination of the latest advanced techniques and methods for identifying and categorizing malware is provided in this review. The review emphasizes the possibility of enhancing the precision and efficiency of malware detection systems by utilizing transfer learning-based approaches.	The focus of the analysis was on comparing the various transfer learning-based approaches and did not involve addressing the difficulties and restrictions linked to using transfer learning for malware detection and classification.
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.	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.	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.

		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.	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.	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. Machine learning techniques have the potential to enhance the precision and effectiveness of identifying malware.	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	propose a cloud-based malware detection system that uses behavioural	The proposed system achieved a high detection rate of 98% and low	It can detect previously unknown malware and	It may not be effective against malware that uses sophisticated

Behavioural Entropy (Kambiz Vahedi ,Khadijeh Afhamisisi)	entropy to detect and classify malware in real-time.	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.	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.	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	propose a method for malware detection using the frequency of API calls made	analysing API calls frequency is an effective method for	It has the ability to detect zero-day malware and doesn't necessitate	it may produce false positives if the program under analysis

(Vidhi Garg, Rajesh Kumar Yadav)	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.	detecting malware. The proposed framework achieved an accuracy of 99.05% in detecting malware samples from the Zoo malware dataset.	access to the program's source code. Additionally, the method is computationally efficient and can be applied in real-time.	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.	The malware detection technique has demonstrated high efficiency in identifying both known and unknown malware in cloud computing environments with a low rate of false positives. 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	propose a method for verifying data integrity in cloud storage services	The system utilizes hash functions for creating an exclusive digest of data blocks	Ensuring the authenticity and integrity of data stored in the cloud.	The proposed approach may have limitations due to its reliance on the API of the cloud storage

(Yindong Chen, Liping Li, Ziran Chen)	using digital signatures. The paper aims to address the challenge of ensuring the authenticity and integrity of data stored in the cloud.	saved in the cloud, while digital signatures are used to guarantee the authentication and integrity of the hash values produced by the hash functions. The management of digital certificates and public keys required for generating and verifying digital signatures is accomplished through the use of a PKI. 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.	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.	service, which may not be accessible or may lack the required features. 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 approach utilizes various techniques, such as data replication, hashing, Merkle tree, secret sharing, and adaptive fault detection, to ensure the integrity of data stored in cloud storage in an effective manner. The adaptive fault detection mechanism used in the methodology can detect both single and multiple	The approach efficiently guarantees the security and authenticity of cloud-stored data by incorporating several techniques, thereby enhancing its resilience and protection. The adaptive fault detection mechanism used in the methodology can detect both single and multiple intrusions in cloud data, providing an	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

		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.	added layer of security. The methodology is flexible and can be customized to meet the specific requirements of different cloud storage systems.	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)	The authors suggest a solution for preserving the authenticity of data stored in cloud storage by utilizing 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	ensuring data integrity in the cloud environment using encryption algorithms.	The AES encryption algorithm is utilized for ensuring data integrity in the cloud environment.	The data stored in the cloud environment is highly secure. The use of encryption algorithms ensures that the	The computational burden linked with the utilization of encryption algorithms and hash functions.

(P.Varalakshmi , Hamsavardhini Deventhiran)		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.	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.	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.	Further research and experimentation are required to assess 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,	propose a solution for ensuring data integrity in cloud computing environments using blockchain technology. The paper introduces a	The proposed system, which utilizes a blockchain-based approach for data integrity verification, has been successfully implemented and offers several	The proposed system offers a secure and tamper-proof way to verify data integrity in a cloud computing environment, along with increased	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

Prateek Kulkarni, Vishnudas M V)	blockchain-based database system that provides a tamper-proof, secure, and efficient method for data integrity verification in cloud computing.	advantages over traditional methods, including enhanced security, transparency, and tamper-proofing.	transparency and decreased dependence on central authorities. The system also provides efficient and cost-effective data integrity verification, which can benefit organizations that deal with large amounts of data.	with many transactions. Moreover, the research paper lacks a comprehensive analysis of the proposed system's performance, and additional investigation is necessary to enhance its efficiency.
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.	The proposed method offers several benefits, including the capability to achieve real-time data integrity verification while maintaining a low computational overhead, its decentralized and tamper-proof nature, and its capacity to provide an effective and scalable solution for data integrity verification in cloud storage.	The practicality and performance of the suggested scheme may be impacted as it has not been put into practice and tested in an actual cloud storage system. Furthermore, the research paper lacks an extensive evaluation of the security and privacy implications of the proposed scheme, indicating a potential research area that needs further exploration
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	The proposed system can bring several advantages, including improved efficiency, transparency, security, and reduced costs. The use of blockchain technology can eliminate intermediaries, reduce	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

		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.	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.	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. Sivananthams, 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 behaviour 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.

References

Ahmed, M., 2022. File Integrity Checkers: Functionality, Attacks, and Protection. *2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2).*

Bander, A., 2017. Lightweight behavioral malware detection for windows platforms. *2017 12th International Conference on Malicious and Unwanted Software (MALWARE).*

Chao, S., 2019. Block Chain-Based Data Audit and Access Control Mechanism in Service Collaboration. *2019 IEEE International Conference on Web Services (ICWS).*

Cristian, G., 2010. Malware detection based on Cloud Computing integrating Intrusion Ontology representation. *2010 IEEE Latin-American Conference on Communications.*

Danyang, B., 2010. Design and implementation for MD5-based data integrity checking system. *2010 2nd IEEE International Conference on Information Management and Engineering.*

Gopalan, C., 2005. Ensuring data integrity in storage. *Proceedings of the 2005 ACM workshop on Storage security and survivability.*

Jambulingam, A., 2019. An Adaptive Methodology for Integrity Checking in Cloud Storage. *International Journal of Engineering and Advanced Technology,* 8(6), pp. 4470-4475.

Jerzy, M., 2008. Modern approaches to file system integrity checking. *2008 1st International Conference on Information Technology.*

Kambiz, K., 2021. Cloud Based Malware Detection Through Behavioral Entropy. *2021 IEEE International Conference on Big Data (Big Data).*

Kang, H. K. K. &. K. J., 2019. Flow-based Malware Detection Using Convolutional Neural Network. *Journal of Information Processing Systems,* pp. 95-105.

Misbah, A., 2020. Security of IoT Using Block chain: A Review. *2020 International Conference on Information Science and Communication Technology (ICISCT).*

Pejman, D., 2021. NLP-based Entity Behavior Analytics for Malware Detection. *2021 IEEE International Performance, Computing, and Communications Conference (IPCCC).*

Priya, S. S., 2023. Review on Malware Classification and Malware Detection Using Transfer Learning Approach. *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT).*

Ravishankar, P., 2020. Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments. *2020 International Conference on Mainstreaming Block Chain Implementation (ICOMBI).*

Sanjeev, Y., 2016. Semantics-Based Online Malware Detection: Towards Efficient Real-Time Protection Against Malware. *IEEE Transactions on Information Forensics and Security,* 11(2), pp. 289-302.

Sivanantham, S., 2022. Reliable Data Storage and Sharing using Block chain Technology and Two Fish Encryption. *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA).*

Suchetha R, P. ,. S. ,. A., 2020. Survey on Data Integrity and Verification for Cloud Storage. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT).*

Varalakshmi, H., 2012. Integrity checking for cloud environment using encryption algorithm. *2012 International Conference on Recent Trends in Information Technology.*

Vidhi, R., 2019. Malware Detection based on API Calls Frequency. *2019 4th International Conference on Information Systems and Computer Networks (ISCON).*

Vinothiyalakshmi, M., 2022. Digitized Land Registration Using Blockchain Technology. *Blockchain Technology,* pp. 73-86.

Wu, P.-x., 2011. Behavior-Based Malware Analysis and Detection. *2011 First International Workshop on Complexity and Data Mining.*

Yeo, K., 2018. 2018 International Conference on Information Networking (ICOIN). *2018 International Conference on Information Networking (ICOIN).*

Yindong Chen, L., 2017. An Approach to Verifying Data Integrity for Cloud Storage. *2017 13th International Conference on Computational Intelligence and Security (CIS).*

Yiran, H., 2022. A Blockchain-Based Efficient Data Integrity Verification Scheme in Multi-Cloud Storage. *IEEE Access,* Volume 10, pp. 105920-105929.

Zhenpeng, Y., 2022. Data Integrity Audit Scheme Based on Blockchain Expansion Technology. *IEEE Access*, Volume 10, pp. 55900-55907.