SN Computer Science (2024) 5:90
https://doi.org/10.1007/s42979-023-02337-4

ORIGINAL RESEARCH

Multi-level Data Integrity Model with Dual Immutable Digital Key Based Forensic Analysis in IoT Network

J. V. N. Raghava Deepthi[1] · Ajoy Kumar Khan[2] · Tapodhir Acharjee[1]

Received: 13 June 2023 / Accepted: 18 September 2023
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2023

Abstract

Over the last decade, the proliferation of Internet of Things (IoT) devices has risen dramatically. The exponential growth of IoT device ecosystems has led to a rise in the risks and cybercrimes associated with the IoT. Because of their malleability, IoT devices are more susceptible to persistent attacks. Forensic investigation of attacks on IoT devices can be challenging for security experts due to the limited processing and memory capabilities of these devices. The existing methods are often complex for IoT environment and providing moderate results. The primary objective of this research is to determine and recommend a secured framework for the IoT that maintains data integrity. This research proposes a simple scheme called Multi-level Data Integrity Model with Dual Immutable Digital Key based Forensic Analysis (MLDIM-DIDKbFA) for securing the IoT data. A digital key is generated which is used for validation of nodes during data transmission. Cyber-attack detection Forensic analysis due to possible cyber-attack is performed if any violation to data integrity is detected in IoT network. The proposed model was contrasted with two state-of-the-art existing models Binary Classifiers for Data Integrity Detection in Wearable IoT Edge Devices (BCDTED) and Distributed Estimation against Data Integrity Attacks in IoT Systems (DEA-DIA). The parameters used for comparison are time and accuracy for node registration, immutable key generation, intermediate node verification, multi-level data integrity verification, forensic analysis etc. It is found that our proposed method is providing better results compared to the other two existing systems. As a result, the proposed forensic system boosts the effectiveness and credibility of IoT environment forensics.

Keywords Internet of Things · Forensics · Data security · Digital key · Cyber-attacks · Data confidentiality · Data integrity

Introduction

The IoT refers to a network of computers and other electronic gadgets that may exchange information and resources safely through the web. The IoT is superior to existing networks because it requires fewer human interactions, provides a broader context, and can be easily expanded [1].

Home automation, wearable tech, smart firefighting, smart metering, improved production, and intelligent structures are just a few examples of the many innovations made possible by the widespread use of IoT [2]. The security of IoT devices is still an issue, despite the fact that their use cases are expanding all the time. Manufacturers of IoT gadgets are more concerned with making their products more appealing to consumers by adding new features and functionalities and streamlining their designs to make the gadgets smarter and more cost-effective. Inadequate protections have led to a rise in cyber-attacks on Internet of Things gadgets in recent years.

This article is part of the topical collection "SWOT to AI-embraced Communication Systems (SWOT-AI)" guest edited by Somnath Mukhopadhyay, Debashis De, Sunita Sarkar and Celia Shahnaz.

✉ J. V. N. Raghava Deepthi
deepthijonnalagadda16@gmail.com

Ajoy Kumar Khan
ajoyiitg@gmail.com

Tapodhir Acharjee
tapacharjee@gmail.com

[1] Department of Computer Science and Engineering, Assam University Silchar, Silchar, Assam 788011, India

[2] Department of Computer Engineering, Mizoram University, Aizawl, India

Published online: 12 December 2023

Smart cities, smart homes, smart healthcare, etc. are just a few examples of where the IoT is gaining traction. As a result of this increased connectivity, several novel uses have been developed for IoT devices. The downside is that IoT devices may be used in public places or even dangerous areas [3], making them vulnerable to a wide range of threats. Because of their inherent simplicity, IoT devices are frequently compromised. Large volumes of data generated by many IoT devices could be utilized to control vital industrial facilities, wearable medical equipment, traffic signals [4], and so on. Data manipulation attacks are among the most damaging that an adversary can execute against an IoT device [5]. An adversary's goal in such an assault is to alter IoT data in a way that causes the system to malfunction and lead to bad control decisions. Incorrect temperature readings, for instance, could lead to the control unit of a factory arbitrarily switching on and off the cooling system, potentially resulting in serious damage to the equipment and even injuries to the workers. As a result, assaults on the IoT that involve data tampering can result in substantial economic loss, damages to infrastructure, and even human injury [6]. This research proposes an integrity detection method to identify data manipulations in IoT devices as a solution to this problem.

To keep data accurate and comprehensive is to ensure its integrity. However, there are several ways in which messages might go wrong during wireless transmission in IoT applications [7], including attenuation, distortion, and the introduction of noise [8]. When there is an error, the receiver is unable to accurately decode the signal and obtain the intended symbol. Error-correcting codes, often known as channel coding [9], are necessary for data security. Error-correcting codes guarantee reliable operation of IoT infrastructure. They protect the reliability of communication channels even when environmental factors such as noise, deformation, and attenuation are present [10]. One of the easiest and most well-known error-detection systems in digital communication uses the parity bit. Information is divided up into chunks [11]. Each block has an extra bit added to it so that the sum of the 11 bits already present in the block, plus the extra bit, adds up to an even number. If there is even a single bit error in the block, the number of ones will be off. Consequently, this enables the isolation of individual mistakes.

Limitations in processing power and memory mean that only a small subset of possible instruction sets can be executed by IoT devices [12]. Therefore, they can't record, track, and analyze data sent by IoT gadgets. Because of this, forensic investigation of attacks on IoT devices has proven challenging for security researchers. Because of these constraints, gathering evidence might be difficult during a forensic investigation [13]. Enhancing the network's resilience and security in an IoT environment calls for specialized tools and methods. More powerful forensic procedures need to be developed and used for the research and examination of IoT devices [14]. The forensic analysis method is an effective tool for avoiding the aforementioned problems. To automate the process of detecting attacks on IoT devices and creating the associated logs and alarms, a forensic analysis framework is proposed in this research. To establish the perpetrator, motive, and effects of a security breach, a thorough post attack investigation known as a forensic analysis is conducted. It is similar to Security Incident Management (SIM) [15], in which security events on a network are identified and then acceptable actions are taken to accommodate for compromised security standards. Network auditing is a pre-examination of the vulnerabilities in a network, while forensic analysis is a post-study of the security breaches that documents how and when something happened [16].

A remote logging server circumvents the framework's data gathering constraints. To facilitate forensic investigation, communication from IoT devices is diverted to a logging server where alarms and logs of malicious attack traffic [17] are generated and kept. A forensic server generates new copies of these logs and analyses them for clues about assaults and their perpetrators [18]. The four phases of any forensic analysis are data collection, inspection, analysis, and documentation and reporting. Information pertinent to a certain assault is gathered in the course of data collecting [19]. The main issue was data collecting, which was hampered by the limited processing capability of IoT devices. When attacks were suspected, no evidence was ever located [20]. The proposed solution employs a monitoring node in the network that maintains logs of malicious traffic and generates alerts to receiving nodes for detecting attacks. The cloud based data integrity verification model is shown in Fig. 1.

To aid with IoT forensic analysis, the proposed system combines a machine learning model with specialized forensic analysis reporting model. In addition to assisting in the creation of rules for attack detection, the forensic server can also provide a fresh alarm whenever malicious traffic is detected in the network [21]. These attack rules are considered into the proposed system. Automated defense against attacks is provided by machine learning. After these steps have been completed, several reports detailing the type, frequency, and potential responses to attacks are generated [22]. With this forensic information, the full picture of the attack can be analyzed and the perpetrators can be tracked down.

The forensic analysis process initially considers the evidence identification process from public datasets and then collection of samples is performed. The data collected will be analyzed for forensic process and then the attribute ranges are documented that is used for future processing. The presentation is performed to use the values for analysis for integrity verification. There are four main steps in the forensic process: finding prospective evidence, collecting

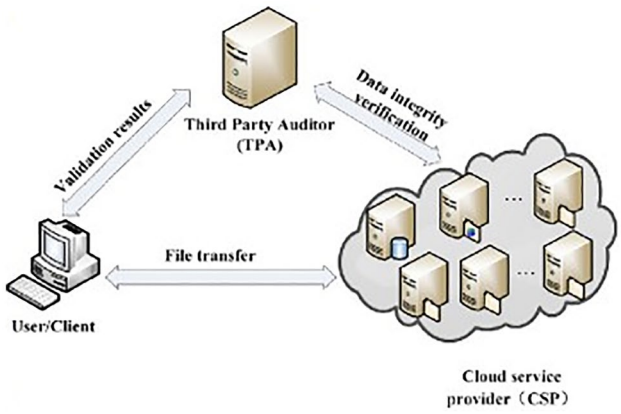

Fig. 1 Cloud based data integrity verification model

it, analyzing it, and writing a report. The forensic analysis process is shown in Fig. 2.

It takes a lot of data analysis and intelligent computation to spot threats and attacks in an IoT environment. To detect threats, these platforms make use of cutting-edge computer systems based on machine learning and smart computing. To discover and reveal the presence of adversaries, digital forensics requires extensive data analysis, such as retrieving and authenticating system logs, assessing information stored in blockchains, etc. To facilitate virtualized resource sharing, it collects and analyses data from access and system logs using blockchain technology. In the early stages, adversary classification and differentiation are aided by the management

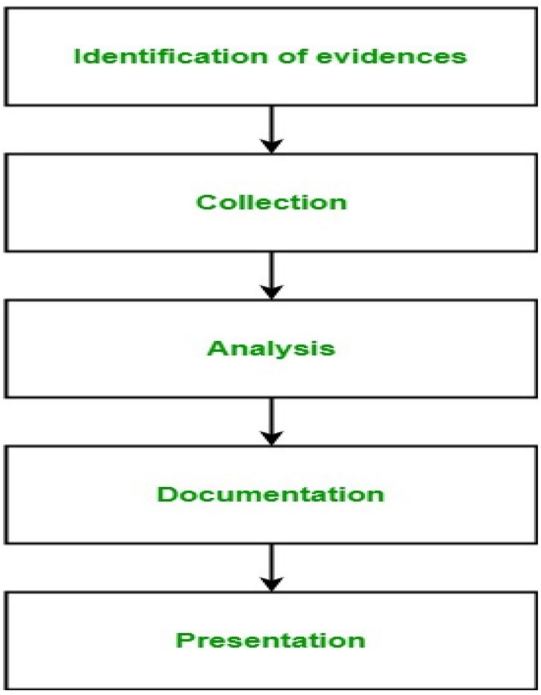

Fig. 2 Forensic analysis process

of system-related records, audit systems, and access controls. This aids in isolating the source of the attack and preventing it from spreading to other systems. Paradigms in machine learning facilitate the differential appraisal and examination of exact information over time and with less complexity. Regular testing and training are carried out over a wide range of data collected from the IoT ecosystem to detect the presence of such threats. Information analysis and consequences based on trust, authorization, and authentication are crucial in making security-related decisions. This research proposes a Multi-level Data Integrity Model with Dual Immutable Digital Key based Forensic Analysis for securing the IoT data.

For digital forensics to be admissible in court, evidence management must adhere to strict legal criteria. The acquired evidence must be shown to be authentic and unaltered. The investigation of computer-related crimes requires the use of specialized computer forensics software and associated toolkits in accordance with generally acknowledged procedures and criteria. By its very nature, digital evidence is fragile, and any mistakes in its treatment or investigation render it unusable as evidence. Because digital evidence can be altered easily, it must be collected, preserved, and documented with the utmost care. Investigators in the field of computer forensics must operate ethically because their work will be scrutinized by a court of law if the case ends up there.

Any system that stores, processes, or retrieves data must be designed, implemented, and used with the utmost care to ensure data integrity during the whole duration of the data's life cycle. Even within the same broad field of computers, the phrase might have wildly varied meanings depending on the exact situation. Data integrity is the safeguarding of data against unauthorized alteration [23]. Data privacy refers to the protection of personally identifiable information while it is accessible to the public. Any group or individual can benefit from adhering to strict data privacy guidelines. Information on a person's life and circumstances is necessarily limited. Users may choose whether or not to share the information. There is little room for privacy for anyone if there is no system in place to protect it. There is a clear distinction between data security, which is generally understood to involve protecting and preserving the information users provide from other unknown persons, and data privacy, which is the act of determining who has access to the data and for what purpose.

Literature Survey

Arlene John et al. [24] evaluated and contrasted a number of AI-based binary classifiers for verifying the authenticity of data collected by IoT-enabled wearable sensors. The amount of information saved and sent can be reduced by detecting data corruption at the network's periphery and

then removing it. As a result, IoT devices can function with less memory and less electricity. In this paper, we look at a number of machine learning-based classifiers for validating ECG data. The feature vectors are computed using Signal Quality Indices (SQIs) that are low-complexity measures of kurtosis and skewness.

The IoT is a cutting-edge innovation that has the potential to revolutionize many different markets by enabling real-time data collecting to boost productivity while cutting costs. The IoT is helping Maritime Transportation Systems (MTSs) prevent ship collisions, boost shipping efficiency, and cut down on revenue loss for harbors and shipyards. IoT-enabled MTS create a vast quantity of real-time data that, when paired with previous data, that was used to efficiently anticipate the future trajectories and concentrations of vessels on the sea. However, the MTS marine traffic data cannot be handled efficiently using conventional big data analysis techniques, and its validity must be verified before it can be used for purposes such as the prediction of vessel paths and high-density zones. Liu et al. [25] designed a data integrity checking scheme for IoT-enabled MTS that is both adaptable and capable of restoring original data. Erasure coding is used to encode vessel data blocks in the proposed approach.

For generic IoT applications, Wu et al. [26] provided a safe distributed estimation approach that is immune to data integrity assaults. A resilient optimal estimation target for protecting the entire IoT system is constructed by capitalizing on the attackers' spatial sparsity. A phony data processor is then created to mitigate the negative outcomes of the assaults. The method's convergence is examined. It demonstrates that under practical conditions wherein all communication connections and sent data are arbitrarily compromised the estimation error will converge to be uniformly constrained for all circumstances. To back up theoretical findings, the author also gave simulation results using a model Internet of Things network consisting of 50 nodes and 145 edges.

As MTSs that take advantage of the IoT continue to evolve, it will become increasingly important to not only store the huge amounts of data created by these systems in a cost-effective and dependable manner, but also to analyze this data as soon as possible. Users can save their information in the Cloud-based Maritime Transportation Systems (CMTS) without having to worry about factors like cost, storage space, physical location, etc. However, CMTS also raises significant security concerns, the most pressing of which is the integrity protection of outsourced data, which is essential to the security, dependability, and efficiency of shipping channels. To address this issue, Li et al. [27] provided a method of auditing CMTS data for integrity that is both dynamic and based on the user's identification. By conducting audits in batches, this system reduces the administrative load associated with key management and boosts auditing efficiency. This approach not only eliminates the communication overhead of the auditing phase, but also has the lowest computing cost across all entities, as demonstrated by a comparison of its performance with that of similar schemes.

Threatening cyber-assaults against the IoT-based smart grid include data integrity attacks (DIA). An attacker has a difficult time obtaining or inferring the branch parameters. Time and circumstance can alter or upset them. Zhang et al. [28] developed the whole category of DIA by designing the zero-parameter-information DIA (ZDIA), which allows the attacker to carry out covert data tampering attacks without knowing the parameters of the branches being targeted. Such an attack can be built with only the cut line's topology information. In addition, the author broaden the scope of ZDIA to include scenarios in which a bus or super-bus has only a few cut lines leading to the outside world.

Yazid et al. [29] discussed a fresh authentication strategy for IoT-based vehicle monitoring systems. The proposed technology, which is based on parallel hash chains, is well suited for low-cost and power-efficient IoT gadgets. The need to send secret keys over the network is eliminated since encryption keys are continuously created on parallel hash chains on both the IoT device and the server. Two transmission handshakes are all that are needed for identification and data transmission with the suggested technique. It eliminates the need for on-device random number generation, which is both hardware intensive and a possible security risk in IoT devices.

Cryptographic hash functions have the critical property of being unable to distinguish between two files if even a single bit of the input is altered. However, a hash function that preserves similarity is essential in computer forensics since it allows for the discovery of previously unknown material. Forensics investigators are having a difficult time figuring out how to use data from these gadgets. For efficient application management, Mahrous et al. [30] introduced a blockchain-based IoT computer forensics architecture that employs both the conventional hash for authentication and the fuzzy hash in order to build the Blockchain's Merkle tree. When compared to traditional hashing methods, fuzzy hashing increases the likelihood that damaging information may be uncovered.

When it comes to the IoT and its capacity for facilitating smart mobility, the Internet of Vehicles (IoV) has emerged as a crucial data sensing and processing platform. Users of the IoV and law enforcement authorities benefit from the combined efforts of both the cameras installed in vehicles and those stationed along the roadways. To provide these forensic services effectively, it is crucial to ensure that data flow between vehicles is both secure and private. In this research, Zhang et al. [31] presented an incentive authentication

scheme (LIAS) that is both lightweight and practical for use in IoV forensic services. The layers of LIAS's architecture are the cloud, the fog, and the user. The privacy and security concerns around forensic services in IoV for ITS inspired this research. The purpose is to strengthen vehicle security and privacy without sacrificing the convenience and efficiency of data sharing across vehicles. To make the most of the capabilities of near-user edge devices and the links between fog nodes and devices, fog-assisted IoV is introduced. The challenges of protecting the privacy and security of automobiles persist, though. Furthermore, information diffusion in automobiles could be easily tracked due to the inherent flaw of wireless communication.

As the amount of data transmitted by email continues to rise, investigators are faced with the formidable issue of extracting the necessary semantic information from the massive amounts of emails, which slows down the investigation. The offender now has an advantage when trying to cover their tracks. Existing keyword-based search algorithms and filtering frequently result in irrelevant, short-sequence emails that bypass important content. To address the aforementioned shortcoming, Hina et al. [32] offered a novel efficient method for multiclass email classification called SeFACED, which makes use of Long Short-Term Memory (LSTM) based Gated Recurrent Neural Network (GRU). SeFACED can process long dependencies of 1000 + characters, not just short ones. By comparing its results to those of more conventional machine learning methods, deep learning models, and state-of-the-art research, SeFACED is able to fine-tune the parameters of LSTM-based GRUs for optimal performance.

The increased use of encryption technology in recent years has presented significant hurdles to computer forensic investigation by making it easier for criminals to conceal damaging data from security regulatory bodies. As a result, research into methods for detecting and analyzing encrypted data is essential. In this study, Li et al. [33] offered an approach to decryption that uses deep convolutional neural networks. The unprocessed information is initially transformed into two-dimensional matrices for use as the network's input. Then, representative features are provided as the input of succeeding layers using the multiscale extraction of features process with different activation functions. The next step is to use the residual learning operation to improve feature discrimination. This method is used to build a network that can automatically learn the global context of encrypted data by extracting it. The proposed technique also reliably identifies encrypted data using a variety of algorithms.

There has been a dramatic rise in the amount of cyber assaults targeting IoT environments recently. The human and monetary costs at all levels of the Internet of Things were high as a result of this. The occurrences of attacks that have attacked the IoT system or its components have become increasingly difficult to identify as cybercriminals have been using anti-forensics activities and deploying strategies and tools to mask their tracks. As a result, the frequency and severity of cyber-attacks against the IoT are both increasing, leading to attacks that are both more efficient and more sophisticated. Conventional safety and forensics solutions, especially in terms of obtaining evidence for attack investigation, are insufficient to prevent and analyze such cyber-attacks. Therefore, there is a pressing want for clearly defined, sophisticated, and sophisticated forensics investigation methodologies to foil anti-forensics methods and identify and apprehend cybercriminals. Jean-Paul A. Yaacoub et al. [34] discussed the rise of anti-anti-forensics as a new forensics defense mechanism against anti-forensics operations and covers the many forensics and anti-forensics approaches that can be implemented in the IoT sector, including tools, techniques, types, and problems. Forensics investigators would benefit from knowing the various anti-forensics tools, methodologies, and techniques used by cybercriminals.

Combining AI with other technologies can boost their efficiency. Smart IoT refers to Internet of Things gadgets that also incorporate artificial intelligence. Wearable devices allow for remote control of smart Internet of Things gadgets. Sensors on wearable electronics like smartwatches and smartbands collect data about their users in order to tailor their services to them. Due to the fact that the generated data are saved in the wearable device's storage, accessing this data from the device can be helpful in solving crimes. Therefore, Kim et al. [35] offered a forensic paradigm for wearable devices that goes beyond indirect forensics and relies instead on direct interactions made by wireless or interfaces. The ecosystem of wearable gadgets served as inspiration for the forensic paradigm, which was then broken down into separate categories for digital and physical investigation. We tested the forensic model on wearables from Samsung, Apple, and Garmin to ensure its versatility.

The literature survey analyzed numerous forensic models for data integrity and also key handling models are analyzed. Based on the analysis done, there are some limitations identified in the traditional models like less key size, using keys for multiple times and easily cracking of keys. The integrity violations are also made even stronger models are designed. The performance levels of the traditional models can be enhanced using strong cryptography models and accurate authentication models that can enhance the performance levels of the cryptography and data integrity models for forensic analysis.

Proposed Method

Traditional digital forensics makes it simpler to track down and identify hacked devices that may contain useful forensic evidence. However, the variety and unique qualities of IoT

devices make forensics an uphill conflict. Including smart appliances, smart meters, smart hubs, virtual assistants, and various wearables, there can be up to seventeen separate possible evidence sources in a modern smart home. In addition, the fluidity of the IoT ecosystem causes borders to blur as devices are continually moving in and out of a particular network, either automatically or because the user has physically relocated them. The devices' mobility across many networks makes it difficult to demarcate cases.

The proposed model data integrity verification and forensic analysis for cyber-attack detection is shown in Fig. 3.

Initially to perform data transmission the nodes in the IoT has to register. After registration for each and every node a digital key will be generated. The digital key will be used only for one time. with the help of digital key the nodes will be authorized. So attacker nodes cannot act as a normal node, as attacker nodes are not provided with digital key. After digital key is generated a random node is selected for data transmission and that node undergoes verification process to prove its authenticity. At multilevel, after the node authentication the data that is transmitted will be verified. Once the verification process completes then the forensics analysis starts if there is any attack on the data or the node. Finally, the list of attackers will be generated based on the attackers within the network.

Any system that stores, processes, or retrieves data must be designed, implemented, and used with the utmost care to ensure data integrity during the whole duration of the data's life cycle. Even within the same broad field of computers, the phrase might have wildly varied meanings depending on the exact situation. Data integrity is the safeguarding of data against unauthorized alteration. Data privacy refers to the protection of personally identifiable information while it is accessible to the public. Any group or individual can benefit from adhering to strict data privacy guidelines. Information on a person's life and circumstances is necessarily limited. Users may choose whether or not to share the information. There is little room for privacy for anyone if there is no system in place to protect it. There is a clear distinction between data security, which is generally understood to involve protecting and preserving the information users provide from other unknown persons, and data privacy, which is the act of determining who has access to the data and for what purpose.

This research proposes a Multi-level Data Integrity Model with Dual Immutable Digital Key based Forensic Analysis (MLDIM-DIDKbFA) for securing the IoT data.

The following algorithm provides the pseudocode for Data Integrity verification and Forensic analysis for cyber-attack detection.

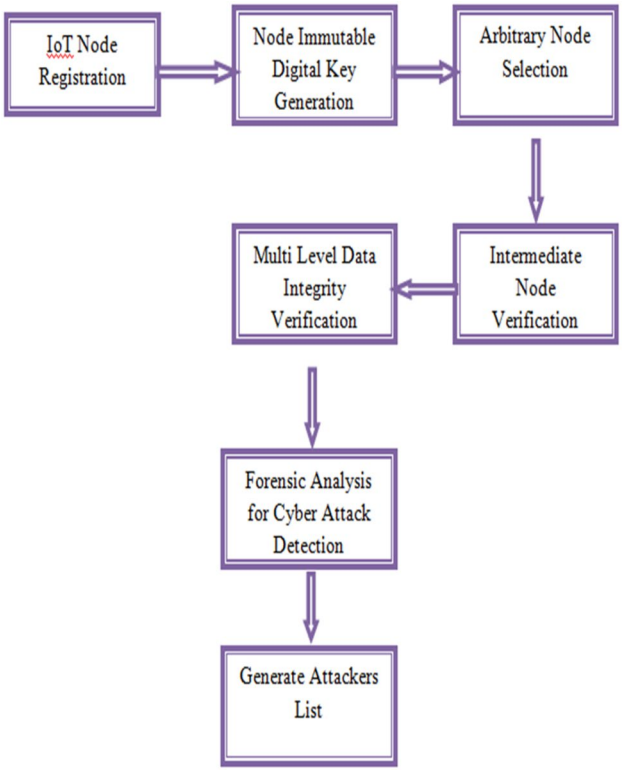

Fig. 3 Proposed framework

Algorithm

Input: Nodes in Network[N= N1,N2....Nm]
Output: Attackers List
step 1: Consider a set of nodes $N=\{N_1,N_2,\ldots\ldots,N_m\}$
step 2: $nodeaddr(n) = getlogaddr(n) \in getphaddr(n)$
step 3: for n in range(1,M)
$NregSet[M] = nodeaddr(n) + timeInst(n) + \frac{getnoderange[n]}{maxNodeaddr(n)} + Th$
step 4: $TimeS \leftarrow getTime(MMSS)$
step 5: $Rval \leftarrow rand(1, getnoderange[M])$
step 6: $Min \leftarrow input(n)\ where\ len(n) > 5$
step 7: $Max \leftarrow input(n) where\ n > min$
step 8: $Ival \leftarrow getPrime(Min, Max)$
step 9: $Sval \leftarrow Input(n) where\ n < Rval\ and\ n > Ival$
step 10: for n in range(1,M)
$Ikset = \frac{Rval \oplus Sval}{Rval || Ival}$
end for
step 11: n=1
while(n<=M)
$Dkey[M] = \frac{Ikset \,\&\&\, val}{Sval \oplus Ikset} \ll 2$
step 12: $Dkey[Status] \leftarrow 1 if (getTime(MMSS) < TimeS + 15$
step 13: for n in range(1,M)
$Anode[M] = Node(NregSet(n)) + \max(\mu(n)) + \max(\delta(n))$
step 14: for n in range(1,M)
$NValid[M] = Node(NregSet(n)) + Dkey(n)$
$\leftarrow Anode(Dkey(n))\begin{cases} set\ T \leftarrow active\ if\ Dkey[Status] == 1 \\ T \leftarrow deactivate \qquad Otherwise \end{cases}$
for n in range(1,L)
$Kreq[L] = getKey(NregSet(n)) \rightarrow AN(Dkey)\begin{cases} if\ Dkey[Status] == 0 \end{cases}$
end for
end for
step 15: n=1
while(n<=M)
$Dintegrity[M] = \frac{simm(D(NregSet(n)), D(NregSet(n+1))}{\lambda} +$
$min(NregSet(D))\begin{cases} set\ Int \leftarrow Max\ if\ NValid == 1\ and\ D(simm) == G\ and\ key \in DKey[M] \\ set\ Int \leftarrow Norm\ if\ NValid == 1\ and\ D(simm) < G\ and\ key \in DKey[M \\ dissimilar \qquad otherwise \end{cases}$
step 16: for n in range(1,M)
$ATKset[M] = \frac{NregSet(n)}{noderange[M]} + \begin{cases} Node(NregSet(n)) \leftarrow Nodeaddr(n) if (Int \leftarrow dissimilar\ and\ Norm) \\ Node(NregSet(n)) \leftarrow if(nodeadrr(n)) \notin NregSet[M] \end{cases}$
end for

Consider a set of nodes $N = \{N_1, N_2, \ldots, N_m\}$ where there can be m number of IoT nodes in a network. Initially the node registrations are performed where each node information is maintained by the network manager for further communication. The node registration is performed as

$$\text{nodeaddr}(n) = \text{getlogaddr}(n) \in \text{getphaddr}(n) \tag{1}$$

$$\text{NregSet}[M] = \sum_{n=1}^{M} \text{nodeaddr}(n) + \text{timeInst}(n) + \frac{\text{getnoderange}(n)}{\text{nextNodeaddr}(n)} + \text{Th} \tag{2}$$

Here nodearr() is the model used to consider the IoT node address and the node entry time for registration is considered using timeInst() of current node n and the maximum range of IoT network is considered using getnoderange() model. Th is the threshold value added during the registration for avoiding attackers to mislead the registration of nodes.

After each node set N_m are registered with the network, each node is assigned with a digital immutable key that cannot be altered in the network. The digital key is used for validation of nodes during data transmission. The digital node is used for only one time by a node. The immutable digital key generation is performed as

$$\text{Ikset} = \sum_{n=1}^{M} \frac{\text{Rval} \oplus \text{Sval}}{\text{Rval}||\text{Ival}} \tag{9}$$

$$\text{Dkey}[M] = \prod_{n=1}^{M} \frac{\text{Ikset} \,\&\&\, \text{Rval}}{\text{Sval} \oplus \text{Ikset}} \ll 2 \tag{10}$$

$$\text{Dkey}[\text{Status}] \leftarrow 1 \text{ if } (\text{getTime}(\text{MMSS})) < \text{Time}S + 15 \tag{11}$$

To monitor the data transmission and behavior of IoT nodes in the network, arbitrary node AN is selected that has best delivery rate and low delay levels. The arbitrary node selection is performed as

$$\text{Anode}[M] = \coprod_{n=1}^{M} \text{Node}(\text{NregSet}(n)) + \max(\mu(n)) + \max(\delta(n)) \tag{12}$$

Here μ is the node computational capability level, δ is the transmission success rate. The node whose computational capabilities and transmission rate are maximum is selected as AN node for monitoring the IoT network.

Each IoT Node N_i will be allowed to initiate data transmission only after validation. The IoT node validation during transmitting and receiving is performed by the AN node that is performed as

$$\text{NValid}[M] = \sum_{n=1}^{M} \text{Node}(\text{NregSet}(n)) + \text{Dkey}(n) \leftarrow \text{Anode}(\text{Dkey}(n)) \begin{cases} \text{set } T \leftarrow \text{active if Dkey[Status]} == 1 \\ T \leftarrow \text{deactivate} \quad \text{Otherwise} \end{cases}$$
$$\text{Kreq}[L] = \sum_{n=1}^{L} \text{getKey}(\text{NregSet}(n)) \rightarrow \text{AN}(\text{Dkey}) \{ \text{if Dkey[Status]} == 0 \tag{13}$$

$$\text{TimeS} \leftarrow \text{getTime}(\text{MMSS}) \tag{3}$$

$$\text{Rval} \leftarrow \text{rand}(1, \text{getnoderange[M]}) \tag{4}$$

$$\text{Min} \leftarrow \text{input}(n) \text{ where len}(n) > 5 \tag{5}$$

$$\text{Max} \leftarrow \text{input}(n) \text{where } n > \text{min} \tag{6}$$

$$\text{Ival} \leftarrow \text{getPrime}(\text{Min}, \text{Max}) \tag{7}$$

$$\text{Sval} \leftarrow \text{Input}(n) \text{where } n \langle \text{Rval and } n \rangle \text{Ival} \tag{8}$$

If T is active, then the sender can transmit and receive cans receive the data. The node validation is performed only when the key status is 1. Otherwise new key request of remaining L nodes is made to the AN node.

It is known that data integrity has been preserved when it is guaranteed to be free of corruption and readily available only by authorized parties. Data integrity, the maintenance and guarantee of accurate and consistent information across all communication channels so as to prevent attackers from modifying the data, must be taken into account in the development, execution, and maintenance of any system that maintains, processes, or retrieves data. The data integrity verification is performed as

$$\text{Dintegrity}[M] = \prod_{n=1}^{M} \frac{\text{simm}(\text{D}(\text{NregSet}(n)), \text{D}(\text{NregSet}(n+1))}{\lambda} + \min(\text{NregSet}(D)) \begin{cases} \text{setInt} \leftarrow \text{MaxifNValid} == 1 \text{and} D(\text{simm}) \text{and} \text{key} \in \text{DKey}[M] = & G \\ \text{setInt} \leftarrow \text{NormifNValid} == 1 \text{and} D(\text{simm}) \text{and} \text{key} \in \text{DKey}[M] < & G \\ \text{disimilar} & \text{otherwise} \end{cases} \tag{14}$$

The cyber-attack detection forensic analysis is performed if an attack is detected in the network causing violation to data integrity. The forensic analysis report is performed and the attack causing nodes ATK{A_1, A_2,…. A_N} are generated as

$$\text{ATKset}[M] = \sum_{n=1}^{M} \frac{\text{NregSet}(n)}{\text{noderange}(M)} + \begin{cases} \text{Node}(\text{NregSet}(n)) \leftarrow \text{Nodeaddr}(n) if (\text{Int} \leftarrow \text{disimilar and Norm}) \\ \text{Node}(\text{NregSet}(n)) \leftarrow if (\text{nodeadrr}(n)) \notin NregSet[M] \end{cases} \tag{15}$$

Experimental Results

This research proposes a Multi-level Data Integrity Model with Dual Immutable Digital Key based Forensic Analysis (MLDIM-DIDKbFA) for securing the IoT data. The proposed model is compared with the traditional Binary Classifiers for Data Integrity Detection in Wearable IoT Edge Devices (BCDTED) [25] and Distributed Estimation against Data Integrity Attacks in IoT Systems (DEA-DIA) [26]. Generally blockchain is used to maintain the security, but to create a block and to insert the data and update the data in a block it is a time consuming process and the time complexity will be increased. So the proposed model consumes less time for maintaining the data integrity, without using blockchain the proposed model is maintaining high data integrity levels when compared to the existing model. The proposed model when compared with the traditional models performs better in node registration time levels. The time taken for node registration of proposed model is less than the traditional models. The immutable digital key generation accuracy levels of the proposed model is high than the traditional models. The intermediate node verification accuracy levels is observed as high than the traditional models. The forensic analysis accuracy levels of the proposed model is high that reflects that the proposed model performance is high in multiple levels.

The proposed model also performs nodes registration to maintain nodes information to the network for node to node communication. The nodes can be easily recognized in the network. The Node registration time levels of the proposed and existing models are shown in Table 1 and Fig. 4. In the IoT network, each node establishes a wireless contact with other nodes for information transmission. Each node in the network have to register with the network administrator so that node information is used and node recognition will be done easily with the unique identity allocated after registration. The proposed model takes only 12 ms for registering 300 nodes and allocating unique identities for future communication. The time it consumes for registration is very less than the traditional models that consumes 18.6 and 21 ms, respectively for 300 nodes.

A immutable digital key is a key that is used by nodes in IoT for the verification to involve in communication that

Table 1 Node registration time levels

Nodes in the network	Models considered		
	Proposed MLDIM-DIDKbFA model	Existing BCDTED model	Existing DEA-DIA model
50	10.2	17	16.7
100	10.5	17.2	17.4
150	11	17.6	18.5
200	11.3	18.1	19
250	11.6	18.4	20
300	12	18.6	21

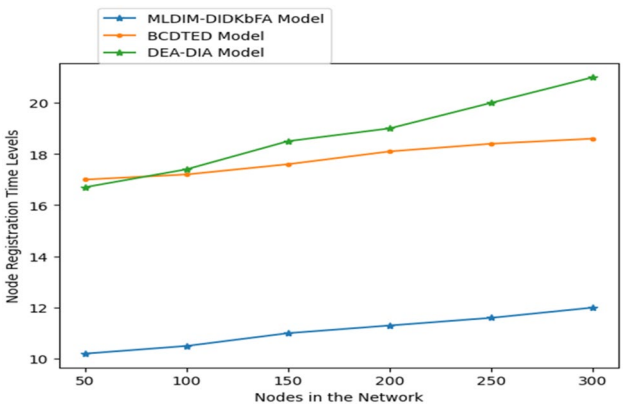

Fig. 4 Node registration time levels

Table 2 Immutable digital key generation accuracy levels

Nodes in the network	Models considered		
	Proposed MLDIM-DIDKbFA model	Existing BCDTED model	Existing DEA-DIA model
50	96.1	91	87.5
100	96.3	91.3	88.1
150	96.7	91.6	88.4
200	97.1	92	89
250	97.4	92.2	89.5
300	97.6	92.6	89.8

Table 3 Intermediate node verification accuracy levels

Nodes in the network	Models considered		
	Proposed MLDIM-DIDKbFA model	Existing BCDTED model	Existing DEA-DIA model
50	96.8	90	93.3
100	97.1	90.5	93.7
150	97.4	91	94.1
200	97.7	91.5	94.3
250	98	92	94.6
300	98.2	92.5	94.8

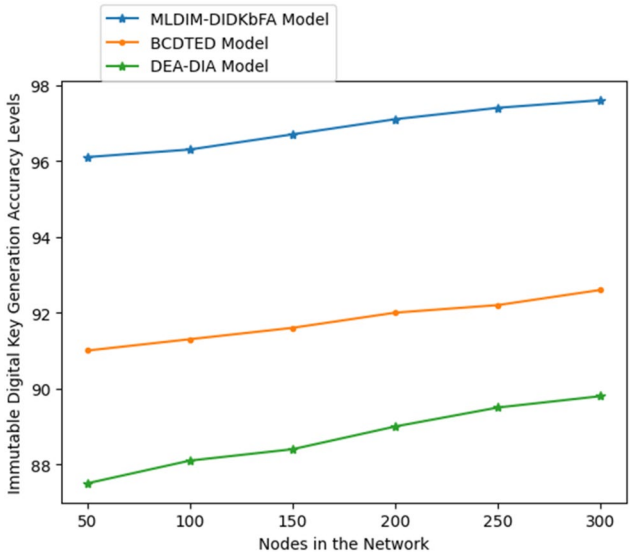

Fig. 5 Immutable digital key generation accuracy levels

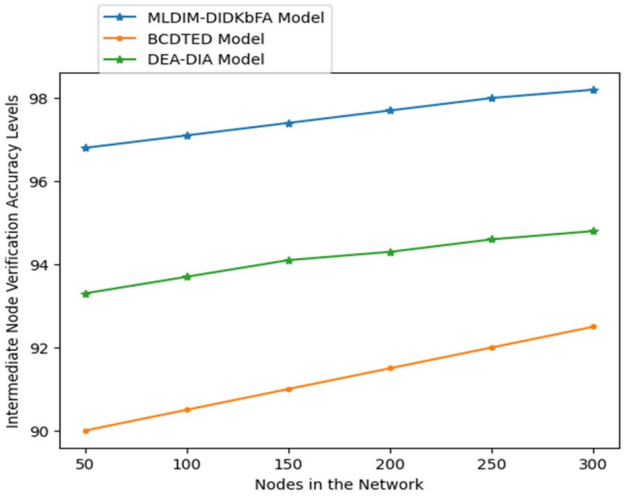

Fig. 6 Intermediate node verification accuracy levels

is used to avoid cyber-attacks in the network. The forensic investigation can be performed for the detection of attacks. For nodes validation, the immutable digital key is helpful. The Immutable Digital Key Generation accuracy levels of the proposed and traditional models are represented in Table 2 and Fig. 5. The proposed model generates immutable keys for the nodes in IoT network for authentication of nodes during transmission. The keys generated are strong and cannot be tampered. The Immutable digital key generation process achieved accuracy of 97.6% for generating keys for 300 nodes. The proposed model uses lightweight cryptography technique for this key generation that is strong and accurate. The traditional models achieved 92.6 and 89.8 percent accuracy that is very less when contrasted with the traditional models.

In IoT, each node will transmit the data to the neighbor nodes for successful data transmission. The nodes in the IoT can be authenticated for maintaining data integrity. The attack detection can be performed with node verification. Table 3 and Fig. 6 show the Intermediate Node Verification Accuracy Levels of the proposed and existing model. The proposed model performs intermediate node verification for analyzing node performance levels. The node performance metrics like loss, delay and transmission rate are analyzed and the proposed model achieved 98.2% accuracy in assessing 300 nodes in the IoT network. The traditional models achieved 92.5% and 94.8% respectively for 300 nodes that is less than the existing models. The integrity model performs better in verification of node performance levels.

There are a number of reasons why it's crucial to safeguard IoT data. One benefit of data integrity is that it guarantees data can be recovered and searched, as well as traced and connected. Stability, performance, and reusability are all boosted by safeguarding data accuracy and validity. Data integrity helps in avoiding modification of data with attacks. The multi-level data integrity verification time levels of the proposed and existing models are shown in Table 4 and Fig. 7. The proposed model performs multi-level data integrity verification

Table 4 Multi-level data integrity verification time levels

Nodes in the network	Models considered		
	Proposed MLDIM-DIDKbFA model	Existing BCDTED model	Existing DEA-DIA model
50	14.8	16.3	18
100	15.2	17	18.4
150	15.5	17.2	19
200	15.8	17.8	19.2
250	16.1	18.2	19.5
300	16.4	18.7	19.8

Table 5 Cyber-attack detection accuracy levels

Nodes in the network	Models considered		
	Proposed MLDIM-DIDKbFA model	Existing BCDTED model	Existing DEA-DIA model
50	97.3	92.6	87
100	97.6	93.1	88
150	97.8	93.6	89
200	98	94	90
250	98.3	94.5	91
300	98.5	95.4	91.2

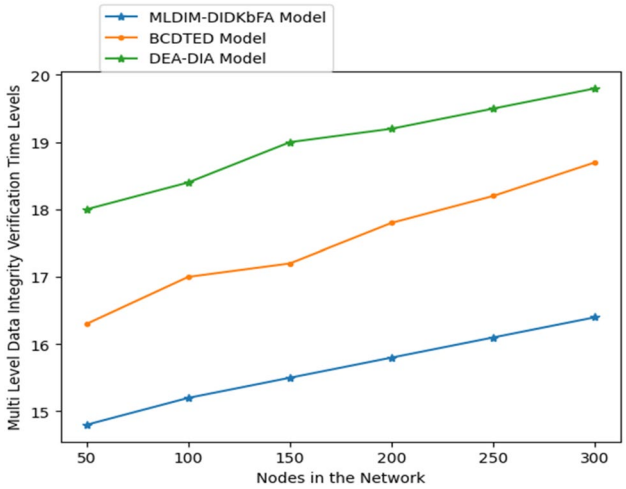

Fig. 7 Multi-level data integrity verification time levels

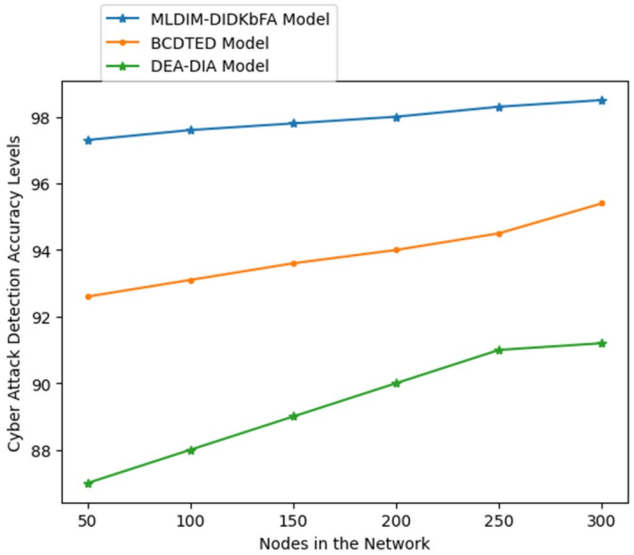

Fig. 8 Cyber-attack detection accuracy levels

for checking if they is any tampering or modifications in the forensic data that helps in strong integrity verification. The proposed model performs multi-level data integrity verification by assessing 300 nodes in only 16.4 ms. The time consumed by the proposed model for multi-level data integrity verification at each node is very less than the traditional models who achieved this in 18.7 and 19.8 ms, respectively.

A cyber-attack is any intrusion into a computer, network, or other electronic device with the intent of obtaining, altering, or deleting data. Malware, social engineering, and systems vulnerabilities are only some of the tools at the attacker's disposal. Any attempt to gain unauthorized access to a network, personal computer, or digital device with the goal to steal, expose, modify, disable, or damage data, applications, or other assets is considered a cyber-attack. The cyber attack detection accuracy levels of the existing and proposed models are shown in Table 5 and Fig. 8. The proposed model detects the cyber-attacks in the network with an accuracy of 98.5% for 300 nodes. The proposed model accuracy levels are very high than the traditional models who achieved 95.4% and 91.2% respectively. The proposed model analyzes each node attributes and change in the attribute ranges and detects the cyber-attacks accurately. Multiple attack patterns can be recognized in the proposed model.

Latency refers to the amount of time IoT network takes for a data packet to travel from its point of origin to its final destination. Latency is typically expressed in terms of milliseconds. A IoT power consumption decreases the longer it is dormant. This also means less chances for nodes to communicate with another and share data. Because of this, the gadget will run more slowly, a phenomenon called as latency. The latency levels of the proposed model is shown in Table 6 and Fig. 9. The proposed model uses light weight cryptography for key generations and performs multi-level integrity for each node in IoT network. The proposed model strategy in forensic analysis and data integration uses simple mathematical models that provides high security and low maintenance. The latency

Table 6 Latency levels

Nodes in the network	Models considered		
	Proposed MLDIM-DIDKbFA model	Existing BCDTED model	Existing DEA-DIA model
50	10.1	15.2	16
100	10.4	15.6	16.5
150	10.7	16	17
200	11	16.5	17.2
250	11.2	17.2	18
300	11.4	17.6	19

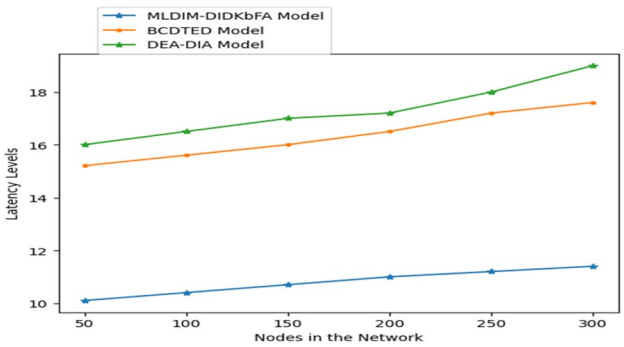

Fig. 9 Latency levels

Table 7 Forensic analysis accuracy levels

Nodes in the network	Models considered		
	Proposed MLDIM-DIDKbFA model	Existing BCDTED model	Existing DEA-DIA model
50	97.6	88	88
100	97.9	89.2	88.2
150	98.1	89.5	88.6
200	98.3	90	89
250	98.5	91	89.4
300	98.6	92	89.8

levels of the model is very less than the traditional models in which proposed model observes 11.4% delay levels that is very less than the traditional models that observes 17.6 and 19 percent latency levels, respectively.

When a security breach occurs or when a node of a network breaks the rules or the law, a forensic analysis is conducted to determine what happened and which node is turned as an attacker node, forensic analysis is frequently associated with the presentation of evidence to the court. The four main steps in any forensic investigation include

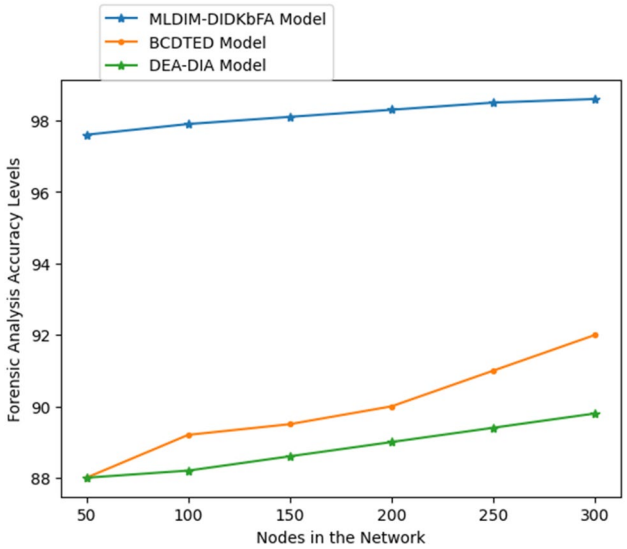

Fig. 10 Forensic analysis accuracy levels

collecting potential evidence, examining that evidence, writing up a report, and presenting the results. Table 7 and Fig. 10 represents the forensic analysis accuracy levels of the existing and proposed models.

Forensic analysis serves the overarching goal of analyzing, recovering, documenting, and preserving evidence for a given case. There are four steps involved in the data forensics process: collection, analysis, reporting, and presentation. Data forensic investigations may employ a number of different methods. Cross-drive analysis is one method used for this purpose since it can connect data found on different drives. There are a number of administrative and legal difficulties that investigators must contend with in addition to the more obvious technological obstacles. Attributing malicious conduct in cyberspace can be challenging due to the intricacies of cyber threats and attacks. There are many different standards for data forensics, but they are not universally accepted, and there is no central authority to guarantee that practitioners are competent and adhering to best practices.

The proposed model performs node authentication, multi-level node integrity verification and analyzes the node attributes and changes observed. The nodes that are causing cyber-attacks can be easily detected with the change in attributes and integrity violations. The proposed model performs node assessments at each level and observes 98.6% accuracy in forensic data analysis. The traditional models accuracy levels are very less that are observed as 92% and 89.8% accurate in forensic data analysis and reporting. The results shows that the proposed model performance in multiple aspects are high that represents that when the proposed model is made

available in real time forensic analysis, it will be useful in data analysis and achieves better accuracy in predictions.

Conclusion

Forensic analysis is the process of thoroughly investigating an attack after the fact to determine what motivated the attacker. The suggested forensic analysis solution gets around the constraints of IoT devices, such as their limited battery life and storage space. The suggested forensic system improves the efficiency and credibility of IoT device forensics in a direct-connected setting. Network traffic is diverted to the logging server and analyzed by comparing it to rules without disrupting connection between devices. The forensics server stores and can recreate these logs of malicious traffic using a variety of methods. When assaults are logged on IoT devices, not only are the logs recreated, but a dataset is also made. Data manipulation attacks are particularly dangerous because they can cause widespread disruption to an IoT system. An adversary's goal in such an assault is to alter IoT data in a way that causes the system to malfunction and lead to bad control decisions. Large volumes of private information are generated by IoT devices. Internet-of-Things devices, however, are open to cyber threats because they rely on the public Internet for data transport. It's possible that widespread harm and outages could emerge from an attack that tampers with or modifies data in order to disrupt data. Multiple reports are then generated to summarize the specifics of the attack, including the sort of attack, the frequency with which it was launched, and any potential next steps. With this forensic information, the full picture of the attack can be documented, and the perpetrators can be tracked down. The proposed model achieved 98.5% accuracy in data integrity verification and optimized forensic evaluation metrics can be applied. More assaults, categorized and sub-categorized, can be added to broaden the scope of this study. To further expand the reach of hybrid machine learning based forensic investigation, the dataset of everyday IoT devices can be utilized and Hash based MAC can also be applied.

Data availability Dataset generated or used during this work is available with the corresponding author and may be provided on reasonable request.

Declarations

Conflict of interest There were no conflict of interests in this manuscript.

References

1. Alam M, Khan E. Edge computing and its impact on IoT. Wesleyan J Res. 2021;14(7):211–22.
2. Xiaoshu Wang Z. Research on data integrity verification technology based on blockchain. J Phys: Conf Ser. 2017. https://doi.org/10.1088/1742-6596/2035/1/012017.
3. Garagad VG, Iyer NC, Wali HG. Data integrity: a security threat for internet of things and cyber-physical systems. International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2020. pp. 244–249. https://doi.org/10.1109/ComPE49325.2020.9200170.
4. Garg N, Bawa S, Kumar N. An efficient data integrity auditing protocol for cloud computing. Future Gener Comput Syst. 2020;109:306–16. https://doi.org/10.1016/j.future.2020.03.032.
5. Raut M, Sable R, Toraskar S. Internet of things(IOT) based smart grid. Int J Eng Trends Technol. 2016;34:15–20.
6. Shuklaa S, Thakur S, Hussaina S, Breslina JG, Jameel SM. Identification and authentication in healthcare internet-of-things using integrated fog computing based blockchain model. IEEE Internet Things J. 2021;15:100422. https://doi.org/10.1016/j.iot.2021.100422.
7. Liang G, Xin J, Wang Q, Ni X, Guo X. Research on IoT forensics system based on blockchain technology. Secur Commun Netw. 2022;2022:4490757. https://doi.org/10.1155/2022/4490757.
8. John A, Panicker RC, Cardiff B, Lian Y, John D. Binary classifiers for data integrity detection in wearable IoT edge devices. IEEE. Open J Circ Syst. 2020;1:88–99. https://doi.org/10.1109/OJCAS.2020.3009520.
9. Liu D, Zhang Y, Wang W, Dev K, Khowaja SA. Flexible data integrity checking with original data recovery in IoT-enabled maritime transportation systems. IEEE Trans Intell Transp Syst. 2023;24(2):2618–29. https://doi.org/10.1109/TITS.2021.3125070.
10. Wu H, Zhou B, Zhang C. Secure distributed estimation against data integrity attacks in internet-of-things systems. IEEE Trans Autom Sci Eng. 2022;19(3):2552–65. https://doi.org/10.1109/TASE.2021.3090416.
11. Li X, Shang S, Liu S, Gu K, Jan MA, Zhang X, Khan F. An identity-based data integrity auditing scheme for cloud-based maritime transportation systems. IEEE Trans Intell Transp Syst. 2023;24(2):2556–67. https://doi.org/10.1109/TITS.2022.3179991.
12. Zhang Z, Deng R, Yau DKY, Chen P. Zero-parameter-information data integrity attacks and countermeasures in IoT-based smart grid. IEEE Internet Things J. 2021;8(8):6608–23. https://doi.org/10.1109/JIOT.2021.3049818.
13. Yazid M, Fahmi F, Sutanto E, Setiawan R, Aripriharta, Aziz M. Simple authentication method for vehicle monitoring IoT device with verifiable data integrity. IEEE Internet Things J. 2022;10(8):7027–37. https://doi.org/10.1109/JIOT.2022.3228926.
14. Mahrous WA, Farouk M, Darwish SM. An enhanced blockchain-based IoT digital forensics architecture using fuzzy hash. IEEE Access. 2021;9:151327–36. https://doi.org/10.1109/ACCESS.2021.3126715.
15. Zhang M, Zhou J, Cong P, Zhang G, Zhuo C, Hu S. LIAS: a lightweight incentive authentication scheme for forensic services in IoV. IEEE Trans Autom Sci Eng. 2023;20(2):805–20. https://doi.org/10.1109/TASE.2022.3165174.
16. Hina M, Ali M, Javed AR, Ghabban F, Khan LA, Jalil Z. SeFACED: semantic-based forensic analysis and classification of e-mail data using deep learning. IEEE Access. 2021;9:98398–411. https://doi.org/10.1109/ACCESS.2021.3095730.

17. Li S, Liu P. Detection and forensics of encryption behavior of storage file and network transmission data. IEEE Access. 2020;8:145833–42. https://doi.org/10.1109/ACCESS.2020.3015080.
18. Yaacoub J-PA, Noura HN, Salman O, Chehab A. Advanced digital forensics and antidigital forensics for IoT systems: techniques, limitations and recommendations. IEEE Internet Things J. 2022;19:100544.
19. Kim M, Shin Y, Jo W, Shon T. Digital forensic analysis of intelligent and smart IoT devices. J Supercomput. 2023;79:973–97.
20. Barbierato L, Estebsari A, Pons E, Pau M, Salassa F, Ghirardi M, Patti E. A distributed IoT infrastructure to test and deploy realtime demand response in smart grids. IEEE Internet Things J. 2018;6(1):1136–46.
21. Azmoodeh A, Dehghantanha A, Conti M, Choo K-KR. Detecting crypto ransomware in IoT networks based on energy consumption footprint. J Ambient Intell Humanized Comput. 2018;9(4):1141–52.
22. Habib MA, Ahmad M, Jabbar S, Ahmed SH, Rodrigues JJPC. Speeding up the internet of things: Leaiot: a lightweight encryption algorithm toward low-latency communication for the internet of things. IEEE Consum Electron Mag. 2018;7(6):31–7.
23. Ucci D, Aniello L, Baldoni R. Survey of machine learning techniques for malware analysis. Comput Secur. 2018;81:123–47.
24. Raja G, Manaswini Y, Vivekanandan GD, Sampath H, Dev K, Bashir AK. AI-Powered Blockchain - A Decentralized Secure Multiparty Computation Protocol for IoV. In: IEEE INFOCOM 2020 - IEEE conference on computer communications workshops (INFOCOM WKSHPS), Toronto, ON, Canada. 2020. p. 865–870.
25. Iwendi C, Maddikunta PKR, Gadekallu TR, Lakshmanna K, Bashir AA, Piran MJ. A metaheuristic optimization approach for energy efficiency in the IoT networks. J Softw Pract Exper. 2020. https://doi.org/10.1002/spe.2797.
26. Wang C, Huang R, Shen J, Liu J, Vijayakumar P, Kumar N. A novel lightweight authentication protocol for emergency vehicle avoidance in VANETs. IEEE Internet Things J. 2021;8(18):14248–57.
27. Xiao Z, Fu X, Zhang L, Goh RSM. Traffic pattern mining and forecasting technologies in maritime traffic service networks: a comprehensive survey. IEEE Trans Intell Transp Syst. 2020;21(5):1796–825.
28. Liu RW, Nie J, Garg S, Xiong Z, Zhang Y, Hossain MS. Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems. IEEE Internet Things J. 2021;8(7):5374–85.
29. Liu D, Shen J, Vijayakumar P, Wang A, Zhou T. Efficient data integrity auditing with corrupted data recovery for edge computing in enterprise multimedia security. Multimedia Tools Appl. 2020;79(15):10851–70.
30. Zhang Z, Deng R, Yau DKY, Cheng P, Chen J. Zero-Parameter-Information FDI Attacks Against Power System State Estimation. In: 2020 American Control Conference (ACC), Denver, CO, USA. 2020. p. 2987–92.
31. Falco G, Caldera C, Shrobe H. IIoT cybersecurity risk modeling for SCADA systems. IEEE Internet Things J. 2018;5(6):4486–95.
32. Javed AR, Rehman SU, Khan MU, Alazab M, Reddy T. CANintelliIDS: Detecting in-vehicle intrusion attacks on a controller area network using CNN and attention-based GRU. IEEE Trans Netw Sci Eng. 2021;8(2):1456–66.
33. Rehman SU, Khaliq M, Imtiaz SI, Rasool A, Shafiq M, Javed AR, Jalil Z, Bashir AF. DIDDOS: An approach for detection and identification of distributed denial of service (DDoS) cyberattacks using gated recurrent units (GRU). Future Gener Comput Syst. 2021;118:453–66.
34. Imtiaz SI, Rehman SU, Javed AR, Jalil Z, Liu X, Alnumay WS. DeepAMD: Detection and identification of Android malware using high-efficient deep artificial neural network. Future Gener Comput Syst. 2021;115:844–56.
35. Li Q, Cheng M, Wang J, Sun B. LSTM based phishing detection for big email data. IEEE Trans Big Data. 2020;8(1):278–88.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.