Received October 1, 2021, accepted November 5, 2021, date of publication November 8, 2021, date of current version November 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3126715

An Enhanced Blockchain-Based IoT Digital Forensics Architecture Using Fuzzy Hash

WAEL A. MAHROUS[1], MAHMOUD FAROUK[2], AND SAAD M. DARWISH[3]

[1]Higher Institute for Tourism, Hotels and Computer, Al-Seyouf, Alexandria 00203, Egypt
[2]Management Information System Department, Agami Higher Institute of Administrative Sciences, Alexandria 21575, Egypt
[3]Department of Information Technology, Institute of Graduate Studies and Research, Alexandria University, Alexandria 21544, Egypt

Corresponding author: Saad M. Darwish (saad.darwish@alexu.edu.eg)

ABSTRACT Due to businesses' growing use of IoT services in their day-to-day operations and the increased use of smart devices, digital forensic investigations involving such systems will need increasingly sophisticated digital evidence collection and processing. The majority of IoT systems are composed of disparate software and hardware components, which may pose security and privacy concerns. Recently, blockchain technology was presented as one of the options for achieving IoT security via the use of an immutable ledger, a decentralized architecture, and strong cryptographic primitives. Integrating blockchain platforms with IoT-based applications, on the other hand, poses a number of difficulties owing to the trustworthiness, integrity, and real-time responsiveness of IoT data. However, certain IoT devices may be incompatible with existing blockchain-based IoT forensic methods for additional digital evidence processing owing to their usage of conventional hash. A critical feature of cryptographic hash functions is that even if just one bit of the input is altered, the output acts pseudo-randomly, making it impossible to identify identical files. However, in the field of computer forensics, it is essential to locate comparable files (e.g., various versions of a file); therefore, we need a hash function that preserves similarity. It is getting more difficult to establish how forensic investigators might utilize traces from such devices. To effectively deal with IoT digital forensics applications, this article presents an improved blockchain-based IoT digital forensics architecture that uses the fuzzy hash to construct the Blockchain's Merkle tree in addition to the conventional hash for authentication. Fuzzy hashing enables the identification of potentially damning documents that might otherwise remain undiscovered using conventional hashing techniques. By comparing blocks/files to all nodes in the blockchain network using fuzzy hash similarity, the digital forensics investigator will be able to verify their authenticity. To support the proof of concept, we simulated the suggested model.

INDEX TERMS Blockchain, Internet of Things, fuzzy hash, IoT forensics, digital examination.

I. INTRODUCTION

Digital forensics is getting increasingly difficult to perform as a result of the exponential growth of computing devices and computer-enabled paradigms, posing new difficulties for remote data processing. The Internet of Things (IoT) is the network of individually identified embedded computing devices that are connected to the current Internet infrastructure. With billions of new and growing devices, the IoT expands the security risks. While the IoT inherits the same monitoring needs as cloud computing, the associated difficulties are exacerbated by the volume, diversity, and velocity of data [1]. Current digital forensic tools, investigative frameworks, and procedures are incapable of addressing the IoT environment's heterogeneity and dispersion characteristics. These features provide significant difficulties for digital forensic investigators and law enforcement agencies. The complexity of the IoT system and the absence of an integrated standard complicate the collection of forensic evidence by security and law enforcement authorities.

As a forensic analyst, this presents challenges since we must devise new methods for collecting and securing this data while ensuring that no evidence has been tampered with. The aim is to identify solutions to these issues by examining how these various types of evidence may be properly seized, kept, extracted, and evaluated. At the moment, there is a defined technique for collecting evidence from hard drives and mobile phones, but no clear protocol for investigating IoT-based

The associate editor coordinating the review of this manuscript and approving it for publication was Theofanis P. Raptis.

devices [1]. While standards for dealing with electronic or digital evidence are being developed, additional supporting disciplines must adapt to help investigators in this new realm and ensure they are educated about appropriate crime scene behavior [2].

From an investigator's viewpoint, the primary difficulties presented by an IoT-based crime scene are as follows: (1) the size of objects of forensic relevance; (2) location - impacts on ease of access, potential connection to other devices, local or cloud-based, etc. (3) The significance of the devices discovered and collected (4) Legal/jurisdictional considerations (5) Ambiguous network boundaries/edgeless networks, i.e. no perimeter or a perimeter that is less clearly defined. (6) Are the tools available sufficient for the tasks? Is the data secure? Is the device a data storage device or is it just middleware? [3].

Existing methods are built for a different generation of evidence sources, with the premise that items of forensic interest would always be available and accessible — while objects of forensic interest in the IoT may not always be available or accessible [3]. Cloud forensics will also be critical in enforcing cybersecurity best practices since all data produced by IoT components will be kept in the cloud for scalability, capacity, and ease. As the number of IoT-connected devices continues to increase, it has become necessary to create a new procedure for investigating IoT-related events. To address security issues, a new era of digital forensics and best practices will be required to authenticate and utilize physical and digital evidence concurrently in a changing regulatory environment [4].

Numerous businesses and academics are increasingly interested in blockchain technology because it offers solutions to the issues connected with traditional centralized architecture. Whether public or private, a blockchain is a distributed ledger that is capable of preserving transaction integrity by decentralizing the ledger across participating users [5]. While the centralized IoT system offers many advantages, it also introduces some difficulties. Integrating IoT and blockchain technologies may help address these issues. Numerous studies utilize the blockchain as a data integrity preservation technique in the digital forensics process since it makes the ledger and internal information visible to all parties, allowing for the verification and preservation of the information's integrity. Current data integrity verification techniques in digital forensics are usually used to gather digital evidence via legal processes and image the disk using professional digital forensic tools [6]. A central authority verifies digital evidence in this manner. However, this centralized approach of preserving evidence's integrity introduces the danger of evidence being tampered with by malevolent insiders or attackers.

In general, the primary constraints of blockchain for IoT [5], [6] are as follows (1) Resource consumption: to secure the blockchain network from attack, the conventional consensus method consumes a large number of resources, which is too expensive for resource-constrained IoT devices. (2) Throughput limitation: Because a new block's capacity is restricted, transactions per second are generally limited to a few dozens, making it unable to keep up with the exponential development of IoT devices. Finally, (3) Confirmation delay: The confirmation delay is too long for IoT applications because of the low access rate of new blocks [40]. As a consequence, there is a need for a lightweight algorithm(s) with trade-offs amongst cost and performance, and security. For resource-constrained IoT devices, lightweight –based blockchain technology is an effective way to force evidence integrity [41].

Fuzzy hashing enables the identification of potentially damning documents that might otherwise remain undiscovered using conventional hashing techniques. The fuzzy hash is similar to the fuzzy logic search in that it looks for documents that are similar but not identical, referred to as homologous files. While homologous files contain similar binary data strings, they are not exact copies. Additionally, fuzzy hashing may be used to compare incomplete files, such as sliced papers, to other documents of interest. Carving may enable the recovery of fragments of documents that may be linked to the original. Fuzzy hashes may also be used to connect a document to a suspect when the document implicating the suspect does not exist in the current file system. If the investigator has access to the original or suspected document, it is possible to hash it fuzzily [7]. It may then be compared to carved objects extracted from the image to ascertain if it was ever in the system.

A. PROBLEM STATEMENT

Due to a lack of security mechanisms, evidence in IoT devices can be altered or destroyed; this can have a detrimental impact on the quality of evidence and possibly render it inadmissible in court. Vendors may not update their devices on a regular basis or at all, and they frequently discontinue support for older devices when releasing new products with new infrastructures. As a result, hackers can attack newly found weaknesses in IoT devices. An investigator must identify and gather evidence at a digital crime scene during the identification phase of forensics. One problem is identifying all IoT devices present at a crime scene, many of which are tiny, harmless, and perhaps switched down. Furthermore, due to the variety of devices and manufacturers' varied platforms, operating systems, and hardware, collecting evidence from these devices is a significant problem. Current digital forensic techniques are not intended to deal with the heterogeneity that exists in an IoT context. Massive volumes of diverse and dispersed evidence created by IoT devices found at crime scenes significantly increase the difficulty of forensic investigations [8]–[12].

Recently, the benefit of blockchain technology in digital forensics is that the examiner may self-verify digital evidence by using the hash function to efficiently create a chain of verifiable evidence. However, the conventional hash method used to ensure data integrity inside blockchain networks is inefficient at dealing with identical files that may arise from

benign or malicious alteration of the IoT sensors examined by the forensic investigator.

B. CONTRIBUTION

The article proposes a fuzzy-enabled blockchain framework for IoT forensics investigation. The proposed framework provides forensic investigation with high levels of authenticity, traceability, and distributed confidence among evidentiary entitles and examiners to deal with the heterogeneity that exists in an IoT context. Within the suggested framework, the evidence items are hashed into a Merkel tree and written into the block using a fuzzy hash. Merkle trees provide a way to prove both the integrity and validity of data and significantly reduce the amount of memory needed to do the above. Furthermore, the required proof and management only needs small amounts of information to be transmitted across networks. Utilizing fuzzy hash functions enables forensic investigators to successfully deal with permissible alteration to digital evidence while using conventional hash methods is ineffective in this situation. The suggested model is feasible for implementing in low power and low memory IoT devices through utilizing a simplified Proof of Work (PoW) consensus (lightweight–based blockchain).

The remainder of this article is structured as follows. Section II provides a thorough overview of current research on IoT forensic analysis, and Section III introduces the suggested blockchain-enabled IoT forensic chain architecture. The experimental results that show the performance of the suggested model and the assessment are given in Section IV. Section V summarizes the paper's research difficulties and trends.

II. RELATED WORK

In recent years, IoT-related research efforts have focused on IoT forensics [13]–[17], which includes the identification, collection, storage, analysis, and dissemination of digital evidence in IoT environments [16], which is very distinct from traditional computer forensics. IoT systems include a varied range of smart devices, heterogeneous networks, and different applications, where massive data volumes and disparate technologies provide new difficulties for forensic investigation [18]–[20]. Since 2017, Digital Forensics (DF) has used developing blockchain technology to record evidence objects, interaction activities, and evidence preservation [21]–[26].

Numerous forensic investigation techniques and analytical models have been suggested based on the knowledge and experiences of forensic investigators and practitioners [20], [27]. However, there are presently no internationally recognized standards that codify these established forensic investigative procedures. Specifically, current forensic investigation techniques have many difficulties in complex digital settings such as IoT, cloud computing, and the networked digital cyber-physical environment. Cebe *et al.* [21] created a blockchain architecture that is optimized for the lightweight application that integrates DF procedures and data privacy to enable effective digital examination of vehicles.

Zhang *et al.* [28] recommended a provenance process model for digital investigations using blockchain technology in a cloud environment, intending to increase stakeholder confidence in cloud forensics. Al-Nemrat [29] examined the feasibility of incorporating blockchain technology into the investigation of financial fraud in e-governance, and their findings indicate that blockchain technologies may effectively fund fraudulent online product evaluations. In untrusted software development, blockchain technology can guarantee integrity, trust, immutability, and authenticity.

The blockchain is utilized in [30] to offer auditability and traceability during software development, and a role-based access control system is created to prevent illegal data access. Hossain *et al.* [31] presented a blockchain-based forensic investigation framework for identifying criminal events in IoT environment and gathering interactions between various IoT entities. Although the suggested framework is effective at modeling interaction transactions, it is inefficient in collecting and analyzing data in large-scale IoT systems. Lone and Mir [32] developed a DF chain based on the widely used Ethereum blockchain technology. The recommended forensic chain concept was implemented on Ethereum, which may offer data gathered from various sources with integrity, transparency, and authenticity. Numerous studies have been conducted on the digital investigation in a heterogeneous environment [20], on lightweight security solutions for IoT devices [26], and digital witnesses [33].

In [10], the authors presented a fog-based IoT forensic framework for early detection and mitigation of intrusions on IoT devices. With the proliferation of IoT devices, the number of security breaches and cyber-attacks is expected to grow. Regrettably, existing forensic techniques are insufficient for collecting forensic evidence in the event of a cyber-attack involving an IoT device. The authors addressed significant difficulties connected with cloud computing and IoT forensics, as well as alternative computing paradigms such as fog computing that may aid in resolving these issues.

In [12], the authors established the blockchain-based forensic investigation framework by taking into account the variety of devices, evidence items, data formats, and more in the complex IoT context. The primary objective is to recover artifacts from IoT devices and then publish them to a blockchain-based IoT forensic chain after evaluating the relationships between evidence items, their provenance, traceability, and auditability.

In [13], the authors described a novel similarity hashing method for use in digital forensics. Their hash is based on the information retrieval concept of TF-IDF (Term Frequency - Inverse Document Frequency). The TF-IDF is a statistical metric that is used to determine the significance of a word in a collection or corpus of documents. Their hash function takes advantage of this concept to determine the most significant pieces (features) of a text. The contribution of a file fragment to the final similarity score is determined by its significance or relevance in this metric.

TABLE 1. Main frameworks and solution that suggested in IoT branch of digital forensics.

Framework	Comments and Limitation	Contribution and Comments
Generic Digital Forensics for IoT [44]	This framework includes the fundamental phases of the digital forensics process, but it lacks a strategy for feedback and assessment, and it makes no reference to privacy or integrity.	It is significant because it incorporates the different safety concepts provided by the ISO standard.
IoT Forensics Framework for Smart Environment [45]	This framework is a privacy-conscious framework that incorporates a set of privacy principles capable of enhancing data privacy but is insufficient for IoT devices with low resources.	The primary objective is to provide a new lightweight version of the IoT forensics framework that can be used to investigate crimes that happened in the IoT environment and is compatible with the nature of IoT devices.
IoT Forensics Meets Privacy Towards Cooperative Digital Investigations [46]	The author presents an adaptive Model for the study of an IoT environment that incorporates a number of privacy and security issues.	The suggested model's concept is to offer a digital witness strategy and methodology that enables citizens to submit sensitive information with investigators through the PROFIT technique.
Application Specific Digital Forensics Investigative Model in IoT [47]	The suggested model comprises of the fundamental phases of the forensic process and does not include any security concepts like as privacy, integrity, or confidentiality.	A forensically significant artifact in three widely accepted IoT application scenarios: smart house, smart city, and wearable.
Privacy Aware IoT Forensics Process Model [48]	This is a privacy-conscious framework that includes a set of privacy principles capable of improving data privacy but is inadequate for low-resource IoT devices.	The suggested framework incorporates all of the fundamental phases of any forensic process model, as well as additional stages such as review, initiation, and feedback.

In [43], the authors provided a comprehensive survey of various frameworks and solutions in all branches of digital forensics with a focus on cloud forensics. Many issues arise as a consequence of the application of digital forensics in crimes committed in a smart environment with IoT devices since the increased number of digital devices linked to the internet results in an ever-growing amount of data. Based on these issues confronting forensics in the age of new technology, a new integration between forensics processes and new technologies such as mining algorithms, security algorithms, data integrity, and authentication algorithms was suggested to address all of these issues. In other more sophisticated cases, new frameworks were offered to address issues that can be resolved through integration. Table 1 shows the main DF frameworks and solutions that were suggested for the IoT environment.

From the literature, it is apparent that the most recent DF analysis and research efforts fall into two categories: (1) those aimed at helping law enforcement, and (2) those aimed at particular forensics applications. The purpose of this study is to create a distributed ledger architecture that may be utilized in complicated cyber settings (such as IoT and cyber-physical systems). The primary difference between the proposed model and the previous blockchain-based IoT digital forensics framework is that the proposed model analyzes the Blockchain validity (evidence items) using fuzzy hashing rather than traditional hashing in order to extend the ability of related work to deal with evidence item modifications caused by benign or malicious IoT environment attacks. When the resemblance between two blocks exceeds 95%, the block is recognized as original evidence.

III. METHODOLOGY

In general, the blockchain can increase transparency at each stage, for example, by assisting the examiner in accurately identifying data sources during the early investigation stage, reducing data storage, and increasing transactional analysis efficiency, all of which can help reduce the investigation's costs. All the studies that discussed IoT digital forensics [1], [3], [5], [8] confirmed that utilizing blockchain technology provides security against attacks as the IoT forensic investigation framework is built on a private blockchain network.

The suggested approach, which incorporates fuzzy hashing into the IoT Blockchain's digital forensic architecture, primarily accomplishes the following goals. (1) Forensic examinations. (2) Continual integrity: critical evidence data was lost or corrupted as a result of insecure evidence systems. A continuous integrity check or validation method is now lacking for the whole evidence chain; the fuzzy hash will be able to resolve this issue. (3) Due to the nature of blockchain technology, it is capable of providing DF immutability and audibility, which are critical characteristics of a DF chain of evidence. (4) The blockchain may be used to track flaws and offer convenient traceability from the scene to the court throughout the evidence chain, thus restricting access to all recorded data (i.e., evidence items, examiners, timestamps,

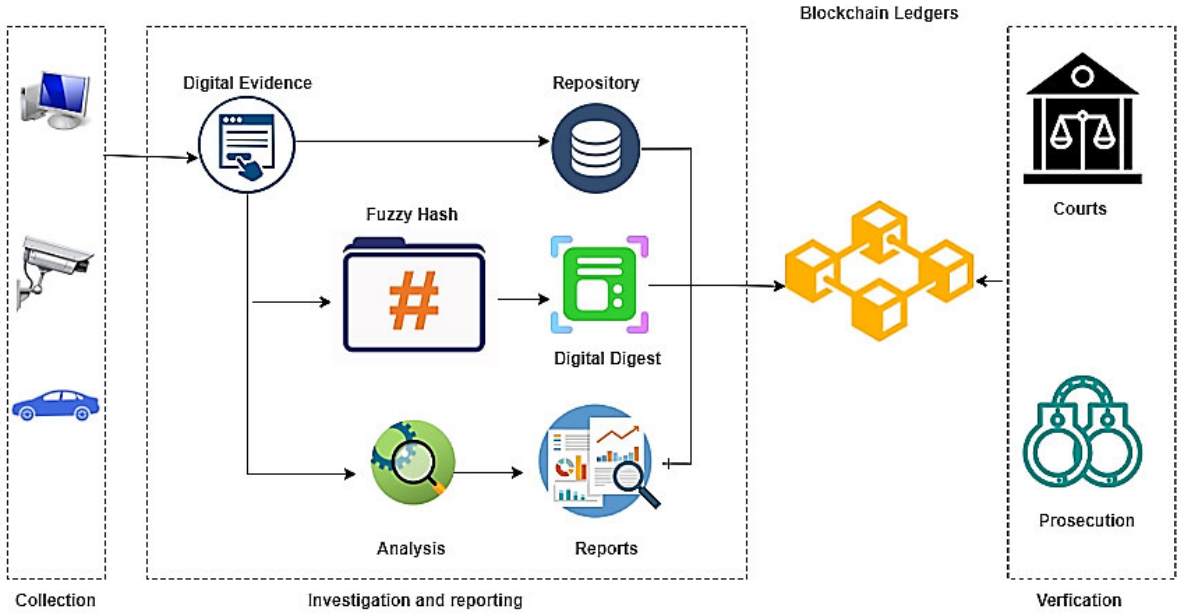

FIGURE 1. Fuzzy hash-based blockchain of IoT evidence management.


```
{
    "device": "Gate7",
    "sensor": "MotionON",
    "time": "2021-02-04 14:57:29.470945",
    "nonce": 43654,
    "Merkleroot": [
        "3:Y8+7xegTJXiQWCAR0V5DX9Py/OPO0f8FykIx:Y8+4gdXpQsdX9aOPC
    ],
    "prev_block": {
        "hash": "0000e61ee9657e290075b6767220c189f2af946756ea5d5d
        "filename": "6"
    }
}
```


FIGURE 2. JSON script for evidence blocks.

and tools) on the blockchain [12]. Fig. 1 depicts the suggested model's major schematic.

A. EVIDENCE IDENTIFICATION AND ACQUISITION

In an IoT context, the vast majority of data is recorded digitally at the point of collection, with proof in the form of digital assets gathered from sensors, devices, cloud storage, and other sources. Restriction of access to a digital asset is problematic in the context of criminal evidence. This stage consists of three major steps:

- The suggested approach identifies and fingerprints digital evidence using a one-way hash algorithm (SHA256). If several versions of digital assets are discovered, each claiming to be definitive, a digital fingerprint is created for each piece of digital evidence; the contents and inspection events are specified as TE records. Fig. 2 illustrates a JavaScript Object Notation JSON) script for a piece of evidence.
- Along with additional metadata and timestamps, the fingerprinted records will be uploaded to the blockchain, as will any identification events/findings throughout this step.
- Each member in the peer-to-peer blockchain network will have a full copy of the evidence blockchain. Once an evidence block is added to the blockchain, each participant may be certain that the data will be accessible and traceable. Each piece of evidence will have an extremely high degree of provenance. For instance, if an evidence item consists of several parts from various sources, each component and its source will be fingerprinted using a hash function to create a TE item in the blockchain [34]. Similarly, the blockchain will be used to create the whole of the complete evidence chains. When TEs need to be ''transferred'' from one party to another, new records will be generated and added to the blockchain using digital signatures.

B. FORENSIC-CHAIN FRAMEWORK

The proposed model is a blockchain-based forensic chain of custody solution for digital investigation. It enables the system to establish a distributed ledger for recording and storing TEs (examining events/findings, and other information). These TEs will be distributed through the blockchain network to all authorized participants. The framework is comprised of the essential components listed below.

- Users and IoT Devices: The term ''users'' refers to those who are involved in this investigation as users, owners, or examiners 35]. All devices, sensors, and IoT infrastructures involved in the case are included in this framework.
- Merkle Tree: A Merkle tree is a hash tree that enables the investigation's TEs to be verified efficiently and securely [12]. It can aggregate all TEs, examine other

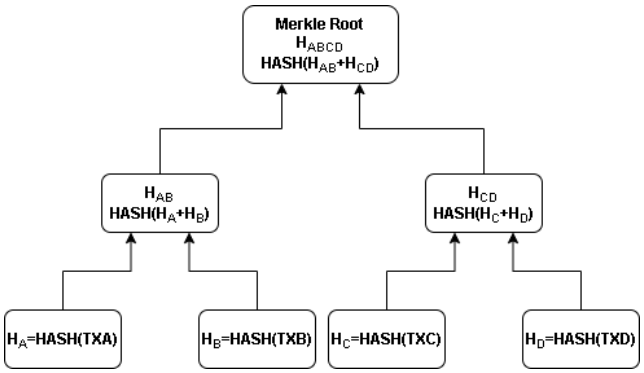

FIGURE 3. Merkle tree structure.

information in a block, and generate a digital signature for the whole collection of objects, allowing a user to verify whether or not a transaction is included in a block. Fig. 3 illustrates a Merkle tree of nodes, where the TE is a file in this instance (it could be folder or memory). A hash tree may be constructed by continually hashing transactional evidence or its hash value until it aggregates into a single root hash; in this work, HT indicates the evidence's hash value.

$$HT_1 = Hash(\text{Transactional\_Evidence\#1}) \quad (1)$$

$$HT_2 = Hash(Transacrional\_Evidence\#2) \quad (2)$$

$$HT_{12} = Hash(HT_1|HT_2) \quad (3)$$

$$H_{\text{root}} = Hash(H_{12}|\ldots) \quad (4)$$

This phase reads all block information and applies the Merkle tree method to get the Merkle root using the fuzzy hash to validate the transactions across all blocks. Finally, this node creates the new block as a new file using the Merkle tree algorithm. Algorithm 1 outlines the stages involved in the construction of a Merkle tree.

Algorithm 1 Merkle_Tree

Input Mined Block Header H : Block payload P
Output Similarity, Boolean valid
H = Extract_nonce_value ( )
B = Calculate_Merkle_Tree ( )
V = Create header_verify(H, P, B)
R = verify(H, B, V)
Similarity = calculate_ssdeep_hashed_value(R)
if (similarity >= 90) **then**
Valid ← TRUE;
else
Valid ← FALSE;
end if
return Valid;
End

- Block: In the proposed model's blockchain network, the evidence item's signature may be validated. Each block's header includes the following attributes: the pre-block hash, the version, the nonce, the timestamp, the block state, and the Merkle root (see Fig.4). The TE item is used to represent the record of the evidence item and is hashed into a Merkle tree.
- Smart Contract: A smart contract, also known as a blockchain contract, is a computer-executable digital contract. Typically, the smart contract is kept on the blockchain network and is overseen by the nodes of the blockchain network. It enables users to exchange information, data, and business processes automatically and without the need for a middleman. Smart contracts may execute, verify, and make decisions automatically in a secure and immutable manner on the decentralized ledger [12]. The following features of the smart contract may help the DF investigation. (1) Autonomy—it may specify the criteria for autonomously locating linked evidence items. (2) Trust evidence items may be encrypted and stored on a distributed ledger. (3) Security—items may be encrypted cryptographically. (4) Speed—when compared to manual processing, smart contracts may substantially decrease examination time. (5) Cost savings—smart contracts eliminate the need for intermediaries such as notaries and witnesses. (6) Accuracy—the automated smart contract operates in a more efficient, accurate, and cost-effective manner.

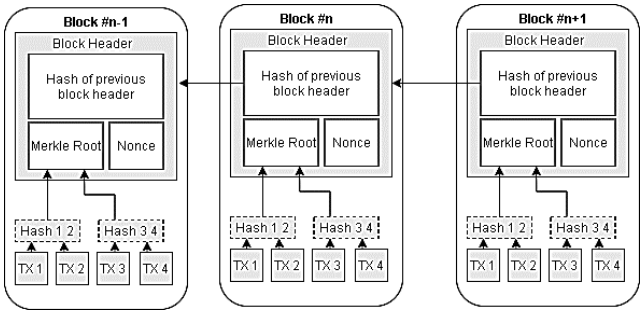

FIGURE 4. Chained blocks.

In our approach, a smart contract begins when the node receives transaction evidence; the node then calculates the nonce using Proof of Work (PoW) consensus and broadcasts it to the blockchain network; it then constructs a Merkle tree depending on the validity of prior blocks. A fuzzy hash of all previous blocks is utilized inside the Merkle tree, and if it is valid, the node will add the new block to the current blockchain in the form of a file.

In our case, reducing the complexity of PoW is needed in IoT. Using a simplified PoW will decrease the time to achieve the consensus between nodes in an IoT network and this is the most used business scenario. For Bitcoin and Ethereum applications, PoW is attributed with high complexity as it requires high computer resources, such as memory and processor as each node requires to generate a hash value that starts with '0000', which will take a long time until find the required hash. In an IoT network, we should reduce the PoW complexity due to IoT resource-constrained. In the suggested model, to achieve the consensus between nodes in an IoT network, we used a simplified PoW algorithm so that the

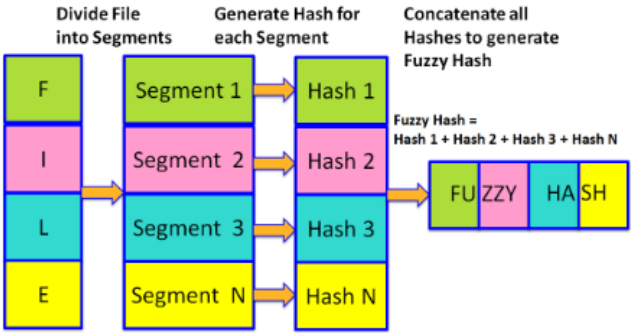

FIGURE 5. Generation of fuzzy hash value.

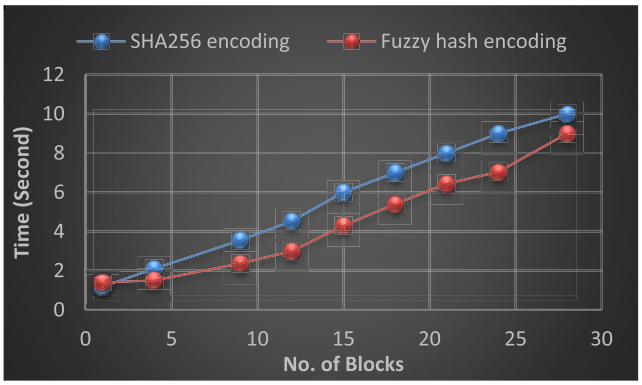

FIGURE 6. Response time for the mined block using SHA256 vs Fuzzy Hash (in case of a small number of blocks).

generated hash started with '00' instead of '0000', which will reduce the time required to create the new block and consequently reduce the complexity.

- Fuzzy Hashing: To ensure that the blockchain is not tampered with at any node, the digital forensic investigator will utilize SSDEEP [36] as a fuzzy hash method to compare file similarities. In security analysis, fuzzy hash methods are used to try to identify file tampering while examining the integrity and similarity of files of interest [36]. The file of interest is split into several blocks and a hash value is computed for each block, with the last step being the concatenation of all block hash values to create the fuzzy hash value as illustrated in Fig. 5. Numerous variables influence the length of the fuzzy hash value, including the block size, the file size, and the output size of the hash algorithm chosen.

Herein, SSDEEP is utilized to build fuzzy hash [36], [37]. SSDEEP is a program that computes context-sensitive piecewise hashes (CTPH). CTPH, also known as fuzzy hashes, is capable of matching inputs that share homologies. These inputs include sequences of identical bytes in the same order, but the bytes between these sequences may vary in content and length. This technique splits a file into a number of chunks according to its content. A rotating hash technique is used to identify the endpoints of these blocks. A rolling hash algorithm generates a pseudo-random value from the input's current context. The rolling hash algorithm operates by preserving a state 6 entirely on the basis of the last few bytes of the input. Each byte processed is added to the state and deleted after a certain number of additional bytes have been processed [38].

C. ANALYSIS

The smart contract will be utilized to generate the analysis results at this step. The digital investigator will analyze the block information to ensure that the similarity rate is more than 90%, ensuring that the blockchain is not tampered with. Utilizing fuzzy hash functions enables forensic investigators to successfully deal with permissible alteration to digital evidence while using conventional hash methods is ineffective in this situation. The digital forensic investigator will pick any node with a random block in the concurrent blockchain, create the fuzzy hash signature and then compare the signatures to the blocks of other nodes using SSDEEP. The result is 100, indicating that the whole block conforms to the block signature and the block data is unaltered.

D. PRESENTATION

As stated before, this step will be based on the analysis stage's results; all evidence can be readily traced back to its source. All reports and presentations will be built on top of the blockchain and will be added to it.

IV. EVALUATION AND DISCUSSION

In this section, we assess the suggested model by conducting several experiments in terms of throughput, response time, and the delay-incurred performance metrics. Our model is implemented using python. For developing Python programs, Atom IDE is used. All the experiments are simulated on Intel Core i5 CPU 2.4 GHz, 4 GB memory, Windows OS. Herein, the miner is deployed for validating the blockchain and the Proof-of-Work concept is used. The proposed model started with building a Merkle tree, validating the blockchain, creating root hash using fuzzy hash, implementing the proof-of-work then creating the text file containing the block information. In our case, the fuzzy hash is used to encode the Merkle tree, and this step comes after applying SHA256 to encode the TE records.

The first set of experiments was conducted to compare fuzzy hash and traditional hash for creating root hash of created Merkle tree in terms of response time. The response time is the time taken by the node to receive the transaction, mining the new block, and create a text file with the mined block information. Fig. 6 shows the response time as a function of the number of minded blocks. The results reveal that the response time is gradually increased with the increase in the number of mined blocks.

The results reveal that the suggested model that utilizes fuzzy hash reduces the response time by an average of 2% compared with the same model that utilizes traditional hash to encode TE records. This confirms that the suggested model can be implemented in real-time digital investigation applications. One possible explanation of these results is that fuzzy hash operates based on the MD5 algorithm and the

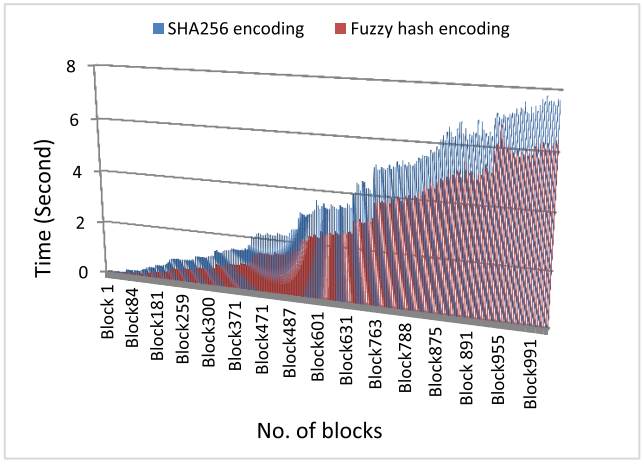

FIGURE 7. Response time for the mined block using SHA256 vs Fuzzy Hash (in case of a large number of blocks).

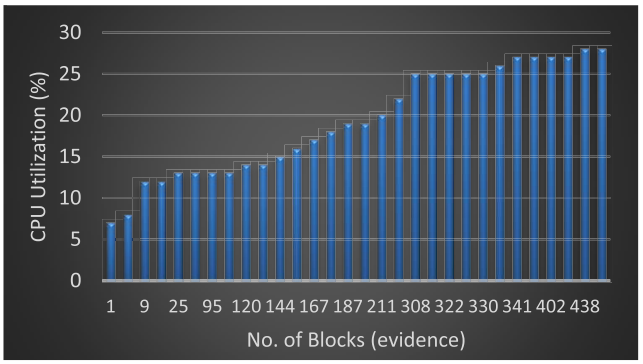

FIGURE 8. CPU utilization with respect to the number of blocks.

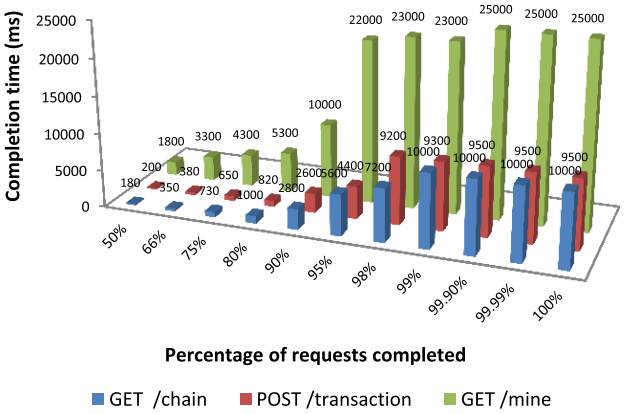

FIGURE 9. Completion time with respect to the percentage of the various completed requests (50 users).

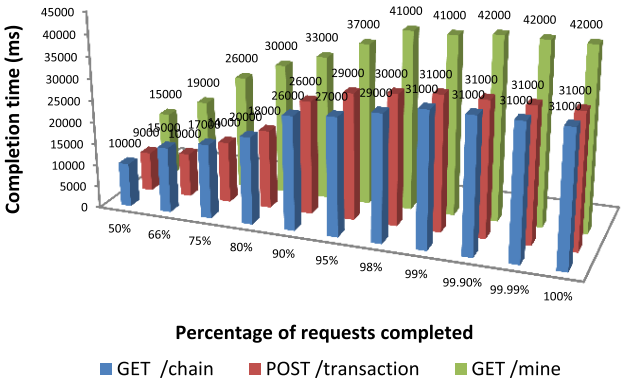

FIGURE 10. Completion time with respect to the percentage of the various completed requests (100 users).

complexity of the MD5 algorithm and SHA256 is equal but the running time of MD5 is faster than SHA256. MD5 produces 32 chars hash while SHA256 produces 64 chars hash. The results shown in Fig. 7 confirm the same fact that was concluded previously but in the case of a large number of mined blocks. The results confirm the feasibility of the suggested model to deal with a large number of blocks in terms of response time for mined blocks.

The second set of experiments was implemented to assess the overall average CPU usage against the number of generated blocks. As expected, from Fig. 8, more CPU resources are needed to mine more blocks. However, as the number of blocks to be mined increases, the amount of CPU usages does not increase with a large amount. According to the computer resources and the number of transactions, services, the CPU utilization varies during producing the new blocks.

In order to assess our model performance a modern load testing framework called Locust [39] was used to test the system infrastructure along with the various APIs. It allows simulating users' behavior using Python scripts. Three scenarios were designed to stress all the Model APIs. They involve a variable number of concurrent users (50, 100, and 150) with a fixed hatch rate of 5 users/sec. Each scenario is run for a duration of 2 minutes during which users perform multiple operations, including GET chain to get the available used blockchain status, POST transaction to the available node, and GET the mined block status. Fig. 9 through 11 depict the percentage of requests completed in a given time interval for the three scenarios. We observe that the completion time increases with the increase in the number of concurrent users, it is also observed that GET/mine requests incur longer service time, this is because the mining process uses computer resources to do the POW consensus and apply Merkle tree root hash.

The final set of experiments was conducted to confirm the ability of the suggested model for execution on IoT devices with low configurations such as The Raspberry Pi that is a low-cost, credit-card- sized computer. Raspberry Pi is widely used in many areas, such as for weather monitoring, because of its low cost, modularity, and open design. It is typically used by computer and electronic hobbyists, due to its adoption of HDMI and USB devices [42]. In our case, the employed Raspberry Pi has the following configuration: Processor type: 32-Bit, Max CPU Speed: 700 MHz, Operating System: Raspbian 5.6.12, Model: 3B, Memory: 512 MB. In this case, was considered Raspberry Pi as a node on our blockchain network. Two important factors for IoT devices are considered during the experimentation: time consumption

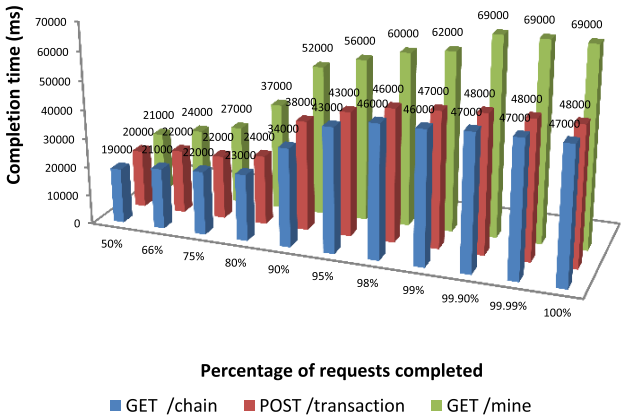

FIGURE 11. Completion time with respect to the percentage of the various completed requests (150 users).

TABLE 2. Time needed to generate the new block and added it to the blockchain network (seconds).

	Laptop High configuration	Raspberry Pi Low configuration
Average Time	3	13

TABLE 3. Power consumption needed to generate the new block and added it to the blockchain network (mW).

	Laptop High configuration	Raspberry Pi Low configuration
Average Power	4	12

and power consumption. We used a wall outlet power meter to measure the power consumption.

The results in Table 2 and Table 3 show the time and power consumption that are required to run the proposed model on two devices with different configurations (one for IoT devices with low configuration and the other for IoT devices with high configuration) respectively. The larger value of time for the Raspberry Pi device is due to the complexity of operations needed to be performed. Moreover, the Raspberry Pi needs fewer resources as compared to a laptop. According to the above comparison in terms of time and power consumption, the suggested model is feasible for implementing in low power and low memory IoT devices.

V. CONCLUSION AND FUTURE WORK

Preservation of data integrity is carried out independently by central authorities in the present digital forensics investigation. This method is sufficiently efficient and convenient procedurally, but the integrity of prospective evidence may be jeopardized if the central authority is attacked by a malevolent attacker. Additionally, human and material resources are expended to maintain the chain of custody and ensure the investigation's integrity. Unlike today, the existing chain of custody method must include a more robust approach to integrity preservation and streamlined processes in order to conduct a thorough digital forensic investigation in large-scale IoT settings.

This article performed a preliminary forensic study on the blockchain-based forensic investigation framework, taking into account the variety of devices, evidence items, and data formats found in the complex IoT environment. We propose a blockchain-based digital forensic framework for the IoT environment in this article to address the heterogeneity and dispersion of the IoT environment, as well as the centralization of current forensic investigations. Additionally, we show the updated block structure and workflow of the suggested framework for investigation by encoding Merkle trees with fuzzy hash to cope with evidence similarities (different version document). In the future work, we will investigate ways to enhance the suggested digital forensic investigation model's execution time and its complexity and apply it to real digital investigations.

REFERENCES

[1] A. MacDermott, T. Baker, and Q. Shi, "IoT forensics: Challenges for the IoA era," in *Proc. 9th IFIP Int. Conf. New Technol., Mobility Secur. (NTMS)*, Paris, France, Feb. 2018, pp. 1–5.
[2] Z. A. Baig, P. Szewczyk, C. Valli, P. Rabadia, P. Hannay, M. Chernyshev, M. Johnstone, P. Kerai, A. Ibrahim, N. Syed, N. Peacock, and K. Sansurooah, "Future challenges for smart cities: Cyber-security and digital forensics," *Digit. Invest.*, vol. 22, pp. 3–13, Sep. 2017.
[3] U. Salama, "Smart forensics for the Internet of Things (IoT)," in *Technical Report, Security Intelligence*. Armonk, NY, USA: IBM, 2017.
[4] B. Cao, Z. Zhang, D. Feng, S. Zhang, L. Zhang, M. Peng, and Y. Li, "Performance analysis and comparison of PoW, PoS and DAG based blockchains," *Digit. Commun. Netw.*, vol. 6, no. 4, pp. 480–485, Nov. 2020.
[5] H. Atlam and G. Wills, *Technical Aspects of Blockchain and IoT Advances in Computers*, vol. 115. Amsterdam, The Netherlands: Elsevier 2019, pp. 1–39.
[6] J. H. Ryu, P. K. Sharma, J. H. Jo, and J. H. Park, "A blockchain-based decentralized efficient investigation framework for IoT digital forensics," *J. Supercomput.*, vol. 75, no. 8, pp. 4372–4387, Aug. 2019.
[7] D. Hurlbut, "Fuzzy hashing for digital forensic investigators," Access Data, Pennsylvania State Univ., State College, PA, USA, Tech. Rep. 1, 2009, pp. 1–13. [Online]. Available: https://citeseerx.ist.psu.edu, doi: 10.1.1.173.6932.
[8] M. Samaniego, U. Jamsrandorj, and R. Deters, "Blockchain as a service for IoT," in *Proc. IEEE Int. Conf. Internet Things*, Dec. 2016, pp. 433–436.
[9] M. C. Xenya and K. Quist-Aphetsi, "A cryptographic technique for authentication and validation of forensic account audit using SHA256," in *Proc. Int. Conf. Cyber Secur. Internet Things (ICSIoT)*, May 2019, pp. 11–14.
[10] E. Al-Masri, Y. Bai, and J. Li, "A fog-based digital forensics investigation framework for IoT systems," in *Proc. IEEE Int. Conf. Smart Cloud (SmartCloud)*, Sep. 2018, pp. 196–201.
[11] P. K. Sharma and J. H. Park, "Blockchain based hybrid network architecture for the smart city," *Future Gener. Comput. Syst.*, vol. 86, pp. 650–655, Sep. 2018.
[12] S. Li, T. Qin, and G. Min, "Blockchain-based digital forensics investigation framework in the Internet of Things and social systems," *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 6, pp. 1433–1441, Dec. 2019.
[13] D. Chang, M. Ghosh, S. K. Sanadhya, M. Singh, and D. R. White, "FbHash: A new similarity hashing scheme for digital forensics," *Digit. Invest.*, vol. 29, pp. S113–S123, Jul. 2019.
[14] T. M. Fernández-Caramés and P. Fraga-Lamas, "A review on the use of blockchain for the Internet of Things," *IEEE Access*, vol. 6, pp. 32979–33001, 2018.
[15] S. Li, S. Zhao, P. Yang, P. Andriotis, L. Xu, and Q. Sun, "Distributed consensus algorithm for events detection in cyber-physical systems," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2299–2308, Apr. 2019.
[16] M. Hossain, Y. Karim, and R. Hasan, "FIF-IoT: A forensic investigation framework for IoT using a public digital ledger," in *Proc. IEEE Int. Congr. Internet Things (ICIOT)*, Jul. 2018, pp. 33–40.
[17] M. M. Hossain, R. Hasan, and S. Zawoad, "Trust-IoV: A trustworthy forensic investigation framework for the Internet of Vehicles (IoV)," in *Proc. Int. Congr. Internet Things*, 2017, pp. 25–32.

[18] M. Chernyshev, S. Zeadally, Z. Baig, and A. Woodward, "Internet of Things forensics: The need, process models, and open issues," *IT Prof.*, vol. 20, no. 3, pp. 40–49, May/Jun. 2018.
[19] D. Quick and K. R. Choo, "Iot device forensics and data reduction," *IEEE Access*, vol. 6, pp. 47566–47574, 2018.
[20] L. Caviglione, S. Wendzel, and W. Mazurczyk, "The future of digital forensics: Challenges and the road ahead," *IEEE Security Privacy*, vol. 15, no. 6, pp. 12–17, Nov./Dec. 2017.
[21] M. Cebe, E. Erdin, K. Akkaya, H. Aksu, and S. Uluagac, "Block4 forensic: An integrated lightweight blockchain framework for forensics applications of connected vehicles," *IEEE Commun. Mag.*, vol. 56, no. 10, pp. 50–57, Oct. 2018.
[22] S. Wu, Y. Chen, Q. Wang, M. Li, C. Wang, and X. Luo, "CReam: A smart contract enabled collusion-resistant e-auction," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 7, pp. 1687–1701, Jul. 2019.
[23] L. V. Der Horst, K.-K. R. Choo, and N.-A. Le-Khac, "Process memory investigation of the bitcoin clients electrum and bitcoin core," *IEEE Access*, vol. 5, pp. 22385–22398, 2017.
[24] H. Ritzdorf, C. Soriente, G. O. Karame, S. Marinovic, D. Gruber, and S. Capkun, "Toward shared ownership in the cloud," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 12, pp. 3019–3034, Dec. 2018.
[25] G. Tziakouris, "Cryptocurrencies—A forensic challenge or opportunity for law enforcement an interpol perspective," *IEEE Secur. Privacy*, vol. 16, no. 4, pp. 92–94, Jul. 2018.
[26] Z. Liu and H. Seo, "IoT-nums: Evaluating nums elliptic curve cryptography for IoT platforms," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 720–729, Mar. 2019.
[27] A. Valjarevic and H. Venter, "A harmonized process model for digital forensic investigation readiness," in *Advances Digital Forensics*. Berlin, Germany: Springer, 2013, pp. 67–82.
[28] Y. Zhang, S. Wu, B. Jin, and J. Du, "A blockchain-based process provenance for cloud forensics," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Dec. 2017, pp. 2470–2473.
[29] A. Al-Nemrat, "Identity theft on E-government/E-governance & digital forensics," in *Proc. Int. Symp. Program. Syst. (ISPS)*, Apr. 2018, pp. 1–5.
[30] D. Ulybyshev, M. Villarreal-Vasquez, B. Bhargava, G. Mani, S. Seaberg, P. Conoval, R. Pike, and J. Kobes, "(WIP) blockhub: Blockchain-based software development system for untrusted environments," in *Proc. IEEE 11st Int. Conf. Cloud Comput.*, Aug. 2018, pp. 582–585.
[31] M. Hossain, R. Hasan, and S. Zawoad, "Probe-IoT: A public digital ledger based forensic investigation framework for IoT," in *Proc. IEEE Conf. Comput. Commun. Workshops*, Apr. 2018, pp. 1–2.
[32] A. H. Lone and R. N. Mir, "Forensic-chain: Ethereum blockchain based digital forensics chain of custody," *Sci. Practical Cyber Secur. J.*, vol. 1, no. 2, pp. 21–27, 2018.
[33] A. Nieto, R. Roman, and J. Lopez, "Digital witness: Safeguarding digital evidence by using secure architectures in personal devices," *IEEE Network*, vol. 30, no. 6, pp. 34–41, Nov./Dec. 2016.
[34] S. Kalber, A. Dewald, and F. C. Freiling, "Forensic application-fingerprinting based on file system metadata," in *Proc. 7th Int. Conf. IT Secur. Incident Manage. IT Forensics*, Mar. 2013, pp. 98–112.
[35] H. Atlam, A. Alenezi, M. Alassafi, A. Alshdadi, and G. Wills, "Security, cybercrime and digital forensics for IoT," in *Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm*. Cham, Switzerland: Springer, 2020, pp. 551–577.
[36] N. Naik, P. Jenkins, N. Savage, L. Yang, T. Boongoen, and N. Iam-On, "Fuzzy-import hashing: A malware analysis approach," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2020, pp. 1–8.
[37] K. Savage, P. Coogan, and H. Lau, "The evolution of ransomware-symantec," Symantec Secur. Response. Symantec Corp., Mountain View, CA, USA, Tech. Rep., 2015, pp. 1–57. [Online]. Available: https://its.fsu.edu/sites/g/files/imported/storage/images/information-security-and-privacy-office/the-evolution-of-ransomware.pdf-Ver1.0
[38] J. Kornblum, "Identifying almost identical files using context triggered piecewise hashing," *Digit. Invest.*, vol. 3, pp. 91–97, Sep. 2006.
[39] G. Cornetta, A. Touhafi, M. A. Togou, and G.-M. Muntean, "Fabrication-as-a-service: A web-based solution for STEM education using Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1519–1530, Feb. 2020.
[40] B. Cao, Y. Li, L. Zhang, L. Zhang, S. Mumtaz, Z. Zhou, and M. Peng, "When Internet of Things meets blockchain: Challenges in distributed consensus," *IEEE Netw.*, vol. 33, no. 6, pp. 133–139, Nov./Dec. 2019.
[41] V. A. Thakor, M. A. Razzaque, and M. R. Khandaker, "Lightweight cryptography algorithms for resource-constrained IoT devices: A review, comparison and research opportunities," *IEEE Access*, vol. 9, pp. 28177–28193, 2021.
[42] U. Khalid, M. Asim, T. Baker, P. C. K. Hung, M. A. Tariq, and L. Rafferty, "A decentralized lightweight blockchain-based authentication mechanism for IoT systems," *Cluster Comput.*, vol. 23, no. 3, pp. 2067–2087, Sep. 2020.
[43] M. Khanafseh, M. Qatawneh, and W. Almobaideen, "A survey of various frameworks and solutions in all branches of digital forensics with a focus on cloud forensics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 8, pp. 610–629, 2019.
[44] V. R. Kebande and I. Ray, "A generic digital forensic investigation framework for Internet of Things (IoT)," in *Proc. IEEE 4th Int. Conf. Future Internet Things Cloud (FiCloud)*, Aug. 2016, pp. 356–362.
[45] L. Babun, A. K. Sikder, A. Acar, and A. S. Uluagac, "IoTDots: A digital forensics framework for smart environments," 2018, *arXiv:1809.00745*.
[46] A. Nieto, R. Rios, and J. Lopez, "IoT-forensics meets privacy: Towards cooperative digital investigations," *Sensors*, vol. 18, no. 2, 492, pp. 1–17, 2018.
[47] T. Zia, P. Liu, and W. Han, "Application-specific digital forensics investigative model in Internet of Things (IoT)," in *Proc. 12nd Int. Conf. Availability, Rel. Secur.*, Aug. 2017, pp. 1–7.
[48] A. Nieto, R. Rios, and J. Lopez, "A methodology for privacy-aware IoT forensics," in *Proc. Trustcom/BigDataSE/ICESS*, Aug. 2017, pp. 626–633.

WAEL A. MAHROUS received the B.Sc. degree in computer science from the Higher Institute of Computer, Abbasiya, Egypt, in 1995. He is a Microsoft Certified Trainer, from 2008 to 2013. He is currently working with the Higher Institute of Tourism, Hotels and Computer, Al-Seyouf, Alexandria, Egypt, as a Lab Teacher, where he is teaching networks, java, digital marketing, and web applications. His research and professional interests include image processing, optimization techniques, security technologies, database management, machine learning, digital forensics, web technologies, and robotics.

MAHMOUD FAROUK received the B.Sc. and M.Sc. degrees from the Military Technical College (M.T.C), Cairo, Egypt, in 1995 and 2001, respectively, and the Ph.D. degree from the Faculty of Engineering, Alexandria University, in 2006. He was an Academic Staff with the Computer Department, M.T.C, from 1997 to 2001. He served as the Computer Specialist and the Computer Educational Assistant for Egypt Armed Forces, from 1995 to 2013. He was a Teacher with King Marriott Academy Alexandria "High Institute for Computer," from 2014 to 2020. He is currently the Head of the Management Information System Department, Agami Higher Institute of Administrative Sciences. He is the author or coauthor for more than 21 national and international papers and also collaborated in several research projects.

SAAD M. DARWISH received the B.Sc. degree in statistics and computer science from the Faculty of Science, Alexandria University, Egypt, in 1995, the M.Sc. degree in information technology from the Department of Information Technology, Institute of Graduate Studies and Research (IGSR), University of Alexandria, in 2002, and the Ph.D. degree from Alexandria University, with a focus on image mining and image description technologies. Since June 2017, he has been a Professor with the Department of Information Technology, IGSR. He is the author or coauthor of more than 50 papers publications in prestigious journals and top international conferences and also received several citations. He has served as a reviewer for several international journals and conferences. He has supervised around 60 M.Sc. and Ph.D. students. His research and professional interests include image processing, optimization techniques, security technologies, database management, machine learning, biometrics, digital forensics, and bioinformatics.