Enhancements in the world of digital forensics

Krishna Sanjay Vaddi¹, Dhwaniket Kamble², Raj Vaingankar³, Tushar Khatri¹, Pranil Bhalerao¹
¹Department of Computer Science and Business Systems, Bharati Vidyapeeth Deemed to be University Department of Engineering and Technology, Navi Mumbai, India
²Department of Computer Science and Engineering, Bharati Vidyapeeth Deemed to be University Department of Engineering and Technology, Navi Mumbai, India

ABSTRACT

Currently, the rapid advancement of computer systems and mobile phones has resulted in their utilization in unlawful acts. Ensuring adequate and effective security measures poses a difficult task due to the intricate nature of these devices, thereby exacerbating the challenges associated with investigating crimes involving them. Digital forensics, which involves investigating cyber crimes, plays a crucial role in this realm. Extensive research has been conducted in this field to aid forensic investigations in addressing contemporary obstacles. This paper aims to explore the progress made in the applications of digital forensics and security, encompassing various aspects, and provide insights into the evolution of digital forensics over the past five years.

Keywords:
Artificial intelligence
Blockchain
Cloud
Deep learning
Machine learning

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

Digital forensics refers to the systematic process of preserving, identifying, extracting, and documenting computer evidence that can be utilized as legal evidence. It involves the application of technological expertise to uncover evidence from various digital media, such as computers, mobile phones, servers, and networks. By providing the forensic team with advanced techniques and tools, digital forensics enables them to effectively solve intricate cases involving digital aspects [1]. In this paper we will analyse the evolution of digital forensics from past few years in information technology industry. How crime rates have been decreased in the industry after implementation of digital forensics techniques. The paper will focus on the study cyber crime in the fields of blockchain, cloud artificial intelligence, machine learning, and deep learning. The four basic steps in digital forensics are: i) Identifying, this is the exercise of sorting and collecting the suspected unique supply or asset believed to comprise proof of crime; ii) Preserving, this is the exercise of making sure the integrity of the accumulated proof and maintaining a "digital trail" of the statistics or media; iii) Analyzing, this is the investigative part of the system in which a forensics practitioner starts off evolved searching for obtained assets or media statistics to discover proof of the suspected crime; and iv) Reporting, this is the system of making a record of findings from the research for presentation to stakeholders and, in a few cases, a lawyer or jury in court. Reporting should additionally be tailor-made to the audience [2].

The cyber crime has affected the world to a tremendous level. The crime rate globally has shown an upward trend in last decade. The average time to find a crime was 212 days back in 2019 [3]. However, those had been estimates primarily based totally most effective on mentioned numbers. Given the prevalence of crime
worldwide, it is imperative to acknowledge the advancements made in the field of digital forensics to combat illicit activities in the realm of information technology [4].

2. METHOD

A passive data gathering process is used in this study, dependent on various literature. To get a sense of current interests in the field of digital forensics, we concentrate on statistical analysis based on patterns that are no older than 2018. Digital forensics is the core topic of a large range of reputable journals that serve both academic and commercial needs.

2.1. Artificial intelligence

This paper proposes a hybrid model of digital forensics platform based on artificial intelligence (AI) for effective crime detection. By integrating AI capabilities, this system surpasses the performance of traditional digital forensics platforms. The process includes data collection, analysis, and reporting, forming a comprehensive approach. The framework presented in this research offers guidance to developers in constructing highly intelligent and automated tools to assist in the overall process of digital forensic investigation. Leveraging artificial intelligence and machine learning, the framework provides enhanced capabilities. However, it is essential to acknowledge the impossibility of achieving a framework that is entirely error-free and intelligent, necessitating the consideration of potential mistakes and false positives. Additionally, alongside conventional investigation methods, data preprocessing, and feature extraction techniques can also be employed [5].

2.1.1. User interface

The system's user interface serves as the means through which users interact with the system. It encompasses various features, including options for selecting the type of crime, generating entry-type reports, and displaying crime matching outputs. Within the crime type option, users can choose the specific category of crime for the reported case, such as assault, burglary, and vehicle theft. Based on the selected crime type, a corresponding form with necessary fields will be generated, ensuring that the relevant database table is populated upon form submission. The crime data entry option provides users with a range of methods to input data into the model, increasing flexibility and efficiency. These methods include form completion, data uploading, document scanning, file importing, digital handwriting, and audio recording. The crime matching output component oversees the display of results from the crime matching task, enabling the detection of new cases and queries. This empowers investigation officers to obtain the required information for their investigations.

2.1.2. Feature extraction module

The crime feature extraction engine employs machine learning methods, including text analysis, optical character recognition, voice recognition, and domain-specific feature extraction algorithms, to extract organized data from unstructured crime report narratives. This approach involves utilizing a look-up table that queries a database as a point of reference for the feature extraction process. The input from crime domain experts can be incorporated into the engine through database tables, Excel files, or similar formats. Additionally, past crime records from external agencies and organizations can be integrated into the system to populate database tables and establish functional reference concepts for the entity extraction technique.

2.1.3. Database

Within the model architecture, the crime database plays multiple roles. It serves as the repository for all the data within the framework, working in conjunction with the feature extraction process to populate and enrich it. Additionally, the crime database provides recommendations for extracting records from future entries, ensuring the ongoing effectiveness of the system. Furthermore, it serves as a valuable data source for the crime data clustering method in a subsequent stage of analysis.

2.1.4. Crime clustering mechanism

The utilization of the crime clustering mechanism ensures the meaningful grouping of current and historical forensic data. The primary objective is to automatically organize the collected forensic data records into a comprehensive inventory of relevant categories. Consequently, in the subsequent phase, when a new record is entered, it can be classified and assigned to one of these clusters. This enables the investigating officer to navigate within the cluster, acquiring further insights into potential perpetrators, victims, future crime occurrences, and other pertinent details. Our analysis incorporates the k-means clustering algorithm to accomplish this task.
2.1.5. Crime matching engine

Upon the logging or querying of a new crime incident in the system, the crime matching engine undertakes the responsibility of classification. It leverages machine learning techniques such as the multi-linear perceptron, K nearest neighbor, and neural networks to predict similar crime instances. These predictions are then presented to the investigating officer as pre-investigation outputs, aiding them in guiding their subsequent investigative operations [6].

2.2. Cloud

2.2.1. Digital forensics

Numerous definitions of digital forensics have been put forth by legal and criminal researchers. As a relatively recent field in the study of computer forensics technology, digital forensic techniques are garnering increasing attention among researchers. Many experts have dedicated their efforts to developing various models for digital forensics. One notable tool employed by investigators is digital forensics triage, which facilitates the rapid retrieval of pertinent evidence. This expedites the investigation process for law enforcement agencies, minimizing the time and resources required.

2.2.2. Cloud forensics

Cloud forensics employs distinct methodologies to address events occurring within cloud environments. Establishing cloud service methods or frameworks to support cloud investigations is considered a crucial area of focus for researchers. Cloud forensics scenarios often encounter challenges related to distributed data across multiple server addresses, maintaining the integrity of the evidence chain, handling unstructured data formats, and other factors that differ from traditional evidence-gathering environments. In a public cloud setting, migrated data may represent a replicated version of actual data stored in cloud servers, and service providers can utilize software to maintain comprehensive usage records for all users, including login details, log files, and IP addresses.

2.2.3. Data mining

The application of data mining approaches proves highly valuable in mining and analyzing data for forensic evaluation, thanks to their sophisticated algorithms. Presently, various data mining techniques are employed to identify relevant digital evidence for the purpose of proving criminal activities and presenting them to public safety agencies or courts. Many studies have utilized classification or clustering techniques to segregate log data in cloud environments. Classification systems can generally be categorized into supervised and unsupervised classes. Supervised classes assign samples to predefined discrete categories, while unsupervised classes allocate samples to categories without predefined labels. Forensic frameworks for cloud environments, based on layered concepts, have demonstrated effectiveness in addressing specific problems across different domains. For cloud forensics analysis, clustering algorithms from the field of data mining are utilized. While individual clustering algorithms may have their own limitations and errors, combining their advantages can lead to improved performance. Data fusion techniques aim to integrate classifiers to produce superior results compared to using a single algorithm [7].

2.3. Blockchain

In general, blockchain technology has the potential to enhance transparency at various stages, including aiding investigators in accurately identifying data sources during the early phases of an investigation, reducing data storage requirements, and improving the efficiency of transaction analysis, thereby reducing investigation costs. Numerous studies focusing on Internet of Things (IoT) digital forensics confirm that the use of blockchain technology on private blockchain networks enhances security against attacks. The proposed approach integrates fuzzy hashing into the forensic architecture of the IoT blockchain with the primary objectives of:

- Conducting forensic investigations effectively.
- Ensuring the ongoing integrity of key evidence data, preventing loss or corruption that may occur within insecure evidence systems. Fuzzy hashing serves as a solution to address this concern.
- Leveraging the inherent characteristics of blockchain technology, such as digital forensic immutability and audibility, which are crucial attributes of a well-maintained evidence chain in digital forensics.

2.3.1. Evidence identification and acquisition

Within the context of the IoT, the vast majority of data is digitally recorded at the point of collection. This includes evidence in the form of digital assets gathered from sensors, devices, cloud storage, and various other sources. However, controlling access to these digital assets poses challenges in the context of criminal evidence. To address this issue, the proposed approach involves three key steps. Firstly, the method utilizes a
one-way hash algorithm to identify and create fingerprints for the digital evidence. In cases where multiple versions of digital assets are discovered, each claiming to be the definitive version, a digital fingerprint is generated for each piece of evidence. The contents and inspection activities are then specified as technical evidence (TE) information.

2.3.2. Forensic-chain framework

The proposed solution for digital investigations is a blockchain-based forensic chain of custody. It enables the establishment of a distributed ledger for recording and storing analysis events, findings, and other relevant information (TEs). These TEs are distributed through the blockchain network to all authorized participants. The framework comprises several key components:

- Users and IoT devices: The term "users" refers to individuals involved in the investigation as users, owners, or examiners. All devices, sensors, and IoT infrastructures associated with the case are included in the framework.
- Merkle tree: A Merkle tree is a hash tree that ensures the verification and security of the investigation's TEs. It can aggregate all TEs, validate data within a block, and generate a digital signature for the entire sequence of objects. This allows for verification of whether a transaction is included in a block.
- Block: In the proposed model's blockchain network, the signature of the evidence item can be validated. Each block's header contains attributes such as the preblock hash, version, nonce, timestamp, chained blocks, block state, and the Merkle root. The TE item represents the evidence item's record and is hashed into a Merkle tree.
- Smart contract: A smart contract, stored on the blockchain network, is a computer-executable digital agreement. It enables automatic exchange data, information, and business processes without intermediaries. In the context of digital forensic (DF) investigation, smart contracts can facilitate autonomous discovery of linked evidence items, secure storage on a distributed ledger, and cryptographic encryption for enhanced security.

The proposed approach involves the node receiving transaction evidence, calculating the nonce using proof of work (PoW) consensus, and broadcasting it to the blockchain network. A Merkle tree is constructed based on the validity of previous blocks, incorporating fuzzy hashing to ensure data integrity. Once validated, the new block is added to the current blockchain as a record. To reduce complexity in IoT environments, a simplified PoW algorithm is utilized, decreasing the time required for consensus among nodes. In terms of data integrity, the virtual forensic investigator employs context triggered piecewise hashing with sliding windows as a fuzzy hashing technique to assess document similarities. This helps ensure that the blockchain remains untampered at any node within the network [8].

2.4. Machine learning

2.4.1. Predicting attacks and crimes

Crime analysts find it difficult, time-consuming, and expensive to determine whether a crime was committed by the same criminal due to the high number of attacks and crimes that take place each year [9]. In order to estimate the potential locations of crime scenes in the future through statistical projections, analytical techniques are applied. Whether it be an online attack or a genuine crime, the perpetrator typically chooses a target, a time, and place for his or her criminal act. It may not be able to forecast crimes using simple approaches [10]-[12].

2.5. Deep learning

Deep learning, an increasingly potent field within artificial intelligence, holds significant potential for various applications in digital forensics. Specifically, deep learning techniques can effectively process substantial volumes of data, enabling the identification of patterns and the generation of predictions for future events. Within the proposed framework, it is evident that forensic video analysis can be broadly classified into two essential categories: video type analysis and video content analysis.

2.5.1. Forensic video type analysis

Within the realm of forensic video type analysis, a key objective is to examine the authenticity of a video, determining whether it has been illegally reproduced or tampered with. This analysis also encompasses tasks such as video source identification and video steganography analysis, aimed at uncovering hidden information within the video. Specifically, video source identification serves as crucial evidence in identifying the source camera or device that captured the video or image. Given the widespread use of smart mobile devices, including smartphones, tablets, and WiFi cameras, the issue of video source identification has become paramount in video forensic analysis (VFA). In the past decade, several approaches based on sensor noise patterns (SNP) have been developed for video/image source identification, leveraging characteristics of the camera sensor.
2.5.2. Video contents forensic analysis

The emphasis is placed on the visual elements conveyed in the video or photograph. In traditional forensic video analysis, investigators dedicate considerable effort to manually scrutinizing the content by playing the video and visually identifying evidentiary objects within it. However, this conventional approach proves to be time-consuming and inefficient, particularly when dealing with large quantities of video footage [13]–[16].

3. RESULTS AND DISCUSSION

Digital forensic investigation is an emerging field within the realm of information technology. It encompasses the diverse techniques employed to address offenses involving computer systems, starting from the initial interaction with end-users and progressing throughout the entire system. In this study, our focus lies in summarizing each section that corresponds to specific keywords mentioned within the content. We acknowledge that these approaches may not operate harmoniously in synchronization. Consequently, we assess and evaluate both long-standing methods and the latest techniques employed, providing a comprehensive summary of our findings.

3.1. Artificial intelligence

The integration of AI holds immense promise for enhancing the field of digital forensics, which involves the collection, analysis, and preservation of digital evidence that can be presented in a court of law. By leveraging input data and training, the accuracy of machine learning models becomes a vital metric for identifying the most effective model in detecting relationships and patterns within a dataset. AI-based algorithms have demonstrated remarkable efficacy in detecting risks, preventing criminal activities, and predicting illegal behavior. This translates to improved forecasting capabilities and deeper insights provided by the models [17], [18].

3.2. Cloud

This study suggests a framework for using the fusion clustering approach for digital forensics that based on voting methodology. The framework that is being given is separated into four layers: an abstract layer, a layer for information gathering, a layer for data storage, and a layer for information fusion. To separate the relevance evidence in the information layer, voting-k-means approach is used. Deep learning techniques will be applied in future research in the cloud computing environment for digital forensics to enhance clustering [19], [20].

3.3. Blockchain

In existing digital forensics investigations, data integrity is maintained by an independent central authority. While this approach offers procedural efficiency and convenience, there is a vulnerability where the integrity of prospective evidence could be compromised if the central authority falls victim to an attack by a malicious adversary. Moreover, significant material and human resources are required to ensure the chain of custody and guarantee the investigation's objectivity. To conduct comprehensive digital forensic inquiries in large-scale IoT environments, it is necessary to enhance the current chain of custody method to include a more robust integrity preservation approach and expedite operations. This research explores a blockchain-based forensic investigation framework that underwent a preliminary forensic analysis, considering the diverse array of devices, evidence items, and data formats encountered within the intricate IoT ecosystem [21].

3.4. Machine learning

Analysis gets challenging as data volumes rise, and producing error-free results is all but impossible. In order to automate the process, machine learning techniques can be helpful in this stage. If individually gathered, the dataset is trained first, if pre-trained, it is filtered in accordance with the specifications. For reliable findings, datasets are crucial. There are many internet resources available. The second phase involves using the machine learning algorithm. If supervised learning is chosen as the method [22], [23].

3.5. Deep learning

It is important to highlight those digital forensic investigations often rely on low-quality CCTV footage to extract potential evidence items. In this study, we have constructed a framework specifically designed for video-based digital forensics investigations. Additionally, we have devised a technique aimed at enhancing video quality to maximize the extraction of evidentiary elements. Our approach focuses on reversibly extracting additional evidence pieces. By employing this technique, it can aid in both anti-crime efforts and swift responses when criminal acts or behaviors are detected. In future endeavors, we aim to establish stronger connections between newly discovered evidence items and existing ones, further strengthening the investigative process [24], [25].

4. CONCLUSION

Within digital forensics a smart contract automatically logs every evidence transfer that takes place throughout a digital forensic inquiry on the blockchain, including the address to which the evidence is sent, its current status, the permission level, the time and date. Cloud computing’s low-cost services are made possible through the use of enormous data centers for storage across many jurisdictions and multi-tenant hosting by virtual servers. ML can be used to spot suspicious or dangerous conduct in connection with several types of crimes. The main goal of this study is to provide a high-level overview of AI and its potential applications in digital forensics. Several of the present-day difficulties in digital forensics can be solved by AI methods.

ACKNOWLEDGEMENTS

The authors would like to thank Professor Dhwaniket Kamble, faculty at Bharati Vidyapeeth Deemed to be University, for guiding us with this topic and for being supportive to the team.

REFERENCES


Enhancements in the world of digital forensics (Krishna Sanjay Vaddi)


BIOGRAPHIES OF AUTHORS

Krishna Sanjay Vaddi is pursuing Engineering in Computer Science and Business Systems in Bharati Vidyapeeth Deemed University, Department of Engineering and Technology, Navi Mumbai. He has interests in machine learning, data science, and artificial intelligence. He can be contacted at email: ks8978082373@gmail.com.

Dhwaniket Kamble is working as an assistant professor in Bharti Vidyapeeth Deemed University, Department of Engineering and Technology, in Computer Science and Engineering Department. His academic qualification is Ph.D. (Pursuing), M.E. (IT), B.E. (IT), Diploma (IT). His research area includes digital forensics, ethical hacking, cyber security and cyber laws, design thinking and software, and engineering. He has published various international journal papers. He can be contacted at email: drkamble@bvucoep.edu.in.

Raj Vaingankar is pursuing Engineering in Computer Science and Business Systems in Bharati Vidyapeeth Deemed University, Department of Engineering and Technology, Navi Mumbai. His current research interests are data science, machine learning, and android development. He can be contacted at email: rajvaingankar2017@gmail.com.

Tushar Khatri is pursuing Engineering in Computer Science and Business Systems in Bharati Vidyapeeth Deemed University, Department of Engineering and Technology, Navi Mumbai. His current research interest is in the field of machine learning, data science, and cybersecurity. He can be contacted at email: khatritushar718@gmail.com.

Pranil Bhalerao is pursuing Engineering in Computer Science and Business Systems in Bharati Vidyapeeth Deemed University, Department of Engineering and Technology, Navi Mumbai. His current research interests are data science, machine learning, and android development. He can be contacted at email: pranilbhalerao01@gmail.com.