IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 November 2023, accepted 12 December 2023, date of publication 15 December 2023,
date of current version 20 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343360

RESEARCH ARTICLE

Cryptographic Techniques for Data Privacy in Digital Forensics

TAIWO BLESSING OGUNSEYI[1] AND OLUWASOLA MARY ADEDAYO[2], (Member, IEEE)
[1]International Faculty of Applied Technology, Yibin University, Yibin, Sichuan 644000, China
[2]Department of Applied Computer Science, The University of Winnipeg, Winnipeg, MB R3B 2E9, Canada

Corresponding author: Oluwasola Mary Adedayo (m.adedayo@uwinnipeg.ca)

This work was supported by The University of Winnipeg (Grant ID: 16792).

ABSTRACT The acquisition and analysis of data in digital forensics raise different data privacy challenges. Many existing works on digital forensic readiness discuss what information should be stored and how to collect relevant data to facilitate investigations. However, the cost of this readiness often directly impacts the privacy of innocent third parties and suspects if the collected information is irrelevant. Approaches that have been suggested for privacy-preserving digital forensics focus on the use of policy, non-cryptography-based, and cryptography-based solutions. Cryptographic techniques have been proposed to address issues of data privacy during data analysis. As the utilization of some of these cryptographic techniques continues to increase, it is important to evaluate their applicability and challenges in relation to digital forensics processes. This study provides digital forensics investigators and researchers with a roadmap to understanding the data privacy challenges in digital forensics and examines the various privacy techniques that can be utilized to tackle these challenges. Specifically, we review the cryptographic techniques applied for privacy protection in digital forensics and categorize them within the context of whether they support trusted third parties, multiple investigators, and multi-keyword searches. We highlight some of the drawbacks of utilizing cryptography-based methods in privacy-preserving digital forensics and suggest potential solutions to the identified shortcomings. In addition, we propose a conceptual privacy-preserving digital forensics (PPDF) model that is based on the use of cryptographic techniques and analyze the model within the context of the above-mentioned factors. An evaluation of the model is provided through a consideration of identified factors that may affect an investigation. Lastly, we provide an analysis of how existing principles for preserving privacy in digital forensics are addressed in our PPDF model. Our evaluation shows that the model aligns with many of the existing privacy principles recommended for privacy protection in digital forensics.

INDEX TERMS Cryptographic techniques, data privacy, digital forensics, forensic readiness, privacy-preserving digital forensics.

I. INTRODUCTION

The rise in the rate of occurrence of cybercrimes can, on one hand, be attributed to the increasing use of interconnected computers and hand-held devices and their ability to store huge amounts of information. On the other hand, it could be attributed to the increased level of digitization [1]. This advancement has made it possible for a user to utilize multiple devices and to access numerous digital services daily [2], which in a way provide digital footprints of the user's everyday life. In addition, this advancement has caused a growing digital dependence and has aided digital evidence in finding its way to the courtrooms [3].

Digital forensics has become a part of many investigations in cases where computers and digital devices have been used to facilitate a crime, where they are the object of the crime, or in cases where they contain information relevant to an incident [4]. The significant development that has been experienced in the field of digital forensics over the last two decades can be attributed to the increased number of scientific research that now exists and the engaging initiatives from Organizations such as the National Institute

The associate editor coordinating the review of this manuscript and approving it for publication was Barbara Masucci.

of Standards and Technology. These initiatives, for example, the Computer Forensics Tools Testing (CFTT), National Software Reference Library (NSRL), and the Computer Forensics Reference Data Sets (CFReDS) [5], [6] have been instrumental in the validation of new tools, and the provision of research data sets that have led to the increased and diverse contributions in the field. Digital forensics as an essential domain of forensics seeks to extract evidence from computers and other digital devices to help uncover crime. Digital contents including audio data, images, videos, logs, emails, metadata, cache data, etc. existing on many devices can be used by law enforcement to understand the details of an event or find supporting evidence during an investigation. In many cases however, these digital devices contain other information including personal, business-related, health-related, financial records, and confidential information that may be exposed during the analysis of the device despite their irrelevance to the event being investigated. Given that forensic investigators usually have full access to devices that are considered pertinent to an incident, access to this information threatens the privacy of those whose information may be on the device [7].

Addressing data privacy in digital forensics seems to be contradicting as the latter involves the extraction of all data for investigation, while the former advocates the need for the control of data access. However, the question remains, if a suspect's privacy is infringed on and personal information is revealed in a bid to uncover a crime, what happens when such a suspect is found innocent? Even so, if the device contains information about the device owner's relationship with other third-party individuals, how is the third party's privacy protected?

Privacy protection involves the right to control one's data, including identity, personal data, and personal activities [8]. Although the collection and analysis of such information may sometimes be important, collecting only relevant data during an investigation while ignoring the non-relevant data is a key point for privacy protection in digital forensics [9]. Balancing the needs of a forensic investigator to support a fair trial with the privacy rights of those being investigated or those associated with them is a quest where both aspects conflict with each other [10], [11]. As a branch of forensic science, digital forensics focuses on the application of scientific methods in the investigation of evidence present in digital devices for understanding and reconstructing the sequence of events that have transpired in the generation of the said evidence [12]. It ensures that the digital forensics processes (depicted in Figure 1) of identification, preservation, acquisition, examination, analysis, and presentation of digital evidence are completed in a legally acceptable manner [13]. Regardless of whether the information on a device is relevant or not, the overall goal of these processes is to provide information that supports or refutes a hypothesis about an incident. Although the amount of non-relevant information is often significantly more than the relevant data, there has been less focus on how to preserve an individual's privacy in such data [14] or the effect that such privacy breaches may have on the subject involved.

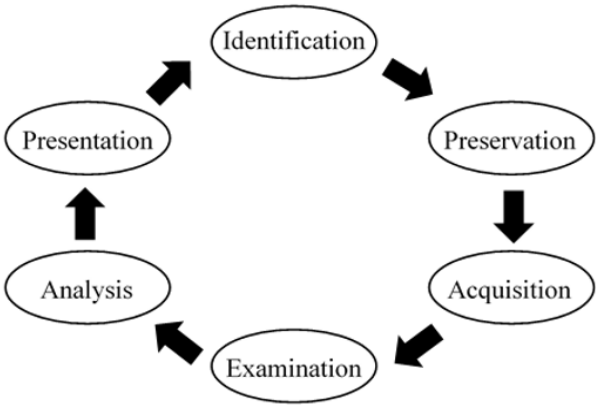

FIGURE 1. Digital forensics processes.

Existing works that have considered the issue of privacy in digital forensics have mostly described principles and investigative guidelines that can be applied to different stages of the forensic processes. The goal of this article is to provide forensic investigators and researchers with a roadmap for developing practical approaches to address some of the data privacy challenges in digital forensics. To achieve this, we examine various privacy techniques that have been proposed to tackle privacy challenges in digital forensics, suggest potential solutions to the drawbacks associated with these techniques, and putting these drawbacks in mind, propose a conceptual model that thrives to preserve user privacy in digital forensics. More specifically, the main contributions of this study are as follows:

- We review the use of cryptographic techniques for privacy protection in digital forensics and analyze their characteristics within the context of whether they support trusted third parties, multiple investigators, and multi-keyword searches. We named these three factors as our analysis factors.
- We propose a simple conceptual model for a privacy-preserving digital forensics investigation process, describe how each of the cryptography-based techniques may be used within the model, present the mathematical representation and algorithm of the model, and examine the model within the context of the identified analysis factors.
- Lastly, we explore some of the analysis factors that may come into play in the use of the model and evaluate how the model aligns with some of the existing principles and guidelines for preserving privacy in digital forensic investigations.

The rest of the paper is organized as follows. Section II gives an overview of privacy challenges and examines how and where privacy concerns may arise in each subdomain of digital forensics during an investigation. In Section III, we discussed the different cryptographic techniques, how they have been utilized for privacy protection in digital forensics, and the drawbacks as well as potential solutions

to applying cryptographic techniques for privacy protection in digital forensics. Section IV presents a privacy-preserving conceptual model for digital forensic investigation, highlighting the entities involved in the model, the investigation model, its analysis factor, and mathematical representation. Section V provides some discussions and evaluation of the model and Section VI contains the conclusions and some future research work.

II. PRIVACY CHALLENGES. WHOSE PRIVACY? WHAT PRIVATE DATA?

To address the issues of privacy in digital forensics and provide some context, it is important to establish how and where private data about individuals may be exposed during investigations. This section discusses privacy challenges from the perspective of whose privacy may be impacted, and what data may be considered private in different domains of digital forensics.

Because digital forensics is mostly concerned with the holistic acquisition and investigation of digital evidence, an underlying principle of digital forensics is that forensic investigations must be reliable, complete, accurate, and verifiable. Existing studies [11], [15], [16], [17] have shown that finding a balance between acquiring case-relevant information and invading users' privacy is a major challenge. In an investigation, data may be retrieved from devices belonging to the suspect, victim, or even witnesses [18]. Despite that the main focus of the digital forensics process is to collect and analyze data that is relevant to the investigation, many digital devices will contain information that belongs to individuals or entities other than the primary user of the device, and an investigator may access such information to even determine their relevancy. Privacy concerns can be viewed with relation to four main entities during investigations: the primary owner of a device, third parties, secondary users, and service providers. These entities are similar to those described by [11].

A. WHOSE PRIVACY?

One aspect of privacy concerns relates to how private information about the primary owner of a device is handled, especially when such information has no benefits to the investigation. Analyzing medical history, browsing and buying patterns, communication metadata, and cell phones, (which tends to hold a greater quantity of information on lifestyles, association, and activities), can provide a complete view of individuals' lives [15]. The direct or indirect exposure of a suspect's information raises privacy concerns if the suspect is later found not guilty. As such, a suspect's confidential information should be kept private by forensics investigators until when found guilty. Satisfying the requirement that there should be no bias in the analysis of digital evidence in an investigation implies that the possibility of being guilty or innocent should be considered, as well as the impact of the investigation on a device owner after the investigation [19].

Even though private data collected on a device will mostly belong to the primary owner in many cases, such data may also expose details about their relationships or interactions with other third parties, that is, an individual who is not directly involved in a crime and/or not a suspect or the victim. For example, communication records and photos or videos about family members, friends, or colleagues may be present on the device. The need for users' informed consent when collecting data from their devices is still a concern in digital forensics and there have been recommendations that informed consent from a user should be simply spelled out, complete, and explicable by the users [20]. While some of the information on a device may be relevant to an investigation, and the examination of data may be done with the owner's consent, such consent does not extend to the private information of third parties. However, obtaining the consent of third parties before viewing their information for relevance or further examination is not feasible in many cases, thus preserving third parties' privacy becomes the responsibility of the investigator.

Another privacy concern relates to data about secondary users (who are unrelated to the investigation) on a shared or multi-user device. In some cases, the device may provide mechanisms to separate each user's data e.g. users with different profiles, however, where this is not the case e.g. in the case of a shared account or when an investigation involves a corporate email server, the privacy of secondary users may be violated during the examination and analysis of data. Such information should be separated and handled in a way that protects the secondary user's privacy.

Privacy concerns may also involve information relating to service providers since their interaction with the device may result in some details of their applications, application interfaces or application functions being stored locally on the device. For example, information about endpoints, confidential information about accounts operated by the primary user, or purchase records may be exposed in the investigation process. Determining whether such information is relevant to an investigation should be done on an individual basis, taking into account the possible implications for both the device owner and the service provider [11].

Lastly, when dealing with service providers or different jurisdictions, the interpretation of privacy is also a challenge. Although privacy is understood in similar ways at a high level, different jurisdictions may understand specific details and interpret privacy differently [20]. Since data may be stored on servers in different countries (e.g., for cloud forensics), what is considered user privacy infringement in one jurisdiction might not be in another jurisdiction. The situation is made more complex when a cloud service provider (CSP) is using services from another CSP located in a separate jurisdiction [21]. Therefore, it is important to be aware of the differences that may exist in such interpretations during an investigation.

B. WHAT PRIVATE DATA?

Tracing back to the origin of digital forensics in the late 1990s, when computer forensics was done by law enforcement personnel with computing expertise, the field has grown to be a significant part of any investigation [22]. Coupled with the growing use of the internet, mobile devices, and other technological advances, different subdomains of digital forensics such as mobile forensics, cloud forensics, network forensics, and Internet of Things (IoT) forensics have emerged to address the challenges of handling various types of data and analysis techniques in different aspects of computing. Despite that the nature of information being examined in different subdomains may differ, the issue of privacy is a concern in almost every subdomain. In what follows, we give an overview of some domains of digital forensics and give a practical indication of how privacy concerns may arise in the different domains as shown in Figure 2.

1) NETWORK FORENSICS

Network forensics is a branch of digital forensics that deals with network-related investigations and may involve the tracking of external and internal network attacks by focusing on inherent network vulnerabilities and communication mechanisms [23]. It involves the identification, capturing, analysis, and reconstruction of network events to discover evidential information about the source of security attacks in a way that preserves the integrity of the data. Network forensics may deal with both dynamic or static data depending on whether the collection and analysis are done on the fly or post-mortem. Some aspects of network forensics include web forensics - which involves the analysis of web browsers and web servers to collect user information; email forensics; and cloud forensics - which focuses on investigating incidents that occur primarily in the cloud environment.

Network data or traffic captured for network forensics contains a lot of information about a user's activities. This may include websites visited, the amount of time spent on each webpage, details of successful and unsuccessful login attempts, unencrypted credentials, records of illegal file download or intellectual property abuse, accessed multimedia files, emails, email attachments, and other documents sent or retrieved over the network [24]. Most of the existing network forensics frameworks focus on the collection of data and have major impacts on the privacy of the primary network user, third parties, and external parties in many cases [25].

2) INTERNET OF THINGS (IoT) FORENSICS

IoT forensics is a relatively new sub-domain of digital forensics [19] that evolved due to the increase in both IoT devices, cybercrimes related to these devices as well as the machine-to-machine (M2M) communication enabled by IoT technology. It focuses on identifying, acquiring, and analyzing evidential information from IoT infrastructures and devices such as wearables, small devices, sensors, connected cars, and RFID for investigative purposes.

IoT forensics processes comprise three levels of forensics namely, device-level forensics, network-level forensics, and cloud-level forensics [26]. Due to the distributed nature and heterogeneity of IoT infrastructures as well as limitations of digital forensics tools, IoT forensics at the network and cloud levels have not been completely used in digital forensics and efforts to design standard frameworks, models, or methods for IoT forensics are still in their infancy [27], [28], [29]. However, there have been attempts to extract and analyze data from devices such as Google Home and Google Assistant apps [30].

Although the fact that many devices only store data for a short time is a challenge for IoT forensics, the traces left behind and the usage nature of IoT devices and frameworks imply that a significant amount of information such as user activities, network traffic, system logs, communication and network usage patterns, and private data about device users can be collected, depending on the device being examined or the scenario. In addition to privacy concerns relating to the primary owner of an IoT device, data collected during IoT forensics may affect the privacy of third parties, external users, and service providers as earlier described.

3) DATABASE FORENSICS

Database forensics is the branch of digital forensics that extracts evidential information from database systems [31]. It is related to the study of metadata and the application of investigative techniques to database contents and metadata. Database forensics investigation focuses on artifacts such as database logs, schema, data structure, metadata (file system), storage engine, etc. [32] and may involve the inspection and validation of the timestamps relating to data updates to validate a user's action.

Because databases are used to store critical and sensitive information in almost all computing systems and applications, they serve as a significant source of information that can be useful for forensic analysis. Much of the critical and sensitive information stored on a database, for example, information about an application, or its operation, an organization, processes on a device, location or transaction histories, etc., creates privacy concerns both for the primary user of the database and other entities that may be involved. Many of the existing process models for database forensics [33] focus on the ability to recover evidential information from a database and do not consider how the privacy of those whose data are stored in the database being investigated may be impacted.

4) MULTIMEDIA FORENSICS

Multimedia forensics involves the attempts to explore, analyze, and retrieve information about multimedia such as images, audio, video, and text. Specifically, it is the analysis of digital multimedia content to produce evidence in the forensics domain [34]. Image forensics as an aspect

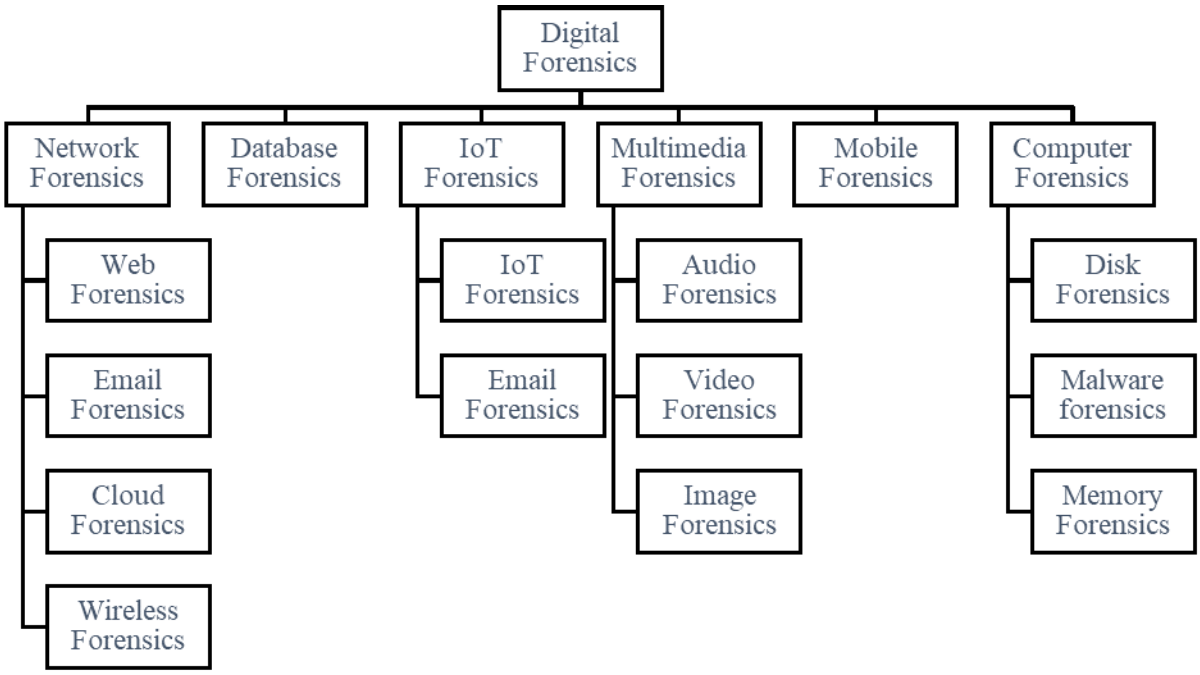

FIGURE 2. Digital forensics subdomains.

of multimedia forensics investigates images by analyzing the authenticity and integrity of data to detect forgeries or manipulations as well as trace the history of the image. In audio forensics, tools and techniques of audio engineering and digital signal processing are applied to study audio data as part of a legal proceeding or for either civil or criminal investigations [35]. Video forensics focuses on the examination, comparison, and evaluation of video for investigative purposes.

While multimedia data being examined may be relevant to an investigation, it may also contain images, conversations, or videos of other individuals or portions that may be irrelevant to an investigation. For example, in audio or video forensics, a recording can provide a real-time, eyewitness account of an event [36] but may also capture personal conversations or events that the entities involved may prefer not to disclose. Also, images being analyzed may look compromising for an entity, even though they may eventually be found to be unauthentic.

5) COMPUTER FORENSICS

Computer forensics focuses on the procedure of obtaining and analyzing computer-related information including data files, hard disk, file storage, hard disk, etc. [34]. Some aspects of computer forensics include disk forensics, memory forensics, and file system forensics. Disk forensics usually involves the process of extracting the content of a file or recovering the contents of a deleted file [36] from a disk drive. The process used to achieve this depends on the data in the disk or its partition.

In memory forensics, live analysis of the RAM can provide detailed information on executed commands in a system, running processes, internet history, system credentials, etc. [37]. File system forensics focuses on the application of knowledge about a file system in discovering evidence and recovering deleted data [38]. Whereas the ability to retrieve metadata, access encrypted data, find hidden information, or extract data from unallocated spaces on a disk may have significant benefits for digital forensics investigations, many of these processes carry the risk of exposing sensitive private data that may turn out to be irrelevant to an investigation.

6) MOBILE FORENSICS

Mobile device forensics involves the analysis of mobile phones and devices to recover digital evidence. From databases that store GPS records, chat history, and messages, to records and logs about applications installed on a phone or mobile device, the collection and analysis of data from mobile devices can expose several private details about the device's user, third parties, secondary users or even service providers, depending on the scenario being investigated.

Lastly, we note that although collection and analysis of evidence (including digital evidence) are expected to follow certain rules in different countries [39], the need to follow all ''reasonable lines of inquiry'' in investigations requires that privacy concerns are considered from the onset of an investigation and built into the digital forensics processes. Unfortunately, many of the existing digital forensics investigation models focus mainly on technical aspects of investigations and do not address the issue of privacy in digital forensics. The works [10] and [11], that have considered the issue of privacy in digital forensics have mostly provided principles and investigative guidelines that can be applied to different stages of the forensic processes to enhance privacy

TABLE 1. Summary of the privacy principles and policies for digital investigation.

S/No	Privacy-Principles [11]	Privacy-Policies [40]
1	Before conducting an investigation, it is crucial to define its scope to ensure it is proportionate and justifiable. Privacy should be safeguarded through appropriate measures, and documented for transparency.	Make two identical hard disk copies and leave one in an environment trusted by the affected party.
2	Full-scale data extraction should only be used when targeted methods could jeopardize the investigation's integrity due to a significant risk.	Remove any unneeded data using specialized erasure tools, such as Evidence Eliminator.
3	Full-scale data extraction must be justified and supported by evidence, aligning with the specific circumstances of the ongoing investigation.	Limit the search for evidence to the goal of the investigation.
4	The investigation scope should adapt as needed, with maximum privacy protection as the initial approach. Examination methods may expand if necessary to serve the investigation's purpose.	Handle time-stamped events in the strictest confidence.
5	Investigators should recognize when they have reached an investigation threshold and halt further probing activities.	Obtain packet acknowledgment via the use of a token rather than the IP address.
6	The use of examined data for investigative purposes should undergo legal and procedural oversight.	Safely stores all internal transaction logs.
7	When targeted data extraction is not feasible, consider screening methods and criteria before a comprehensive data review. Evaluate their suitability and use them before resorting to a full data examination, as needed.	Preserve event logs in external nodes.
8	The investigating authority must document its decision-making processes in creating an examination strategy and make them accessible for third-party evaluation to ensure transparency.	Ensure that organizational policy describes actionable items related to attacks.
9	Investigating authorities should take steps to define and carry out digital device investigations consistently.	Establish policies to safeguard backed-up data relevant to an investigation.
10	In cases of a suitable relationship between the investigating authority and the device owner, prompt communication of investigative strategies and any alterations is recommended.	Handle disposal of data in a secure manner.

preservation during analysis but there are only a handful of solutions that implement or support these guidelines practically. In Section V, we discuss how some of these principles align with our proposed conceptual approach for privacy preservation in digital forensics.

III. PRIVACY-PRESERVING TECHNIQUES FOR DIGITAL FORENSICS

Generally, techniques for data privacy protection for digital forensics can be categorized into policy-based, non-cryptography-based, and cryptography-based approaches. Policy-based approaches give data owners insight into how their private data should be collected, used, and disclosed if needed [9]. It also provides insight into how policies are developed to achieve privacy objectives without hindrance to law enforcement agents during criminal investigations and to restrict access to unrelated files [40]. Existing works that have addressed privacy concerns in digital forensics through the specification of principles and guidelines provide details that can be incorporated into the definition of policies relating to privacy in digital forensics. For instance, [11] suggested a set of privacy-preserving data processing principles that define conduct that is indicative of privacy protection. These principles contain a set of investigative behaviors designed to balance the requirement for effective investigative processes with the need to prevent unnecessary invasion of privacy, particularly during the data extraction and examination stages of digital forensics. The principles are summarized in Table 1. Similarly, [40] proposed policies that could protect privacy, both from the user's perspective and the investigator's perspective, without hindering law enforcement investigations of crimes as depicted in Table 1.

Non-cryptographic-based techniques for privacy protection in digital forensics mainly focus on the use of blockchain technology [41]. Blockchain technology as an emerging technology has been recently adopted for privacy protection in digital forensics, particularly in, cloud forensics [42], [43], IoT forensics [44], [45], mobile forensics [46], multimedia forensics [47], and data management (for maintaining a chain of custody) [48]. This is due to its decentralized and transparent information-sharing approach. However, the adoption of this technology for privacy protection in digital forensics, especially in the traditional digital forensics subdomain, is still in its infancy stage [49].

Cryptography-based techniques involve the use of encryption and have been adopted in many aspects of computing to provide confidentiality, integrity, authentication, and non-repudiation [50], so the application of these techniques in preventing or limiting access to information, as well as protecting private data is not new. However, the application of cryptographic techniques in digital forensics is relatively new with only a small number of studies focusing on different domains of digital forensics.

To provide a road map for understanding the use of privacy-preserving techniques in digital forensics, we describe existing studies that have deployed cryptographic techniques that allow the examination and analysis of digital evidence while preserving the privacy of those involved. The description of these studies is based on analysis factors such as: (i) does the study require computation from a trusted or an untrusted third party to function properly, (ii) does it permit multiple investigators to access forensics data, and lastly (iii) does the study support multi-keyword searches. These analysis factors are essential for a feasible privacy-preserving digital forensic solution to consider. The following sections

TABLE 2. Summary of the analyzed privacy-preserving digital forensics studies.

Authors	Year	CT Used	DF Domain	Trusted Third Party	Multiple Investigators	Multi-keyword search
[53]	2019	HE	IoT Forensics	No	Yes	No
[54]	2020	HE	IoT Forensics	No	No	No
[55]	2013	HE	Log Forensics	No	No	No
[56]	2011	HE	Computer Forensics	Yes	No	Yes
[58]	2011	Commutative Encryption & HE	Computer Forensics	Yes	No	No
[59]	2013	Commutative Encryption & HE	Computer Forensics	Yes	No	Yes
[61]	2017	Secret Sharing	Cloud Forensics	Yes	No	No
[62]	2017	Secret Sharing	Cloud Forensics	Yes	No	No
[63]	2013	Secret Sharing	Computer Forensics	Yes	Yes	No
[64]	2013	Secret Sharing	Computer Forensics	Yes	Yes	Yes
[66]	2015	Searchable Encryption	Email Forensics	Yes	Yes	Yes
[67]	2016	Searchable Encryption	Disk Forensics	No	No	No
[68]	2021	IBE	Cloud Forensics	Yes	Yes	No

CT - Cryptographic Techniques, DF - Digital Forensics, IBE - Identity-Based Encryption

provide an overview of each cryptographic technique that can be considered and how they have been applied in digital forensics. We also describe the drawbacks seen in these studies and suggest possible solutions to support the use of these techniques in digital forensics. A summary of the relevant literature is provided in Table 2.

A. HOMOMORPHIC ENCRYPTION

Homomorphic encryption (HE) allows computation on encrypted data without the need to first decrypt the data, learning neither the inputs nor the computed results. HE is classified into three categories; fully homomorphic encryption (FHE), somewhat homomorphic encryption (SWHE), and partial homomorphic encryption (PHE) [51]. HE may be used in digital forensics for data protection, by allowing evidential information to be encrypted and analyzed without first decrypting the data, thus ensuring data privacy of those involved. This technique provides resilience in situations where computations are carried out by an untrusted or potentially compromised party [52].

One of the common use of HE for digital forensics has been for the preservation of logs, particularly in cloud environments or where the privacy of individual logs records need to be preserved while allowing such record to be analyzed together. Log files contain useful information crucial to forensics investigation, but may contain private information that needs to be protected. Reference [53] proposed a logging scheme that considers log segmentation and distributed storage to collect logs from distributed edge nodes and protect log confidentiality by taking into account edge-cloud characteristics. The system utilized a multi-index-chain (MIC) technique and distributed storage cluster to acquire forensics data without relying on a service provider. The index files include information on the distributed log block being shared with MIC peers through its network. Thus, allowing forensics investigators to collect related log blocks based on index files and distributed storage clusters. To ensure log privacy, the authors implemented the partial homomorphic encryption scheme for the proposed system.

In another study, [54] proposed an efficient privacy-preserving IoT-based log management system for digital forensics that captures and preserves forensics logs continuously for IoT devices in a cloud environment. Through the use of homomorphic encryption, the authors designed an automated and secure log collection model that is capable of preserving smart environment logs from distributed edge nodes in a fog-enabled cloud environment. To preserve and transfer logs securely from IoT devices to the cloud with consideration for delay sensitivity and task offloading, they introduced a fog layer amid the IoT and cloud layer. The three layers offer different security controls for log secrecy and privacy to tackle multi-stakeholder (multi-tenancy) issues and log alteration. Similarly, [55] utilized the (partial) homomorphic encryption scheme for secure log management using a Tor network. The study designed a secure logging system that ensures the confidentiality and integrity of the logs by applying HE for encrypted operations on the logs while the Tor network improves the privacy and security of log data while in transmission.

In terms of handling other types of forensic data, [56] designed a privacy-preserving multiple keyword search system over encrypted data which keeps the investigator subject confidential and protects irrelevant data from the investigator. The system requires a server administrator who is in charge of a suspect's encrypted data to receive some set of case-related encrypted multiple keywords from an investigator. The administrator searches these keywords against an encrypted suspect's data and returns the resulting data to the investigator who then decrypts them for investigation. More specifically, their scheme supports both conjunctive and disjunctive keyword searches over encrypted data to help generate robust investigation data. The conjunctive keyword search returns documents containing all of the several keywords, while the disjunctive keyword search returns aggregated documents containing either one of the keywords or all of them. While most of the studies examined do not require a trusted third party, which suggests that either the systems do not require any input from a third party, or could work with an untrusted third party while still preserving privacy, they lack a multi-keyword search. Therefore, investigators can not submit multiple search keywords to retrieve the most relevant data.

B. COMMUTATIVE ENCRYPTION

Commutative encryption enables plaintext to be encrypted more than once using different users' public keys and does not require decryption before the encryption/re-encryption process [57]. Commutative encryption can be used for privacy protection in digital forensics by allowing both investigator and server administrator (or a device owner) to encrypt evidential information using different encryption keys. Hence, ensuring that an investigator has access to case-related information only and that the relevant data comes from the server (or device) without alteration. In general, this technique may be used to address the challenge of getting relevant data in situations where information (relevant and irrelevant) is stored with a third-party service provider.

A design for privacy-preserving forensic investigations for shared servers was proposed by [58]. The proposed system allows the server administrator to encrypt all data of interest that are stored on the server; this prevents the investigator from learning any data. The investigator then encrypts case-related keywords and sends them to the server administrator. The administrator searches for the relevant keywords from the encrypted data and returns the relevant data to the investigator and the investigator decrypts the data to perform analysis on such relevant data. [59] improved the proposed system to include a verification technique such that the authenticity and integrity of the collected encrypted data from the administrator can be verified to know whether the presented evidence is actually from the server without any alteration. However, both systems require a trusted third party to function well, which could be sometimes infeasible. In addition, the systems do not support queries from multiple investigators, which could be a challenge in a situation where two or more investigators are required.

C. SECRET SHARING

Secret sharing as a cryptographic tool can be applied in a situation where access to sensitive information has to be protected by more than one party. It is a scheme in which different shares of a secret are distributed to parties such that only a fixed subset of parties can reconstruct the secret [60]. Secret sharing may be used to preserve privacy in digital forensics by distributing a secret (suspect's data) into n pieces/data files and storing them in different locations to prevent leaking. Each data file holds no intelligible information about the secret, and the original secret cannot be reconstructed from any one separate file. This may be particularly useful in situations involving the cloud environment or service provider, or where the data of interest is maintained by a server administrator.

Focusing on cloud forensics, [61] proposed a solution based on secret sharing and message authentication codes (MAC) for robust logging of cloud events for forensic investigations. Since there is at least one logging server in a cloud environment, the proposed system attached a MAC to every data written to the log file by the log server, thus creating a chain to avert an attacker from modifying events without being detected. As a further security buffer, each event of the log is divided into n shares and circulated among random nodes in the cloud, then the events are recorded into an immutable database. Using the search space reduction technique in a cloud environment, [62] designed a secure cloud forensic solution, based on the set of inputs that define the historical activity data for the virtual machine, that efficiently searches for forensic evidence within cloud and edge environment without compromising the privacy of non-target. Leveraging standard metering, network logs, and a secret sharing scheme, the study proposes a privacy-preserving solution that reduces the digital forensics target search space during an investigation in the cloud.

Aiming to improve investigation efficiency and privacy of data, [63] proposed the use of a secret sharing scheme secure for keyword searching and matching procedures. The authors treated data files managed by a server administrator as a sequence of words, with each word and keyword treated as secret and divided into n pieces of secret shares. A third party is then required to match each word in a file with each keyword from the investigator. Precisely, the third party matches the shares of each word to the shares of each keyword given by the investigator until a match is found. Once a match is found, t shares of all remaining words of the same file are forwarded to the investigator to reconstruct the whole file based on the principle of (t,n)-threshold secret sharing. Data integrity and authenticity were guaranteed in the system by utilizing a digital signature [64]. The proposed system can also be queried by multiple investigators. One major

downside is the inability of an investigator to query forensics data using multiple keywords.

D. SEARCHABLE ENCRYPTION

Searchable encryption (SE) is an encryption technique that allows search operations over encrypted data. SE can either be searchable symmetric encryption (SSE) or Public-Key Searchable Encryption (PKSE) [65]. SE can be used to preserve privacy in digital forensics by allowing case-relevant keywords to be searched over encrypted evidential information, without direct access to suspect personal information. Targeting email forensics, [66], presented a privacy-preserving email forensics system that analyzes email data in a corporate environment. The system enabled non-interactive threshold keyword searches on encrypted emails by utilizing searchable encryption. More specifically, an investigator with the encrypted data searches the encrypted data for selected keyword searches. The search process reveals the content of an email if it contains at least t number of keywords amongst those that the investigator is searching for. Otherwise, the investigator learns nothing about the content of the email or whether any of the selected keywords are contained in the encrypted data. As a follow-up to [66], the authors [67] proposed an improvement to the above-described system and implemented it for disk image forensics. The system included an additional step of pre-processing disk images before applying a protection mechanism (encryption).

E. IDENTITY-BASED ENCRYPTION (IBE)

Identity-based encryption is a public key encryption in which a user/sender can generate a public key from a known unique identifier such as the email address of the receiver, and a trusted third-party server calculates a corresponding private key from the public key. Reference [68] proposed an IBE-based secure cloud storage system that is compatible with cloud forensics and supports digital forensics investigations, by using multiple public-key generators (PKG) to generate the (encryption) keys. The system permits legal authority or an investigator to act as a party in the key generation in collaboration with another trusted key generation authority which acts as the other PKG. Whenever the need for a forensic investigation arises in the cloud environment, the legal authority collaborates with the trusted key generation authority to re-generate the private key and decrypt the file contents, then provide the decrypted files for further forensics analysis. Since neither the legal authority nor the trusted key generation authority can act alone to generate the private key for decryption, data access can be controlled. However, the shortcomings are the file content has to be decrypted first before any forensics analysis and the scheme only permits single keyword search which makes it impractical when large file content is to be analyzed.

F. DRAWBACKS OF DEPLOYING CRYPTOGRAPHIC TECHNIQUES AND POTENTIAL SOLUTIONS

As seen from the literature reviewed and Table 1, very few studies have explored the use of cryptography-based techniques for mitigating data privacy challenges in digital forensics. This is in part due to the following reasons as seen in existing studies. Possible ways of addressing the identified issues are also described below.

- The resulting size of ciphertext from encrypted evidential information is too large. The collection stage of the digital forensics process often results in an enormous dataset and encrypting this dataset produces ciphertext with a substantially even bigger size. This has the potential of making an investigation less thorough and requiring more resources for data processing. A possible solution for this challenge is to encrypt only case-relevant data, implying that the verification of keywords to determine data relevancy should be included in digital forensics models to ascertain whether compiled keywords are case-relevant or not. While this will not fully eliminate some of the privacy concerns earlier described, it has the potential to limit their occurrence [69].
- With the increased rate at which data of interest in a forensic investigation may be generated, investigators now require more resources and time to collect, examine, and analyze forensic data, regardless of whether the data is encrypted or not. To address this challenge, investigators must work with relevant authorities and service providers to understand where to look for relevant evidence. To achieve this, it may be important to focus on who, what, when, where, why, and how questions when reviewing possible sources of evidence.
- Some of the cryptography-based schemes lack a verification method to ascertain whether the keyword searches are case-relevant or not. Case-relevant keywords are instrumental in reducing the volume of data collected and encrypted and reducing the potential access to private data. A careful selection of the keywords through collaboration between investigators and other stakeholders (e.g., law enforcement or a service provider) would be necessary to address some aspects of this challenge. In addition, models for privacy-preserving digital forensics should integrate keyword verification techniques, to determine what data should be collected and/or encrypted and ultimately reduce investigation time.
- The management of encryption and decryption keys in existing cryptography-based techniques is a challenge that can limit the usability of some solutions. While this challenge may be addressed simply by minimizing the transfer of keys between parties involved in an investigation, a dedicated entity that generates key pairs, encrypts data, and decrypts the ciphertext should be integrated into privacy-preserving digital forensics models.

With all these drawbacks and solutions in mind, we propose a simple conceptual privacy-preserving digital forensics model that utilizes a cryptography-based scheme in the following section. We then explain how the different cryptographic techniques earlier described may be incorporated into this conceptual model and other factors that may be considered in their use.

IV. PRIVACY-PRESERVING DIGITAL FORENSICS MODEL

This section describes our conceptual model for privacy-preserving digital forensics (PPDF) as shown in Figure 3. The model employs the use of encryption in the handling of evidential data throughout the digital forensic investigation process. In what follows, we first describe the entities involved in the model and their responsibilities, then discuss the model based on the digital forensic investigation processes. Furthermore, we examine how the model overcomes the shortcomings highlighted in Section III and lastly present the mathematical description of the model. The model is discussed in light of how the different cryptographic techniques discussed in Section III may be used at each stage.

A. PRIVACY-PRESERVING MODEL ENTITIES

The main goal of the privacy-preserving digital forensics model is to allow the identification, preservation, acquisition, analysis, documentation, and presentation of evidence from digital devices while preserving the privacy of the individuals involved. The entities that exist in the privacy-preserving model, as well as the assumptions relating to each entity within the model, are described as follows.

- User: The user is an individual who is involved in a case that requires a digital forensics investigation. The user could be a suspect, an accused person, a victim, or a third party. Data that belongs to a user, especially those that are non-relevant to the investigation should be kept private. Hence, it is important for investigators to perform their investigative role without compromising the user's privacy.
- Investigator: The investigator may be a law enforcement agent (LEA) or someone involved in the examination and analysis of digital evidence and may interact with both the user(s) and the service provider. The investigator is mostly responsible for generating an asymmetric key pair (Pri_{key}, Pub_{key}), and making the public key (Pub_{key}) available to the Service Provider (SP) but the creation of the key pair may be delegated to a cooperating user if necessary. The investigator also determines case-relevant keywords, required to ascertain the data that should be extracted. The investigator is also required to work closely with the SP(s) to retrieve evidence within the SP's jurisdiction.
- Service Provider (SP): The service provider is charged with searching the suspect's data, in their custody, for case-relevant keywords, encrypting relevant data, and sending it to the investigator(s). We assume that this entity only receives input from the investigator, and does not collude with the user, hence, they can be trusted.

B. CONCEPTUAL MODEL BASED ON THE DIGITAL INVESTIGATION PROCESSES

The digital forensics investigation process is divided into six stages (as described in Section I). However, in this section, we categorize the processes into four stages based on the four privacy-preserving stages depicted in Figure 3 and discuss the possible applications and limitations of the cryptographic techniques described in Section III at each stage of the privacy-preserving model.

1) STAGE 1: PREPARATION / KEY GENERATION

The first step of the digital forensic process is the identification stage. At this stage, an investigator recognizes the nature of an incident, prepares the tools and equipment needed during the investigation, and defines the tasks to be accomplished during the investigation. Furthermore, and while this may be a difficult task, they also work with other entities (users and service providers), to ensure data and user privacy. This stage can also be referred to as the forensics readiness stage. As part of the preparation for the investigation, the investigator is expected to obtain the necessary approval or warrant for investigation and devise appropriate data preservation, and chain of custody mechanisms. The goal at this stage is to ensure that the right resources, both material and human are employed for evidence preservation and to prevent irretrievable damage to digital evidence due to their volatility.

No cryptographic technique is required at this stage since there is no interaction with data. However, measures to ensure data privacy should be integrated into the stages right from the beginning and included as part of a readiness plan. This should include generating encryption keys to facilitate privacy preservation throughout the investigation process. In the case of a cooperating user (i.e., a user who shows a willingness to cooperate with the investigator during the investigation process e.g. the victim in an incident), the investigator may decide to delegate the responsibility of generating the encryption key to the user to limit access to their personal information and foster their involvement in maintaining their privacy during the investigation. The public key is made available to the investigator and/or the service provider as the case may be. On the other hand, if the user is an uncooperating user, either because they are not found at the crime scene, could not be located or for any other reason, the investigator generates the encryption key pair and makes the public key available to a service provider if necessary. In practice, law enforcement agents who deal with digital evidence are trained to properly utilize recent technology to enhance their investigation [70], [71]. Therefore, it is natural to assume that LEAs can generate and manage the public key, generated either by the user or the investigator.

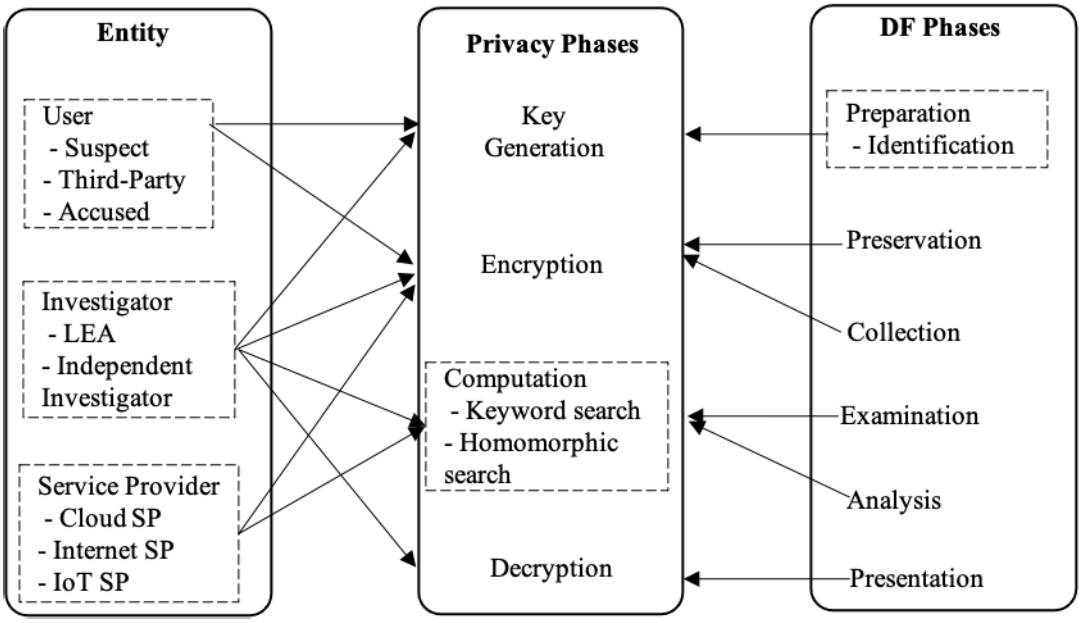

FIGURE 3. An overview of the proposed PPDF model.

The decision regarding whether or not a user generates the key pair should be made on a case-by-case basis depending on well-defined factors that have been identified based on the incident type, user request, or other conditions relevant to the case. In a situation where a user is delegated to generate the public key, the custody of the corresponding private key is discussed in Section V-A. Public key sharing is done based on the location of the evidential information, either in the cloud or non-cloud environment. For preparation in a cloud-based environment, the public key is shared with a cloud service provider, while the LEA manages the non-cloud environment. The key generation step in this stage is an important step that is required regardless of which encryption techniques described in Section III are eventually used. This step is also required regardless of the data acquisition method later used, either static or live acquisition. A flowchart representing the steps involved in this stage is depicted in Figure 4.

2) STAGE 2: PRESERVATION / ENCRYPTION

Next to the identification and preparation stage of the digital forensics process is the preservation and acquisition stage. This stage focuses on the isolation, preservation, and collection of evidential data from the digital crime scene. Collected data typically includes both case-relevant and non-relevant data since it may be almost impossible to determine which information is relevant right from the crime scene [72]. As shown in Figure 5, data acquisition and preservation may involve making an image copy of disk drives and other digital objects found at the crime scene, or running data collection tools and writing the output to external storage for servers and other critical devices that cannot be powered off [73].

For our conceptual PPDF model, this stage also includes the generation of keywords that are related to the crime being investigated. The investigator(s) who have oversight of

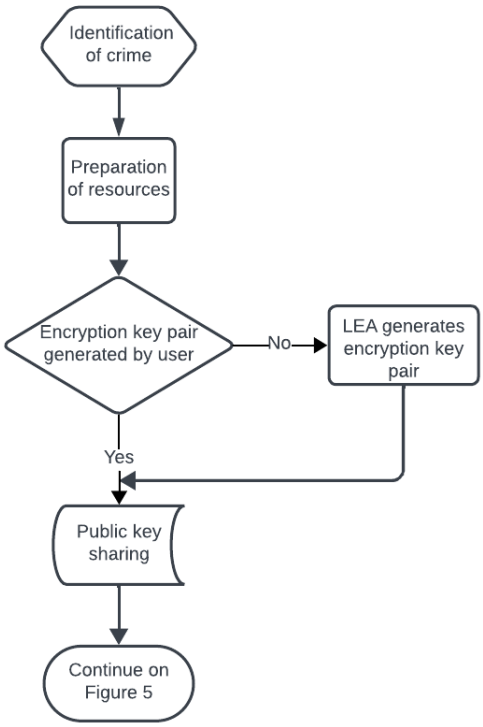

FIGURE 4. A flow diagram of the preparation and key generation stage.

the case, establishes a set of case-related keywords, known as the first keywords. A search of the first keywords is conducted on the collected data without any encryption of the collected data or the keywords. The data retrieved from the first keyword search are then encrypted using the appropriate public key earlier generated (Pub_{key}). To ensure that as much data as possible can be gathered, the first keyword search should use disjunctive keywords as described in [74]. This helps to generate more independent and non-interrelated

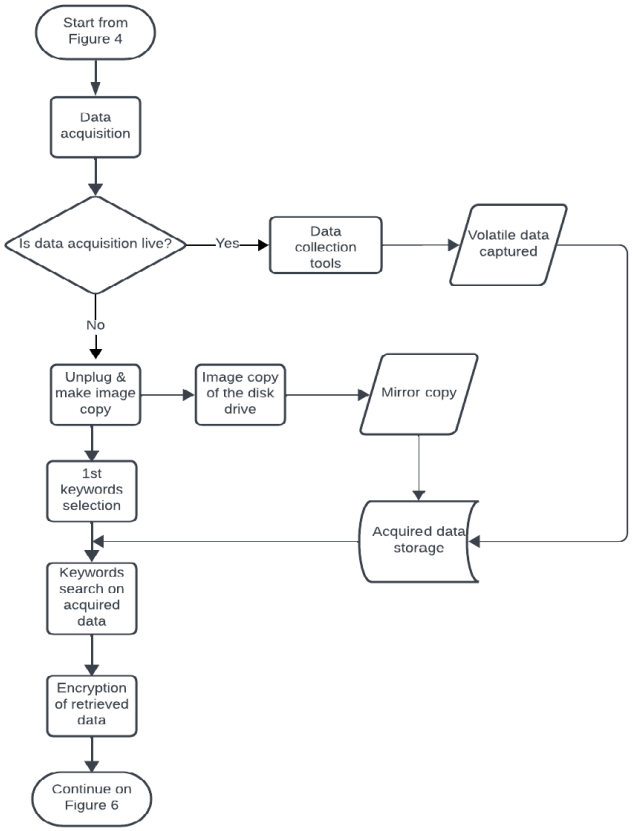

FIGURE 5. A flow diagram of the preservation and encryption stage.

data [75], which gives the investigator access to more case-related information. However, for evidential information with a service provider (e.g. cloud service provider), the first keywords are encrypted by the investigator before being sent to the service provider as further discussed in Section V-A. It is important to note that we differentiate between our use of 'case-related' and 'case-relevant or pertinent' data. Case-related data is any data that can be linked to the criminal case but may not be necessarily relevant and it is associated with the first keyword search, while case-pertinent data are relevant and are associated with the second keyword search described in Stage 3 below.

Apart from ensuring that the data retrieved is related to the case being investigated, one of the goals of this step is to ensure that non-related data is not encrypted, thus significantly reducing the size of the ciphertext generated which needs to be further analyzed. Thinking about privacy concerns, we note that it is possible that information considered to be case-related or case-relevant may still contain private data, however, this must be included as part of the analysis to ensure a holistic view of the data and/or the investigation. For example, if the date of an incident is considered as a search keyword, all files created on this date may be retrieved as related data but their relevance is yet to be determined. Data considered to be relevant is still included in the analysis even though it may contain some private information.

The main privacy-related task in this stage is the encryption of results from the disjunctive keyword search. From the privacy techniques earlier discussed in Section III, homomorphic encryption, commutative encryption, and searchable encryption schemes fit this purpose. This is because the encryption of search results can be easily performed by the investigator alone when any of these three schemes is utilized. The secret sharing scheme requires a suspect's data to be shared among multiple parties, thus, this may not be viable for the conceptual PPDF model as the entities involved in the model (i.e. investigator, user, and service provider) are separated, and would not typically be able to hold portions of data independently or share data. For identity-based encryption, the encryption key is generated based on a user's identity, and a centralized server is required to generate the private key.

To utilize either secret sharing or identity-based encryption for privacy protection in digital forensics, a trusted third party or central server will be required to manage the distribution of suspect's data into n shares and generate a private key for decryption. This requirement can be seen in all related solutions discussed in Section III, where all applications of secret sharing or identity-based encryption require a trusted third party as depicted in Table 1. Thus, the use of encryption techniques for privacy protection in digital forensics requires that realistic assumptions be made and the practical feasibility of a proposed solution be considered.

3) STAGE 3: EXTRACTION / COMPUTATION

After the evidence encryption comes the extraction of case-pertinent data and analysis of the dataset. This stage examines the encrypted case-related evidence to extract data relevant to the investigation and seeks to ensure data confidentiality. To achieve this, another set of keywords, known as the second keywords, that are designed to be very case-specific are established and searched on the encrypted case-related data as described in Figure 6. The main objectives of this stage are to extract case-pertinent information from encrypted data thus having relevant pieces of evidence to work with; and to perform analysis, such as pattern recognition and classification on the encrypted case-pertinent evidence in order to determine the significance of the data, identify evidence patterns, and make conclusions about the case. This is also the stage investigators test their hypothesis.

The second keyword search is performed to obtain a fine-grained result that contains the case-pertinent data on the investigation and reduces the dataset that needs to be examined or analyzed further. The second keyword searches are also performed on the encrypted case-related evidence received from a cloud service provider if any. The resulting ciphertext from the search is then analyzed to discover their relationship to the case investigated. An example of such analysis on the encrypted data includes pattern recognition where evidence can be classified based on similar features

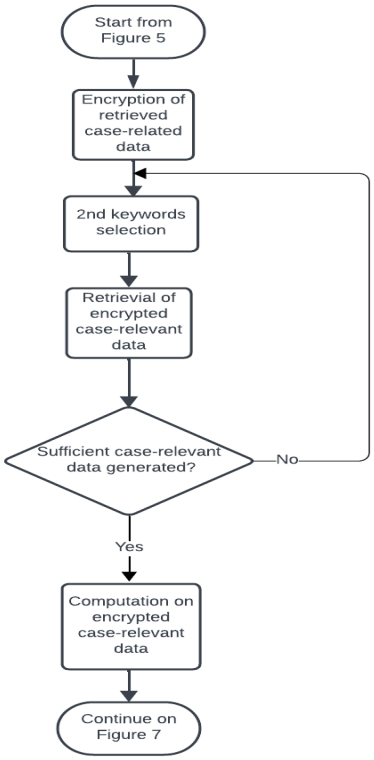

FIGURE 6. A flow diagram of the extraction and computation stage.

as demonstrated in [76]. The second keyword selection and search for case-relevant data may be repeated several times to aid the confirmation or refuting of a hypothesis. This stage of the model achieves data confidentiality and user privacy as investigators only get to work on encrypted data. Moreover, other (non-evidential and non-case related) information is not examined further but stored. An investigator does not have access to this information unless there is a justification for such access, with permission given by a superior investigator, this approach is in line with the third principle of privacy-preserving digital investigation described in [11].

Considering the cryptographic schemes earlier discussed, only the homomorphic encryption and searchable encryption schemes can support this stage of the model. Other cryptography techniques such as commutative encryption, secret sharing scheme, and identity-based encryption, in most cases, would still need to be used in conjunction with the homomorphic encryption scheme to allow computation on encrypted data without decrypting first. Also, the use of secret sharing and identity-based encryption requires the use of a central server or a dedicated third party as earlier discussed.

4) STAGE 4: PRESENTATION / DECRYPTION

The last stage involves the summarization and description of findings, as well as further validation or refuting of hypotheses made during an investigation as shown in Figure 7. The resultant ciphertext from the encrypted analysis is then decrypted using the corresponding private key of the public key scheme used for encryption.

Regardless of whether the key pair used for encryption is generated by the user or an investigator in stage 1, it is the sole responsibility of the investigator to decrypt the encrypted outcome from the search using the associated key. Hence, this justifies why the corresponding private key, for a public key generated by a user, is stored in a private key storage system as further described in Section V-A. It is important to note that several factors come into play in the application of the conceptual model described. In the following section, we examine the factors that may be considered with regard to the model.

An example of a scenario where the model is applicable could be a crime scene where a suspect has been accused of defrauding someone who made an online purchase. We assume that the suspect was apprehended and their smartphone confiscated by an LEA. Following the four stages delineated in our conceptual PPDF model, at the first step of preparation, the LEA defines the tasks to perform during the investigation, acquires a forensic image of the smartphone's media storage, prepares the needed tools and equipment to extract case-related data from the smartphone, and obtains the necessary warrant/approval. In the second stage, case-related keywords are curated based on the crime at hand, this could involve searching text messages, emails, call logs, documents, social media posts, browsing history, and other content for pertinent information. The keywords are searched and the result is encrypted using the public key generated, either by the user or the LEA.

Using a set of case-relevant keywords (second keyword), the encrypted case-related data from the smartphone is further searched to obtain encrypted pertinent data to the investigation. Investigators may refine the keyword searches based on initial findings or narrow down the results as described above. The encrypted case-relevant data is analyzed to uncover patterns and make findings related to the crime. Analysis such as metadata pattern analysis, network traffic analysis, file structure analysis, frequency and location pattern analysis, and correlation analysis could be carried out on the encrypted case-relevant data to discover their relationship to the case investigated. Lastly, the result obtained from all the analyses is decrypted, and the investigator's hypotheses are either validated or refuted.

C. CONCEPTUAL MODEL ANALYSIS FACTORS CONSIDERED

In our examination of the conceptual model, we address the underscored challenges outlined in Section III. These challenges encompass: (i) the conceptual model's capacity to accommodate untrusted third parties, such as service providers, or to function autonomously without external involvement; (ii) the model's handling of queries from

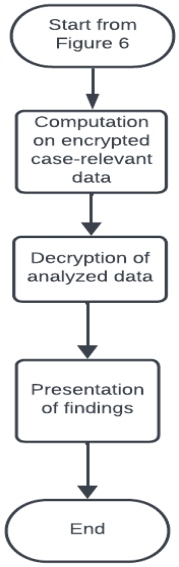

FIGURE 7. A flow diagram of the presentation and decryption stage.

multiple investigators, particularly in scenarios necessitating such interactions; and (iii) the extent to which the model facilitates multi-keyword searches. Subsequently, we explicate the model's performance within these contexts.

The conceptual model presumes that a portion of the user's evidential data resides with a cloud service provider, an assumption commonly made due to the ubiquity and cost-effectiveness of cloud storage. However, it is unwise to rely on the trustworthiness of the cloud service provider in safeguarding evidential information. Consequently, user information is encrypted with a public key shared with the service provider. Furthermore, a set of encrypted keywords is sent to the service provider. Utilizing these encrypted keywords, the service provider executes a disjunctive keyword search on the encrypted dataset, and subsequently transmits the resulting dataset to the investigator, as presented in stage 2. Consequently, the model demonstrates its operability in the presence of an untrusted third party, enhancing its feasibility.

The concept of multiple investigators denotes that two or more investigators can simultaneously access and manipulate the same dataset. This scenario may arise when the investigation spans different regions or countries. Within the conceptual model, diverse investigators can access and collaborate on the same case by sharing pertinent resources, including sets of keywords (both first and second), any encrypted dataset provided by the service provider, and the encrypted resultant dataset obtained from the initial keyword search.

Lastly, the inclusion of multi-keyword searches authorizes investigators to submit multiple keywords for retrieving more comprehensive data, in contrast to the limited results yielded by single-keyword searches. The conceptual model accommodates both single and multi-keyword searches.

TABLE 3. Variables and notations used in the model description.

Symbol	Definition
Pub_{key}	Public key for encryption
Pri_{key}	Private key for decryption
F_{kywd}	First set of keywords searched from the user data
S_{kywd}	Second set of keywords searched from case-related data
U_D	User data (both relevant and non-relevant)
Cr_{kw}	Case-related keywords retrieved from user data
Cr'_{kw}	Encrypted case-related keywords
Crv_{kw}	Case-relevant keywords retrieved from case-related data
Crv'_{kw}	Encrypted case- relevant keywords

The initial keyword search, i.e., the disjunctive search, constitutes a multi-keyword search that yields independent and case-relevant data. Conversely, the second keyword search entails a conjunctive keyword search, retrieving case-pertinent information, as elucidated in Stage 3. In addition to other considerations, the incorporation of keyword searches in the proposed model aims to mitigate the encryption of all user data, both relevant and non-relevant. This approach is necessitated due to the substantial computational complexity associated with cryptography-based privacy techniques and the extensive data typically possessed by users. In light of the aforementioned analysis factors, the conceptual model offers a robust solution to these challenges, thereby furnishing a substantiated proof of concept model.

D. FORMAL SPECIFICATION OF THE CONCEPTUAL MODEL

This section presents the formal description of the conceptual PPDF model. The definition of notations used is presented in Table 3, and the mathematical representation and algorithms of the model are also presented.

Using a generic homomorphic encryption key generation scheme defined in [77], we describe the key generation process depicted in Algorithm 1 for the conceptual model. Algorithm 1 illustrates the key generation process which takes as input $n =$ the dimension of the ideal lattice and $t =$ the bit length of the coefficient with the size being 128 bits and outputs the public key and private key. For the keyword generation, let $k_{n+1} = \{k_1, k_2, \ldots, k_{n+1}\} \in U_D$ be the set of keywords that can be produced from user data and let $k_n = \{k_1, k_2, \ldots, k_n\}$ be the set of first keywords from user data denoted F_{kywd} as shown in table 3. A disjunctive keyword search will be $\{k_1 \cup k_2 \cup \ldots \cup k_n\} \in U_D$. Therefore, $Cr_{kw} = \{k_1 \cup k_2 \cup \ldots \cup k_n\} \in U_D$. Similarly, the second keyword search will be $\{k_1 \cap k_2 \cap \ldots \cap k_m\} \in Cr_{kw}$, and $Crv_{kw} = \{k_1 \cap k_2 \cap \ldots \cap k_m\} \in Cr_{kw}$. Note that $Cr_{kw} \in U_D \gg Crv_{kw} \in U_D$, which implies that $Cr'_{kw} \in U_D \gg Crv'_{kw} \in U_D$ hence $Crv'_{kw} \in Cr'_{kw}$.

The encryption stage involves the encryption of both case-related keywords retrieved through the first keyword search and the encryption of the case-relevant keywords retrieved through the second keyword search as illustrated in Algorithm 2. For the encryption of case-relevant keywords,

Algorithm 1 : Key Generation

Input: dimension n, bit length t
Output: Pub_{key} = (d, r), Pri_{key} = (w_i)
1:Choose a random vector v from
$v(x) = \sum_{i=0}^{n-1} vix^i$: vi is a random t-bit length and $\sum_{i=0}^{n-1} vi \equiv 1 \bmod 2$
2:Compute the resultant d of $v(x)$ and $f(x)$, and the coefficient w_1 of the linear term of $w(x)$
Where $w_x v_x = d \bmod f(x)$
3: If $gcd(wi, d) \neq 1$ then
4: Go to 1
5: else
6: Compute w_0 and $r = \frac{w_0}{w_1} \bmod d$
7: Compute an odd w_i via $w_i = rw_i + 1 \bmod d$ and w_0, w_1. The subscripts are modulo n.
8: Output: $Pub_{key} = (d, r)$, $Pri_{key} = w_i$
9: end

the same process in Algorithm 2 is followed but with different input Public key (d, r) and case-relevant keywords (Crv_{kw}) and output (Crv'_{kw}).

Algorithm 2 : Encryption of Case-Related Keywords

Input: Public key (d, r), case-related keywords (Cr_{kw})
Output: Encrypted case-related keywords (Cr'_{kw})
1: To encrypt plaintext $Cr_{kw} \in \mathbb{Z}$ with $pk = (d, r)$
2: Choose a random noise vector $\overrightarrow{u} = (u_0, u_1, \ldots, u_{n-1})$ with $u_i \in (0, \pm 1)$
3: Compute the resultant ciphertext of Cr_{kw} as
$c = Enc(Cr_{kw}, pk) = [b + 2.\sum_{i=0}^{n-1} uir^i]_d \in [\frac{-d}{2}, \frac{d}{2})$.
4: Set $\overrightarrow{a} = 2\overrightarrow{u} + Cr_{kw}\overrightarrow{e}_1 = (2u_0 + Cr_{kw}, 2u_1, \ldots, 2u_{n-1}) \in \mathbb{Z}^n$ with $\overrightarrow{e}_1 = (1, 0, \ldots, 0)$.
5: Output Cr'_{kw}
6: end

For the investigator(s) to accept or refute some of the hypotheses, the encrypted case-relevant keywords have to be analyzed and conclusions made based on the outcome of the analysis. Following a similar pattern described in [78], we delineated the classification of the encrypted data. Specifically, we outline the K-Means classification algorithm on forensics data, an approach presented in [79], to determine the significance of analyzed data as shown in Algorithm 3, while the decryption algorithm is presented in Algorithm 4.

V. CONSIDERATIONS AND EVALUATION OF THE PPDF MODEL

In this section, we examine some of the factors that may be considered at different stages of the conceptual model. We also discuss an evaluation of the model with respect to some of the existing principles for preserving privacy in digital forensics.

Algorithm 3 : K- Means Computation on Encrypted Data

Input: An encrypted case-relevant keywords $Crv'_{kw} = Enc[k_1 \cap k_2 \cap \ldots \cap k_m]$, the number of clusters q, and the termination condition μ.
Output: Encrypted Cluster $C' = (c'_1, c'_2, \ldots, c'_q)$
1:Randomly selects q data records $C_{(l)} = c_1^{(l)}, c_2^{(l)}, \ldots, c_q^{(l)}$ as the initial clusters where $l = 1$.
2:For each k_i in Crv'_{kw} assign it to the closest cluster $Crv_{i,j}^{(l)}$
3: Compute the counts $m_{i,j}^{(l)}$ of each cluster $m_{i,j}^{(l)} = (1 \leq j \leq q)$ and q local centers $w_i^{(l)} = (w_{i,1}^{(l)}, w_{i,2}^{(l)}, \ldots, w_{i,q}^{(l)})$ where $w_{i,j}^{(l)}$ is a d dimensional point.
4: If $maxDist(c_{i,j}^{(l)}, c_{i,j}^{(l+1)})1 \leq j \leq q > \mu$ the algorithm iterates, otherwise and output the final results
5: Output: $C' = (c'_1, c'_2, \ldots, c'_q)$
6: end

Algorithm 4 : Decryption

Input: Private key $Pri_{key} = w_i$, encrypted cluster $C' = (c'_1, c'_2, \ldots, c'_q)$
Output: A set of clusters in plaintext
1:Compute $[C'.w_i]_d \pmod 2$
2:Let $\overrightarrow{a} \in \mathbb{Z}$, then we recover the plaintext since $\| \overrightarrow{a} \times W \|_\infty < \frac{d}{2}$
3: Output: $C = (c_1, c_2, \ldots, c_q)$
4: end

A. CONSIDERATIONS OF THE PPDF MODEL

As mentioned earlier, when a cooperating user is delegated to generate the asymmetric key pair (Pri_{key}, Pub_{key}), the public key is shared among the entities involved such as the investigator and/or the service provider. The corresponding private key for decryption is stored in a private key storage system by the LEA. This is primarily to prevent private key loss and to enable the investigator to access the key for decryption after the analysis of the encrypted case-relevant data, without having to wait for the user. However, it is worth mentioning that for the investigator to access the private key, there has to be permission issued by a senior investigator who has oversight of the case.

In a case where the user's data is stored with a service provider, an encrypted set of the case-related keywords (first keywords) is sent to the service provider to ensure that they do not become privy to the investigation details. This assumes the use of homomorphic encryption, thus allowing the service provider to perform the search. The service provider performs this encrypted case-related keyword search on the evidence within their jurisdiction. The service provider then encrypts the retrieved data from the first keyword search with the public key shared by the investigator. Afterwards, the service provider sends the encrypted retrieved case-related data to the investigator who then queries it to extract

case-relevant data. The encryption of the first keywords is only necessary when evidential information is with a service provider (cloud environment) and not in a non-cloud environment.

When dealing with a cloud environment, the extraction and computation of data in stage 3 of the PPDF model may be challenging, particularly if the cloud service provider is not willing to grant the investigator access to necessary data. This is a common issue that may be encountered both for privacy-preserving and non-privacy-preserving digital forensics process models. It is also related to the drawback of accessing information in data centers which may be in a jurisdiction different from that of the investigator. For a non-cloud environment, the investigator, in most cases, has the essential devices and/or information to perform the investigation. Therefore, investigators can perform the identified tasks in this stage.

B. EVALUATION OF THE PPDF MODEL

To evaluate the proposed conceptual PPDF model, we consider how each stage of the model aligns with many of the existing principles for privacy protection in digital forensics models. Table 4 shows the categorization of the existing principles into the four stages of the conceptual PPDF model, depicting the alignment of the processes in each stage to each of the principles.

In [9] the author proposed the classification of evidential information into different privacy levels to prevent encrypting entire user's data and to reduce the investigation cost in terms of time and resources. Considering both the user's and the investigator's perspectives, the authors first classified this information into private and non-private for users and relevant and non-relevant for investigators. This resulted in these four groups: non-private and non-relevant, non-private and relevant, private and non-relevant, and lastly private and relevant. Subsequently, they defined three privacy levels for evidential information to enable more efficient privacy-preserving digital forensics investigation. The privacy levels are; direct accessible data (DAD), privacy-preserving accessible data (PAD), and non-accessible data (NAD). For DAD, the data is relevant and non-private so it can be directly extracted and analyzed. The PAD signifies relevant and private data, hence, privacy-preserving technique(s) must be applied during data extraction and analysis, while NAD implies that the data is not relevant to the case and is not accessible to the investigator. This is in line with stages 2 and 3 of the conceptual PPDF model in which only related and relevant data are extracted and the appropriate privacy-preserving technique is applied.

Similarly, using cryptographic techniques and blind signatures, the authors in [10] proposed a system involving a sequential release of private information in digital forensics investigation based on prior knowledge of the private information and proof of a hypothesis. In other words, it explores the feasibility of protecting more sensitive information until knowledge about less sensitive information has been demonstrated. To balance the efficacy of the investigation against user privacy, they propose a scale for partitioning information into privacy-accurate levels where (L_1) denotes partitioned information with a low privacy-accurate level and (L_4) implies a high privacy-accurate level. The four-level privacy-accurate scale is described as follows: (L_1) – the evidence does not divulge any personal information, (L_2) – evidence may refer to personal information, (L_3) – evidence may infer personal information, and (L_4) – evidence undeniably divulges personal information. This somewhat supports the four possible groups of forensics data posited in [9] and is in line with our conceptual PPDF model's case-related and case-relevant approach to evidential information discussed in stages 2 and 3.

TABLE 4. Classification of existing privacy-preserving digital forensics principles.

Ref. No	Preparation Key Generation	Preservation Encryption	Extraction Computation	Presentation Decryption
[9]	–	DAD, PAD, NAD	–	–
[10]	–	L_1, L_2, L_3, L_4	–	–
[11]	PD1, PD4, PD9	PD2, PD3, PD7, PD10	PD5, PD6, PD8	–
[40]	PP3, PP8, PP9, PP10	PP1, PP2, PP4, PP5, PP6, PP7	–	–

L_1-L_4 - privacy accuracy level of information classified based on its private level, PD1-PD10 - Privacy-preserving data processing principles represent the principles for consideration when conducting digital forensics extraction and examination of data from a digital device, while PP1 - PP10 - privacy policies for protecting user information in DF.

Another study in [11] proposed a set of ten privacy-preserving data processing principles for consideration during the extraction and examination of evidential information from digital devices in digital forensics investigation, represented as PD1 - PD10 and summarized in Table 1. Emphasizing the need for balance between the requirement for effective investigative processes with the need to prevent unnecessary invasion of privacy, the principles highlighted the concerns regarding potential privacy invasion caused by the examination of digital devices in criminal investigations. We classified these ten principles (PD1 - PD10) under the first three stages of the conceptual model, as depicted in Table 4, because the principles only consider user privacy concerns from the extraction to the examination stage of evidential information. Notably, the first principle (PD1) which states that *"the scope of any investigation should be defined and evaluated before its implementation to ensure that it is both proportionate and justifiable"* is in line with stage 1 of the conceptual PPDF model, where we underscore that an investigator must define the investigation scope and ensure that data privacy is integrated as part of a readiness plan. Furthermore, the third principle (PD3), *"where a need for the extraction and examination of all available data from a given digital device is established, this need must be both evidenced and justifiable with regards to the current*

investigation scenario" aligns with stage 3 of PPDF model in which investigator does not have access to non-case related information unless there is a justification supported with permission from a superior investigator with an oversight on the case.

Lastly, the author in [40] identified the significance of privacy policies in protecting users' private information and, hence, posited that such privacy-preserving policies should restrict an investigator from analyzing user's private data. The author defined ten privacy-preserving policies, denoted by PP1 - PP10 in Table 4 and summarized in Table 1, from both the user's and the investigator's perspectives, which covers the first three stages of the digital forensics process model. Five of the ten policies were based on the investigator's perspective, while two and three were based on the users', and both investigator and user perspectives, respectively. These policies are in alignment with the processes involved in each stage of our PPDF model. For example, of particular interest among the ten policies is the third policy (PP3) to *"limit the search for evidence to the goal of the investigation"* which aligns with the first stage of our conceptual PPDF model. In general, the conceptual PPDF model aligns with the key existing privacy-preserving principles and privacy levels outlined in the literature.

VI. CONCLUSION

In this study, we discuss the various cryptographic techniques that can be utilized for privacy protection in digital forensics, with an analysis of relevant studies that have utilized any of the techniques for privacy protection. We provide a summary of the findings for each study, highlight some drawbacks to the use of each cryptographic technique for privacy protection in digital forensics, and recommend potential solutions to address the highlighted drawbacks. Moreover, we proposed a conceptual model for a privacy-preserving digital forensics model that is based on cryptographic techniques and consider how and where each encryption technique may be used in the model. We present the mathematical representation and algorithm of the model and examine how the model may perform within the context of some identified analysis factors. We also examine some of the factors that may be considered at each stage of the model in specific situations and evaluate the model via a comparison with existing principles for preserving privacy in digital forensics investigations.

This study provides digital forensics investigators and researchers with a roadmap for addressing data privacy challenges in digital forensics, specifically by using cryptographic techniques. Our evaluation of the conceptual model shows that it performs well within the context of the analysis factors and it supports all the key privacy principles that have been suggested in the literature for privacy preservation in digital forensics. The model focuses on the overall digital forensics process but can be adapted to specific scenarios and sub-domains of digital forensics as necessary. In future work, we plan to implement and conduct an empirical study to determine its practicality and performance. Extension of the model to address some of the challenges in specific digital forensics subdomains will also be explored.

REFERENCES

[1] J. K. Malik and S. Choudhury, "A brief review on cyber crime-growth and evolution," *Pramana Res. J.*, vol. 9, no. 3, p. 242, 2019.
[2] F. Casino, T. K. Dasaklis, G. P. Spathoulas, M. Anagnostopoulos, A. Ghosal, I. Böröcz, A. Solanas, M. Conti, and C. Patsakis, "Research trends, challenges, and emerging topics in digital forensics: A review of reviews," *IEEE Access*, vol. 10, pp. 25464–25493, 2022.
[3] J. Sammons, "The basics of digital forensics: The primer for getting started in digital forensics," *J. Digit. Forensics, Secur. Law*, vol. 9, no. 1, p. 83, 2012.
[4] R. Kaur and A. Kaur, "Digital forensics," *Int. J. Comput. Appl.*, vol. 50, no. 5, pp. 5–9, 2012.
[5] J. R. Lyle, R. P. Ayers, and D. R. White, "Digital forensics at the national institute of standards and technology," U.S. Dept. Commerce, Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep., 2008.
[6] K. Kent, S. Chevalier, T. Grance, and H. Dang, "Guide to integrating forensic techniques into incident response," U.S. Dept. Commerce, Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. SP 800-86, 2006.
[7] R. Verma, J. Govindaraj, and G. Gupta, "Data privacy perceptions about digital forensic investigation in India," in *Proc. IFIP Int. Conf. Digit. Forensics*, 2016, pp. 25–45.
[8] L. Englbrecht and G. Pernul, "A privacy-aware digital forensics investigation in enterprises," in *Proc. 15th Int. Conf. Availability, Rel. Secur.*, Aug. 2020, pp. 1–10.
[9] W. Halboob, R. Mahmod, N. I. Udzir, and M. T. Abdullah, "Privacy levels for computer forensics: Toward a more efficient privacy-preserving investigation," *Proc. Comput. Sci.*, vol. 56, pp. 370–375, Jan. 2015.
[10] N. J. Croft and M. S. Olivier, "Sequenced release of privacy-accurate information in a forensic investigation," *Digit. Invest.*, vol. 7, nos. 1–2, pp. 95–101, Oct. 2010.
[11] G. Horsman, "Defining principles for preserving privacy in digital forensic examinations," *Forensic Sci. Int., Digit. Invest.*, vol. 40, Mar. 2022, Art. no. 301350.
[12] S. Raghavan, "Digital forensic research: Current state of the art," *CSI Trans. ICT*, vol. 1, no. 1, pp. 91–114, Mar. 2013.
[13] R. McKemmish, "What is forensic computing," *Trends Issues Crime Criminal Justice*, vol. 118, pp. 1–6, Jun. 1999.
[14] S. Saleem, O. Popov, and I. Bagilli, "Extended abstract digital forensics model with preservation and protection as umbrella principles," *Proc. Comput. Sci.*, vol. 35, pp. 812–821, Jan. 2014.
[15] E. Vincze, "Challenges in digital forensics," *Police Practice and Research*, vol. 17, no. 2, pp. 183–194, 2016.
[16] N. M. Karie and H. S. Venter, "Taxonomy of challenges for digital forensics," *J. Forensic Sci.*, vol. 60, no. 4, pp. 885–893, Jul. 2015.
[17] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, "A survey on the Internet of Things (IoT) forensics: Challenges, approaches, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1191–1221, 2nd Quart., 2020.
[18] S. E. Goodison, R. C. Davis, and B. A. Jackson, "Digital evidence and the U.S. criminal justice system. Identifying technology and other needs to more effectively acquire and utilize digital evidence, priority criminal justice needs initiative," Rand Corp., Santa Monica, CA, USA, Tech. Rep., 2015.
[19] E. Casey, "Clearly conveying digital forensic results," *Digit. Invest.*, vol. 24, pp. 1–3, 2018.
[20] A. Nieto, R. Rios, and J. Lopez, "Privacy-aware digital forensics," in *Security and Privacy for Big Data Cloud Computing and Applications*, 2019, pp. 157–195.
[21] J. I. James and Y. Jang, "Practical and legal challenges of cloud investigations," 2015, *arXiv:1502.01133*.
[22] E. Casey, "Digital evidence and computer crime," in *Forensic Science, Computers, and the Internet*. Cambridge, MA, USA: Academic Press, 2011.
[23] S. Khan, A. Gani, A. W. A. Wahab, M. Shiraz, and I. Ahmad, "Network forensics: Review, taxonomy, and open challenges," *J. Netw. Comput. Appl.*, vol. 66, pp. 214–235, May 2016.

[24] L. F. Sikos, "Packet analysis for network forensics: A comprehensive survey," *Forensic Sci. Int., Digit. Invest.*, vol. 32, Mar. 2020, Art. no. 200892.
[25] A. Al-Dhaqm, R. A. Ikuesan, V. R. Kebande, S. A. Razak, G. Grispos, K. R. Choo, B. A. S. Al-Rimy, and A. A. Alsewari, "Digital forensics subdomains: The state of the art and future directions," *IEEE Access*, vol. 9, pp. 152476–152502, 2021.
[26] H. F. Atlam, E. E.-D. Hemdan, A. Alenezi, M. O. Alassafi, and G. B. Wills, "Internet of Things forensics: A review," *Internet Things*, vol. 11, Sep. 2020, Art. no. 100220.
[27] T. Janarthanan, M. Bagheri, and S. Zargari, "IoT forensics: an overview of the current issues and challenges," in *Digital Forensic Investigation of Internet of Things (IoT) Devices*. Cham, Switzerland: Springer, 2021, pp. 223–254.
[28] I. Yaqoob, I. A. T. Hashem, A. Ahmed, S. M. A. Kazmi, and C. S. Hong, "Internet of Things forensics: Recent advances, taxonomy, requirements, and open challenges," *Future Gener. Comput. Syst.*, vol. 92, pp. 265–275, Mar. 2019.
[29] A. A. Boozer, A. John, and T. Mukherjee, "Internet of Things software and hardware architectures and their impacts on forensic investigations: Current approaches and challenges," *J. Digit. Forensics, Secur. Law*, vol. 16, no. 2, p. 4, Sep. 2021.
[30] A. Akinbi and T. Berry, "Forensic investigation of Google assistant," *Social Netw. Comput. Sci.*, vol. 1, no. 5, pp. 1–10, Sep. 2020.
[31] O. M. Adedayo and M. S. Olivier, "Ideal log setting for database forensics reconstruction," *Digit. Invest.*, vol. 12, pp. 27–40, Mar. 2015.
[32] R. Chopade and V. K. Pachghare, "Ten years of critical review on database forensics research," *Digit. Invest.*, vol. 29, pp. 180–197, Jun. 2019.
[33] A. Al-Dhaqm, S. A. Razak, S. H. Othman, A. Ali, F. A. Ghaleb, A. S. Rosman, and N. Marni, "Database forensic investigation process models: A review," *IEEE Access*, vol. 8, pp. 48477–48490, 2020.
[34] G. M. Jones and S. G. Winter, "An insight into digital forensics: History, frameworks, types, and tools," in *Cyber Security and Digital Forensics*, M. M. Ghonge, S. Pramanik, R. Mangrulkar, and D.-N. Le, Eds. Wiley, Jan. 2022, pp. 105–126.
[35] R. C. Maher, "Overview of audio forensics," in *Intelligent Multimedia Analysis for Security Applications*. Berlin, Germany: Springer, 2010, pp. 127–144.
[36] A. R. Javed, W. Ahmed, M. Alazab, Z. Jalil, K. Kifayat, and T. R. Gadekallu, "A comprehensive survey on computer forensics: State-of-the-art, tools, techniques, challenges, and future directions," *IEEE Access*, vol. 10, pp. 11065–11089, 2022.
[37] R. Kumars, M. Alazab, and W. Wang, "A survey of intelligent techniques for Android malware detection," in *Malware Analysis Using Artificial Intelligence and Deep Learning*. Cham, Switzerland: Springer, 2021, pp. 121–162.
[38] B. Carrier and E. Spafford, "An event-based digital forensic investigation framework," *Digit. Invest.*, 2004.
[39] J. C. Deprez, C. Ponsard, and N. Matskanis, "A goal-oriented requirements analysis for the collection, use and exchange of electronic evidence across EU countries," in *Proc. IEEE 24th Int. Requirements Eng. Conf. Workshops (REW)*, Sep. 2016, pp. 106–113.
[40] S. Srinivasan, "Security and privacy vs. computer forensics capabilities," *Inf. Syst. Control J.*, vol. 4, pp. 1–3, 2007.
[41] S. Kumar, A. Singh, A. Benslimane, P. Chithaluru, M. A. Albahar, R. S. Rathore, and R. M. Álvarez, "An optimized intelligent computational security model for interconnected blockchain-IoT system & cities," *Ad Hoc Netw.*, vol. 151, Dec. 2023, Art. no. 103299.
[42] Pallavi and V. Bharti, "A comprehensive review of cloud forensics and blockchain based solutions," in *Proc. 6th Int. Conf. Electron., Commun. Aerosp. Technol.*, Dec. 2022, pp. 749–754.
[43] G. Ragu and S. Ramamoorthy, "A blockchain-based cloud forensics architecture for privacy leakage prediction with cloud," *Healthcare Anal.*, vol. 4, Dec. 2023, Art. no. 100220.
[44] G. Kumar, R. Saha, C. Lal, and M. Conti, "Internet-of-Forensic (IoF): A blockchain based digital forensics framework for IoT applications," *Future Gener. Comput. Syst.*, vol. 120, pp. 13–25, Jul. 2021.
[45] M. Kim, Y. Shin, W. Jo, and T. Shon, "Digital forensic analysis of intelligent and smart IoT devices," *J. Supercomput.*, vol. 79, no. 1, pp. 973–997, Jan. 2023.
[46] J. Gu, B. Sun, X. Du, J. Wang, Y. Zhuang, and Z. Wang, "Consortium blockchain-based malware detection in mobile devices," *IEEE Access*, vol. 6, pp. 12118–12128, 2018.
[47] M. Kerr, F. Han, and R. V. Schyndel, "A blockchain implementation for the cataloguing of CCTV video evidence," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2018, pp. 1–6.
[48] S. H. Gopalan, S. A. Suba, C. Ashmithashree, A. Gayathri, and V. J. Andrews, "Digital forensics using blockchain," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2, pp. 182–184, 2019.
[49] T. K. Dasaklis, F. Casino, and C. Patsakis, "SoK: Blockchain solutions for forensics," in *Technology Development for Security Practitioners*. Cham, Switzerland: Springer, Jun. 2021, pp. 21–40.
[50] M. A. Khan, M. T. Quasim, N. S. Alghamdi, and M. Y. Khan, "A secure framework for authentication and encryption using improved ECC for IoT-based medical sensor data," *IEEE Access*, vol. 8, pp. 52018–52027, 2020.
[51] A. Viand, P. Jattke, and A. Hithnawi, "SoK: Fully homomorphic encryption compilers," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2021, pp. 1092–1108.
[52] T. B. Ogunseyi and T. Bo, "Fast decryption algorithm for Paillier homomorphic cryptosystem," in *Proc. IEEE Int. Conf. Power, Intell. Comput. Syst. (ICPICS)*, Jul. 2020, pp. 803–806.
[53] J. Park and E.-N. Huh, "ECLASS: Edge-cloud-log assuring-secrecy scheme for digital forensics," *Symmetry*, vol. 11, no. 10, p. 1192, Sep. 2019.
[54] K. Janjua, M. A. Shah, A. Almogren, H. A. Khattak, C. Maple, and I. U. Din, "Proactive forensics in IoT: privacy-aware log-preservation architecture in fog-enabled-cloud using holochain and containerization technologies," *Electronics*, vol. 9, no. 7, p. 1172, Jul. 2020.
[55] M. M. Rathinraj, M. J. R. Rajalakshmi, and M. M. Saranya, "Partial homomorphic encryption for secure log management using Tor network," *Int. J. Eng. Res. Technol.*, vol. 2, no. 12, pp. 3231–3235, 2013.
[56] S. Hou, T. Uehara, S. M. Yiu, L. C. K. Hui, and K. P. Chow, "Privacy preserving confidential forensic investigation for shared or remote servers," in *Proc. 7th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Oct. 2011, pp. 378–383.
[57] K. Huang and R. Tso, "A commutative encryption scheme based on ElGamal encryption," in *Proc. Int. Conf. Inf. Secur. Intell. Control*, Aug. 2012, pp. 156–159.
[58] S. Hou, T. Uehara, S. M. Yiu, L. C. K. Hui, and K. P. Chow, "Privacy preserving multiple keyword search for confidential investigation of remote forensics," in *Proc. 3rd Int. Conf. Multimedia Inf. Netw. Secur.*, Nov. 2011, pp. 595–599.
[59] S. Hou, R. Sasaki, T. Uehara, and S. M. Yiu, "Double encryption for data authenticity and integrity in privacy-preserving confidential forensic investigation," *J. Wireless Mobile Netw., Ubiquitous Comput., Dependable Appl.*, vol. 4, no. 2, pp. 104–113, 2013.
[60] T. B. Ogunseyi and C. Yang, "Survey and analysis of cryptographic techniques for privacy protection in recommender systems," in *Proc. Int. Conf. Cloud Comput. Secur.*, 2018, pp. 691–706.
[61] G. Weir, A. Amuth, M. Whittington, and B. Duncan, "Cloud accounting systems, the audit trail, forensics and the EU GDPR: How hard can it be?" in *Proc. Brit. Accounting Finance Assoc. (BAFA) Annu. Conf.*, Aug. 2017.
[62] A. Odebade, T. Welsh, S. Mthunzi, and E. Benkhelifa, "Mitigating anti-forensics in the cloud via resource-based privacy preserving activity attribution," in *Proc. 4th Int. Conf. Softw. Defined Syst. (SDS)*, May 2017, pp. 143–149.
[63] S. Hou, S. Yiu, T. Uehara, and R. Sasaki, "Application of secret sharing techniques in confidential forensic investigations," in *Proc. 2nd Int. Conf. Cyber Secur., Cyber Peace Fare Digit. Forensics*, 2013, pp. 69–76.
[64] S. Hou, S. M. Yiu, T. Uehara, and R. Sasaki, "A privacy-preserving approach for collecting evidence in forensic investigation," *Int. J. Cyber-Secur. Digit. Forensics*, vol. 2, no. 1, pp. 70–78, 2013.
[65] M. I. Mihailescu and S. L. Nita, "A searchable encryption scheme with biometric authentication and authorization for cloud environments," *Cryptography*, vol. 6, no. 1, p. 8, Feb. 2022.
[66] F. Armknecht and A. Dewald, "Privacy-preserving email forensics," *Digit. Invest.*, vol. 14, pp. S127–S136, Aug. 2015.
[67] K. Afifah and R. S. Perdana, "Development of search on encrypted data tools for privacy preserving in digital forensic," in *Proc. Int. Conf. Data Softw. Eng. (ICoDSE)*, Oct. 2016, pp. 1–6.
[68] D. Unal, A. Al-Ali, F. O. Catak, and M. Hammoudeh, "A secure and efficient Internet of Things cloud encryption scheme with forensics investigation compatibility based on identity-based encryption," *Future Gener. Comput. Syst.*, vol. 125, pp. 433–445, Dec. 2021.

[69] T. B. Ogunseyi and O. M. Adedayo, "Cryptographic techniques in digital forensics," in *Proc. Amer. Acad. Forensic Sci. (AAFS) Annu. Sci. Conf.*, 2023, p. 366.
[70] H. Chen, J. Schroeder, R. V. Hauck, L. Ridgeway, H. Atabakhsh, H. Gupta, C. Boarman, K. Rasmussen, and A. W. Clements, "COPLINK connect: Information and knowledge management for law enforcement," *Decis. Support Syst.*, vol. 34, no. 3, pp. 271–285, Feb. 2003.
[71] B. Martini and K.-K.-R. Choo, "An integrated conceptual digital forensic framework for cloud computing," *Digit. Invest.*, vol. 9, no. 2, pp. 71–80, Nov. 2012.
[72] H. I. Bulbul, H. G. Yavuzcan, and M. Ozel, "Digital forensics: An analytical crime scene procedure model (ACSPM)," *Forensic Sci. Int.*, vol. 233, nos. 1–3, pp. 244–256, Dec. 2013.
[73] F. Y. Law, K. P. Chow, M. Y. Kwan, and P. K. Lai, "Consistency issue on live systems forensics," in *Proc. Future Gener. Commun. Netw. (FGCN)*, vol. 2, 2007, pp. 136–140.
[74] Y. Zhang, Y. Li, and Y. Wang, "Conjunctive and disjunctive keyword search over encrypted mobile cloud data in public key system," *Mobile Inf. Syst.*, vol. 2018, pp. 1–11, Mar. 2018.
[75] S. Tahir, L. Steponkus, S. Ruj, M. Rajarajan, and A. Sajjad, "A parallelized disjunctive query based searchable encryption scheme for big data," *Future Gener. Comput. Syst.*, vol. 109, pp. 583–592, Aug. 2020.
[76] R. Altschaffel, R. Clausing, C. Kraetzer, T. Hoppe, S. Kiltz, and J. Dittmann, "Statistical pattern recognition based content analysis on encrypted network: Traffic for the TeamViewer application," in *Proc. 7th Int. Conf. IT Secur. Incident Manage. IT Forensics*, Mar. 2013, pp. 113–121.
[77] Y. Zhang, R. Liu, and D. Lin, "Improved key generation algorithm for Gentry's fully homomorphic encryption scheme," in *Information Security and Cryptology—ICISC 2017* (Lecture Notes in Computer Science), vol. 10779, H. Kim and D. C. Kim, Eds. Cham, Switzerland: Springer, 2017.
[78] Y. Zhu and X. Li, "Privacy-preserving k-means clustering with local synchronization in peer-to-peer networks," *Peer-Peer Netw. Appl.*, vol. 13, no. 6, pp. 2272–2284, Nov. 2020.
[79] N. L. Beebe and L. Liu, "Clustering digital forensic string search output," *Digit. Invest.*, vol. 11, no. 4, pp. 314–322, Dec. 2014.

TAIWO BLESSING OGUNSEYI received the Ph.D. degree from the Communication University of China, Beijing, China, in 2020. He is currently an Assistant Professor with Yibin University, Sichuan, China. His research interests include applied cryptography, privacy-enhancing technologies, data privacy, information security, cybersecurity, and machine learning.

OLUWASOLA MARY ADEDAYO (Member, IEEE) received the Ph.D. degree in computer science from the University of Pretoria, South Africa, in 2015. She is currently an Assistant Professor with The University of Winnipeg, Manitoba, Canada. Her research interests include databases, database forensics, digital forensics, cybersecurity, and privacy. She is a member of the Canadian Society of Forensic Science and an Associate Member of the American Academy of Forensic Sciences.