Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.024110
Article

Tech Science Press

Computer Forensics Framework for Efficient and Lawful Privacy-Preserved Investigation

Waleed Halboob[1,*] and Jalal Almuhtadi[1,2]

[1]Center of Excellence in Information Assurance, King Saud University, Riyadh, 11653, Saudi Arabia
[2]College of Computer and Information Sciences, King Saud University, Riyadh, 11451, Saudi Arabia
*Corresponding Author: Waleed Halboob. Email: wmohammed.c@ksu.edu.sa
Received: 05 October 2021; Accepted: 27 November 2021

Abstract: Privacy preservation (PP) in Digital forensics (DF) is a conflicted and non-trivial issue. Existing solutions use the searchable encryption concept and, as a result, are not efficient and support only a keyword search. Moreover, the collected forensic data cannot be analyzed using existing well-known digital tools. This research paper first investigates the lawful requirements for PP in DF based on the organization for economic co-operation and development OECB) privacy guidelines. To have an efficient investigation process and meet the increased volume of data, the presented framework is designed based on the selective imaging concept and advanced encryption standard (AES). The proposed framework has two main modules, namely Selective Imaging Module (SIM) and Selective Analysis Module (SAM). The SIM and SAM modules are implemented based on advanced forensic format 4 (AFF4) and SleuthKit open source forensics frameworks, respectively, and, accordingly, the proposed framework is evaluated in a forensically sound manner. The evaluation result is compared with other relevant works and, as a result, the proposed solution provides a privacy-preserving, efficient forensic imaging and analysis process while having also sufficient methods. Moreover, the AFF4 forensic image, produced by the SIM module, can be analyzed not only by SAM, but also by other well-known analysis tools available on the market.

Keywords: Digital forensics; digital evidence; AFF4; privacy preservation; selective imaging

1 Introduction

Digital forensics (DF) is a cybersecurity discipline that deals with extraction digital evidence from digital media. It has five main steps, namely digital evidence identification, collection, preservation, analysis, and presentation. Nowadays, it has several branches, such as computer, network, mobile, software, database, IoT, and cloud forensics [1–4]. For instance, computer forensics deal with gathering digital evidence from computer storage devices (e.g., hard disks, flash memories, DVDs, etc.,). Network forensics consider the collection and analysis of digital evidence from network traffic and logs (e.g., routers, switches, etc.). Mobile forensics is used when investigating mobile systems and apps, and so on.

Privacy preservation (PP) in DF is an essential issue for many reasons. For example, to enforce privacy protection acts while dealing with forensic data. Digital forensics examiners have to consider the relevant laws as the investigation result will end up in court [5]. Additionally, collected forensic data may contain files irrelevant to the crime and/or relevant to other unrelated users. As a result, finding a balance between the DF and PP is highly required [6–11]. It is clear that the process of digital forensics interferes with the privacy rights [12]. It imposes a privacy threat to both victims and suspects [13]. Such balance requires addressing several legal and technical requirements, and they are not clear yet. Addressing the privacy issue in digital forensics still an open issue and more research is required to address it [13–17].

Privacy risks in digital forensics are different from those in different computing areas [13]. In computing, private data includes the data owner's (or user's) identity, information, and activities, and these should only be disclosed to authorized parties [6]. On the other hand, in DF, the data owner's or user's identity must be revealed to the investigator, otherwise it is impossible to link a suspect to any crime under investigation. Therefore, in relation to DF, data privacy only needs to consider the confidentiality of the user's information and activities while excluding the user's identity. Last but not least, the privacy of the data owner's information and activities cannot be achieved by making them inaccessible to the examiner. The investigator should be able to check all data and decide which are relevant to be collected. As a result, the target of PP in DF is only finding a balance between the privacy and investigation needs in order to collect only relevant data even if it is private. However, finding the required balance is also difficult so there is a need to offer a balance based on a universal and well-known privacy guideline.

Another issue relevant to the privacy is the data volume. The forensic data is increased over the time which leads to increasing the investigation cost in terms of resources and time [5,13]. By the way, PP in DF mainly uses cryptography techniques which make the investigation cost, especially in terms of time, worse. Since applying the hashing, and encryption/decryption algorithms while imaging and analyzing the forensic data will require more time than before. Therefore, it is very important to consider the data volume issue while preserving the privacy in DF.

The motivation behind this research is to preserve the privacy while reducing the volume of the forensics data to minimize the investigation cost. This is achieved by first selectively imaging only relevant data and, secondly, encrypting only private and relevant data using a symmetric cryptography, instead of encryption the whole data using asymmetric cryptography.

This research paper first investigates the lawful requirements for PP in DF based on a global privacy guideline. Then, it proposes a privacy-preserving computer forensics framework that has two main modules, which are: selective imaging module (SIM); and selective analysis module (SAM). Based on the selective imaging concept, the SIM first collects or images relevant data only while ignoring irrelevant data. Second, the relevant digital evidences are classified into private and non-private. The private data are imaged in an encrypted format, using advanced encryption standard (AES), while normally (in plain text format) imaging the non-private. This leads not only to preserving privacy but reducing the investigation cost (required time and storage) as only the relevant data are imaged and the encryption is used only with the private relevant data. The SAM is used for analyzing the collected data using several analysis methods, basically keyword-based and attributes-based searches. Additionally, the collected forensic image–by the SIM module–can be also analyzed using the existing popular digital forensics tools such as X-Ways [18] and FTK AccessData [19].

The proposed SIM and SAM forensics modulus are implemented using Java language and with the assistance of AFF4 [20] and SleuthKit [21] open source digital forensics frameworks. This is to ensure that the proposed framework is developed and evaluated in a forensically sound manner. To be more specific, the SIM and SAM modules are executed as part of the SleuthKit framework to be tested in a

forensically sound environment. Then, the SIM and SAM modules are refined based on the implementation stages. Finally, both modules are evaluated and compared with other related works.

This research paper covers the computer forensics branch only. On the other hand, it considers the privacy issue while investigating computing storage devices (e.g., hard disks, CDs/DVDs, flash memories, SD memories, etc.). For that reason, other DF branches (e.g., network forensics, mobile forensics, etc.) are not covered here. However, concepts, methodologies, and designs provided by this research paper can be adopted in other digital forensics branches. Also, the proposed framework considers three main technical DF steps, namely, collection, preservation, and analysis. These steps are totally executed using DF tools. However, the non-technical DF steps (identification and presentation) are not covered by this research paper.

The outline of this paper starts in Section 2 with presenting the selective imaging concept followed in Section 3 by introducing the related works. The suggested lawful technical privacy requirements in DF are discussed in Section 4. The architecture of the proposed framework is presented in Section 5. Sections 6 and 7 introduce the proposed SIM and SAM modules, respectively. Section 8 discusses the implementation of the proposed framework, and then, in Section 9, research results and discussion are presented. Section 10 concludes the presented work and highlights potential future works.

2 Selective Imaging Concept: Concept and State of the Art

Privacy in DF and DF imaging are two different topics but related. Making a bit-by-bit forensic image for the whole storage device has already been proven to be an undesirable solution, especially when investigating large disk volumes, shared server, or distributed storage [5,13,17,22]. In DF labs, many cases are pending due to the huge volume of relevant data [23]. The typical amounts of storage and data involved is increased over time, which increases the required investigation cost. The cost here includes the required imaging time and resources. To be more specific, the resources are mainly the required storage for storing the collected data [24]. For this reason, the selective imaging concept is introduced to image only relevant forensic data and, as a result, reduce the imaging cost.

Nevertheless, imaging only the relevant data using the selective imaging concept is the key point for addressing the privacy issue in DF as mentioned earlier in this research paper. The selective imaging concept was first proposed by Turner [25], and it is the most acceptable and used solution until now. The idea is to use a pre-analysis step to identify the data that are relevant to the case. At that time, only the selected data are imaged, instead of imaging everything in the targeted storage. Accordingly, when using the selective imaging concept, the forensics analysis task is executed in two different phases: a pre-analysis phase and normal analysis phase. The latter is performed at the DF Lab and after selectively imaging the relevant data. According to Turner [26], the relevant data can be selected through three different approaches, namely: manual, semi-automatic, and fully automatic selections. Using the first approach, the relevant data are selected manually. With the second approach, the keyword-based or/and attribute-based searches are used for selecting the relevant data. Finally, the last approach requires some intelligent algorithms for selecting the relevant forensic data. This should be based on pre-identified parameters entered by the investigator. In anyway, the fully automatic selection still has several shortcomings, and further research efforts are required to develop a more practical, fully automatic method [22,27].

There are several selective imaging methods in the literature such as block-based compression [27], hash-based disk imaging [28], risk sensitive digital evidence collection [24], digital evidence bags [25–26,29–31], sifting collector [32], and filtering techniques [23].

The advanced forensic format 4 (AFF4), proposed by Garfinkel et al. [33] has become a standard selective forensic image format. The AFF4 is currently the most suitable forensic format for storing selectively imaged data because it supports multiple objects, encryption, and metadata [22,29]. Most, if not all, DF tools support the selective imaging concept based on the AFF. For instance, EnCase imager [34] supports the imaging of a single file or folder. FTK imager [35] supports also imaging some selective files and folders.

However, the AFF4 format is also further studied by several researchers. In [22,25], a selective imaging model is proposed based on the AFF4 forensic format. The authors also implemented a prototype and showed how it can be integrated with the Digital Forensics Framework (DFF). The efficiency of the proposed prototype is also evaluated. In addition, in [36], the effects of ordering the selected forensics files, based on their offsets or physical addresses, on the imaging efficiency is evaluated in both magnetic hard disks and electronic memories (such as flash memories, SD cards, and SSD hard disks). They suggest ordering the forensics files, especially when dealing with the magnetic hard disks.

It can be concluded that the selective imaging concept is proposed for reducing the forensic imaging cost (in terms of time and resources). However, it is very important to mention here that even though the investigation cost and PP are two different issues, but related, as both of them are addressed basically using the selective imaging concept.

3 Related Works

As mentioned earlier, there is a need for solving the conflict between DF and PP since "privacy is a right but not a cover for crime," as stated by Caloyannides [37]. This conflict has been studied widely in the literature. At the beginning, Burmester et al. [6] studied the conflict between: anonymity and accountability; privacy and authentication; PP and DF; and free speech and liability/copyright. The authors recommended applying auditing, policies, and cryptography for solving this conflict.

Then, in [9], several digital forensics issues are discussed, including the privacy issue. The study advised following existing privacy acts so the investigation process can preserve privacy in a lawful manner. Saboohi [38] considered the issue of relevant data selection in which the investigator decides which data are relevant. Since selecting the relevant data using the current DF tools requires that the investigator see all of the entire data first to check which data are relevant, the author realizes that selecting the relevant data requires changing the existing laws, labs, and procedures. The state of privacy and DF in the United States is studied by Adams [39], according to the Fourth Amendment to the United States Constitution. This study suggested obtaining a search warrant/authorization letter and using an audit trail for auditing the forensics investigation process.

Fahdi et al. [40] studied several technical, legal, and resources challenges in DF. In terms of privacy, this study found that the DF tools did not consider the legal requirements for PP during investigation. Nieto et al. [41] discussed the privacy issue in DF in depth. Preserving the privacy in different DF branches is discussed. The study concluded that private and public sectors should understand the privacy issue in DF and help in resolving this issue.

Recently, the privacy issue in DF stills widely discussed in the literature both from legal and technical sides. In [5,12–14], the issue is reviewed and, as a result, more efforts are still required to address these issue.

However, existing solutions can be classified into three different directions, namely: policy-based approaches; privacy-aware digital forensics (DF) guidelines; and cryptographic approaches. These directions are reviewed in the following sub-sections.

3.1 Policy-based Approached

In fact, this approach is used to identify the requirements of the private data collection, management, use, and disclosing. Srinivasan [42,43] proposed four policies for the digital evidence collection step only. Therefore, the others digital evidence steps (e.g., analysis, preservation, etc.) must be covered too. In [44], four privacy levels for computer forensics are suggested. These levels are: 1) non-private and non-relevant; 2) private and non-relevant; 3) non-private and relevant; and 4) private and relevant. With the first and second levels, the data are not relevant and should not be investigated. The relevant data are investigated in two different ways. The relevant and non-private data are investigated normally but the relevant and private data are investigated in a privacy-aware manner. In [45], ten privacy policies for computer forensics are presented. These policies cover all computer investigation steps. However, it is clear that existing policy-based approaches are few and cover only computer forensics and there is a need to extend these works to other digital forensics branches such as network forensics.

In [13], a privacy impact assessment (PIA) methodology for large scale digital forensics is presented. The authors studied the relevant assessment methodologies and suggested three of the most relevant methodologies. They come out also with a privacy measure to evaluate the compliance level.

3.2 Privacy-aware Digital Forensics Framework

Digital forensics (DF) frameworks are guidelines that describe the investigation steps and their sub-steps. Each DF framework can differ from other frameworks on the number of steps as well as sub-steps along with adding new concepts or requirements (like encryption, auditing, etc.) in any step or sub-step.

In the literature, there are three privacy-aware DF frameworks. First, Gupta [46] proposed a framework called a Privacy Preserving Efficient Digital Forensic Framework Investigation (PPEDFI). The author relied only on the selective imaging concept discussed here early. The relevance is determined through some parameters entered by the investigator. These parameters can be file name, extension, size, etc. In fact, this work tried to address the issues of privacy and efficiency in DF in a general or high-level framework. Saleem et al. [47] proposed an abstract model (or guideline) that preserves the privacy of data owners based on another abstract model proposed by Reith et al. [48]. This model just concerned the general or high-level steps of DE protection as well as PP.

In 2017, Nieto et al. [49] proposed a methodology for DF in IoT forensics based on the ISO/IEC 29100:2011 standard [50]. The ISO/IEC 29100: 2011 standard, which provides the best practice in privacy preservation, is an optional standard for the private and public sectors. It is also different from the privacy acts, which are more required to be enforced in DF. In 2020, Ferguson et al. [51] proposed a framework called PRECEPT, which provides a guideline for ethically dealing with digital evidence during the investigation process. Like the works proposed in [46], this framework presents general or high-level guidelines. Finally, Englbrecht et al. [52] presented an entropy-based approach for identifying the private data in enterprise forensics so such data are excluded during the forensic imaging step. Such methods can help organizations ignore private data while internally responding to a cybersecurity incident.

3.3 Cryptographic Approaches

Regarding cryptographic approaches, Law et al. [53] proposed a solution that makes a bit-by-bit image from the targeted storage, builds and encrypts an index for each file, and finally prepares and encrypts keywords to search for relevant data in the index files during the analysis stage. However, this solution supports only text-based data files (e.g., txt, doc, pdf, etc.) which can be indexed using their contents (or words) while the non-text-based files (such as photo and video files) cannot be indexed. Also, the investigation cost, in terms of time, is very high.

In Hou et al. [54], two searchable encryption schemes are proposed. With the first scheme, the suspect encrypts his data, and then the examiner encrypts a single search keyword and passes it to the suspect. The suspect searches for the relevant data and submit them to the examiner. The main issue with this scheme is that the suspect can hide any data and it cannot be trusted. Therefore, the authors suggest another scheme to address this issue using a Third Trusted Party (TTP). The TTP is used to search for relevant data but even trusting a third party is not a good idea in DF. In Hou et al. [55,56], the first scheme proposed in [53] is improved first to be able to work with search keywords, instead of one, and second to ensure preservation of the collected forensic data while supporting different investigators.

Another cryptographic-based approach is proposed for cloud computing by Jayaraman et al. [17]. The presented work effectively provides a protection against several kinds of attacks (such as spoofing, tampering, denial of service, and privilege escalation) and digital forensic readiness which makes the proposed technique ready for any investigation case in a privacy-aware manner.

These works [53–56] have several drawbacks, including the following: i) the investigation cost, in terms of time, is very high as all data must be indexed and encrypted; ii) they cannot support non-text data such as photos, videos, etc.; iii) the forensic analysis uses only keyword search and such analysis is not sufficient, especially for non-text data files. More analysis method, such as attributes-based search, must be also supported.; and iv) existing well-known digital forensic analysis tools (e.g., EnCase, FTK AccessData, etc.) cannot be used for analyzing the collected data.

4 Lawful Technical Privacy-Preserving Requirements in Digital Forensics

In this Section, the lawful requirements for privacy preservation (PP) in digital forensics (DF) is investigated. In fact, any DF solution or tool must be lawful, as proven by a court of law [56], and by taking into account existing acts. However, different countries have different privacy acts. For that reason, the most important guidelines are found in the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [57]. Based on the OECD guideline, this research paper investigates the lawful requirements for PP in DF. This is because the most countries' privacy acts, including those of the United States, Europe, and the Asia-Pacific Economic Cooperation (APEC) countries [58], build their privacy acts to be consistent with the core values of the OECD guideline.

The OECD guideline has eight privacy protection principles that must be followed while collecting, using, and disclosing data. These eight principles can be used for evaluating the lawfulness of any DF tools in term of PP and whenever any OECD-based or OECD compatible privacy act is used in both the private and public sectors. Except in some internal cases in which the organization's policies are followed during an investigation unless the case is reported to the police or court of law. Another important issue that must be addressed by law enforcement is: what actions must be taken if an investigator violates the privacy act? Actually, such actions cannot be discussed here as it is outside the scope of this research paper. Even the OECD privacy principles didn't discuss this issue and it seems that actions are taken by the investigation authority (e.g., in the court of law) against the investigator and according to laws.

Tab. 1 concludes how the eight OCED requirements can be applied in digital forensics. The security controls required for enforcing each privacy requirement are identified. Finally, PP in DF requires applying eight security controls or requirements: selective imaging; non-repudiation; encryption-based access control; integrity; privacy policies; identical copy; authenticity; and auditing.

Table 1: Privacy preservation requirements in digital forensics

#	OCED requirements	Meaning	Requirement	Requirement type
1	Collection limitation principle	Data collection must be limited and performed in a lawful and fair manner as well as with the knowledge of the data owner.	Court order (or search warrant).	Legal requirement
2	Data quality principle	The investigator should collect only relevant data.	Selective imaging concept	Security requirement
3	Purpose specification principle	The investigator should specify the purpose of collecting the data before starting the collection process.	The investigation's goal and scope must be specified in the court order.	Legal requirement
4	Use limitation principle	The collected data should not be used with anything not relevant to the data collection purpose.	Non-repudiation, and encryption-based access control	Security requirement
5	Security safeguards principle	The collected data must be protected.	Integrity, and encryption-based access control; and authenticity	Security requirement
6	Openness principle	A policy should exist and be used to specify how the data collector and processor deal with the collected data.	Privacy policies	Security requirement
7	Individual participation principle	The data owner must know which data related to him or her are collected, as well as be able, at any time, to challenge the data collector, in relation to whether the data collected and used are really related to him or her.	A second forensic image is provided to the suspect as suggested in (Srinivasan, 2006 and 2007), but this leads to increasing the investigation cost and revealing the investigation confidentiality.	Security requirement
8	Accountability principle	The data collector is accountable for complying with all the previews requirements.	Auditing the investigation process	Security requirement

5 Architecture of the Proposed Framework

The architecture of the presented framework consists of two modules, as shown in Fig. 1, which are SIM and SAM. Mainly, the SIM module collects all relevant data into a partial or selective forensic image while ignoring the irrelevant data. The SAM module is used to analyze the selective forensic image. Before the imaging process is started, a pre-analysis process is executed using other suitable forensic recovery tools to first scan the suspect's data for any deleted files and then select the relevant and/or private forensic data files. In fact, this research paper assumes that scanning the targeted storage and selecting the relevant and private data are outside its scope. However, the existing methods for this purpose are studied. Some recovery tools such as CnWRecovery [59] can be used. These tools can be used for scanning the suspect and selecting the relevant data.

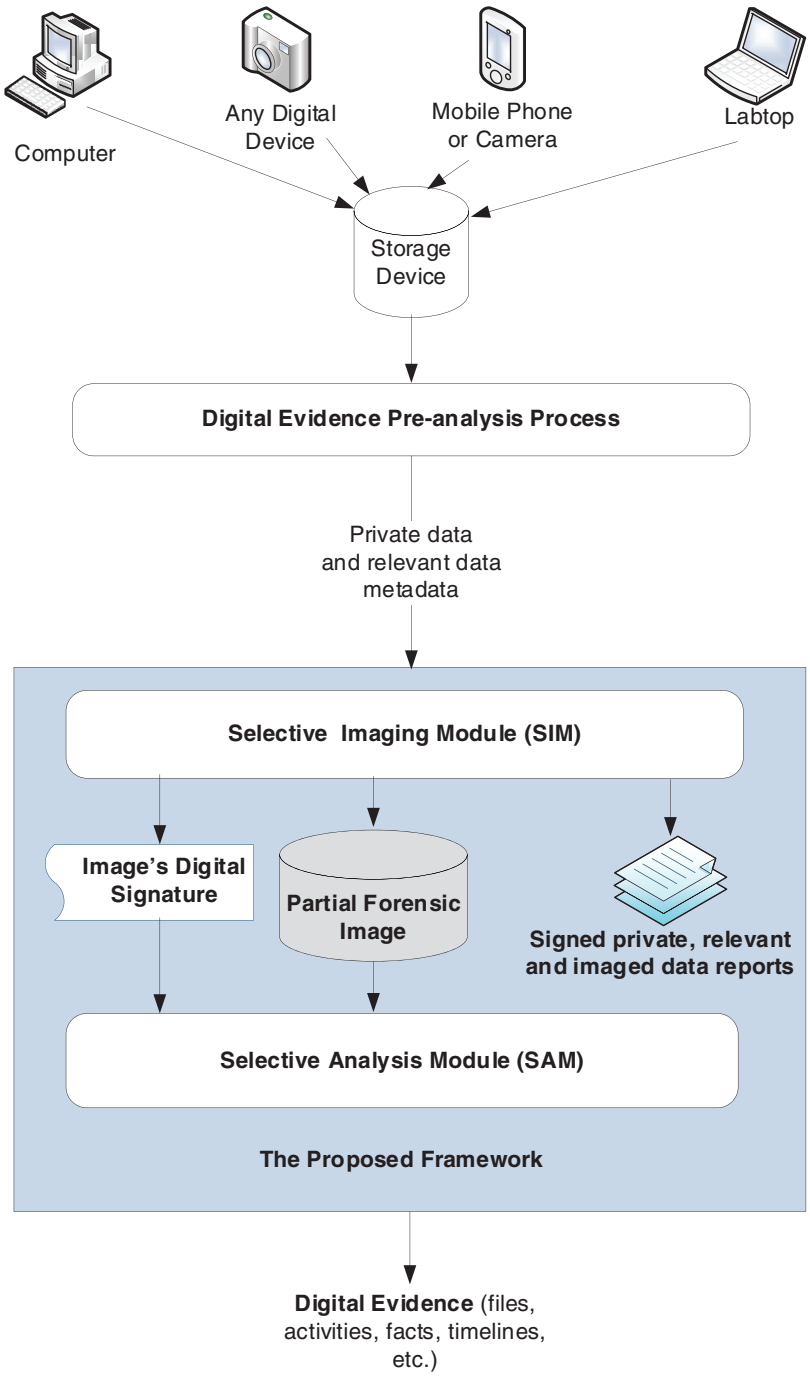

Figure 1: General architecture of the proposed framework

The CnWRecovery [59] tool is used and verified in details by this research paper. Its search results can be reported in common-separated values (csv) files, which then is used by the SIM to image all reported files inside the .csv files. In additions, the forensic data are selected and saved into several metadata files. The selected private data files are reported into one or more csv files, which are called private data metadata reports (PDMRs). The selected relevant data files are also reported into one or more csv files, which are

called relevant data metadata reports (RDMRs). Each metadata report contains several information about each selected data file (e.g., name, path, size, type, create date, modify date, etc.). For the public/private keys, they are used for ensuring the forensic data integrity and authenticity along with providing the non-repudiation security feature. Each investigator has a key pair (public and private key). The investigation authority (e.g., court of law) or the digital forensics lab can serve as a certificate authority (CA). The CA needs to establish a public key infrastructure (PKI) to generate a public/private key pair for each investigator. The use of these keys will be discussed later.

6 Selective Imaging Module (SIM)

Fundamentally, this module is used for imaging all relevant data (private or not) while ignoring irrelevant data using two main processes, namely imaging and verification processes. The following sub-sections present these two processes in more details.

6.1 The Imaging Process

First of all, during this process, the following rules are applied:

- The non-relevant data are ignored, private or not.
- The relevant non-private data are directly imaged without encryption. But the other security requirements or controls presented in Section 4 are enforced.
- The relevant private data are imaged in a privacy-preserving manner by enforcing all security requirements discussed in Section 4.

To improve the imaging efficiency, the imaging process starts also with running the ordering algorithm proposed in [36]. This algorithm mainly orders the selected relevant forensic data files, according to their offsets (or physical addresses) in the targeted storage. The goal is to reduce the imaging cost in terms of time. The time consumed by the storage to read data files is less if the data files are ordered according to their offsets, especially if the storage is a magnetic hard disk due to reducing the time required by the heads to move from one cylinder to another while reading the data files. However, the ordering algorithm receive both PDMRs and RDMRs reports, which discussed early, and come out with a re-port called an ordered relevant data metadata report (ORDMR).

Consequently, as shown in **Algorithm 1**, the core inputs for the imaging process is the ORDMR, targeted storage, and finally the investigator's public and private keys. The main outputs of this process are selective or partial forensic images, and its hash value along with the investigator's digital signature. Up to now, the proposed framework ensures the selective imaging, integrity, non-repudiation, and authenticity. For the forensic image format used by this research paper, our work needs a forensic format that supports the selective imaging concept. Several forensic image formats are now widely used, including the raw image format (RAW/DD), SGZIP, EnCase forensic image file (EO1), advanced forensic format 3 (AFF3), and advanced forensic format 4 (AFF4). The AFF4 is used by this research paper because it supports enforcing the eight PP security requirements or controls in DF, which are presented in Section 4.

The first imaging processing step, as shown in Line 10, is reading forensic files from the storage device. If the file is non-private, it is directly written into the AFF4 image as shown in Line 15 and, in this case, the file's hash value is, by default, generated (using AFF4 API) and saved into the AFF4 image, along with the file content at the same segment. But, in this research paper, the AFF4 API is modified by replacing MD5 with SHA-1 to have a more secure hashing method.

In a case where the current file is private, the forensic file hash value is calculated using SHA-1 and saved temporarily in a byte array called CurrentHash (Line 12). Then, the file's content is encrypted

using the AES system (Line 13) and the encrypted file and its hash value (or CurrentHash) are stored inside the AFF4 image (as in Line 14). Finally, and after imaging all selected files, a hash value and digital signature are generated from the partial AFF4 forensic image as presented in Line 19 and Line 20, respectively. The SHA-1 and investigator's private key are used for hashing the entire AFF4 image and signing the hash value, respectively.

Algorithm 1: Execution of imaging process

Input: Ordered Relevant Data Metadata Report (ORDMR) and Storage Device

Output: Partial AFF4 Image, AFF4HASH, AFF4DIGSIGNATURE, and Encrypted s

```
1   LET i = 1
2   LET n = number of all data items inside ORDMR
3   LET s = a new AES secret key
4   LET CurrentHash, AFF4Hash = new byte arrays
5   Ipri = investigator's private key
6   LET Ipub = investigator's public key
7   LET AFF4DigSignauter = a new byte array
8   LET Partial AFF4 Image = a new AFF4 image
9   WHILE i is less than or equal to n THEN
10      READ file[i] from the storage
11      IF file[i] is private THEN
12          CurrentHash = Hash (file[i])
13          DO encrypting file[i] with s
14          WRITE CurrentHash, and file[i] into the Partial AFF4 Image
15      Else WRITE file[i] into the Partial AFF4 Image
16      End If
17      DO increasing i by 1
18  END WHILE
19  AFF4Hash = Hash (Partial AFF4 Image)
20  AFF4DigSignauter = Encrypt (AFF4Hash) with Ipri
21  DO encrypting s with Ipub
22  DO save the Partial AFF4 Image, AFF4Hash, AFF4DigSignauture, and Encrypted s
23  End
```

6.2 The Verification Process

Unlike the traditional verification procedure, in which the targeted storage is imaged, bit-by-bit, and two hash values are generated from the targeted storage and forensic image. Then, the two hash values are compared, and if they match, the storage is correctly imaged. However, with the selective imaging concept, there is a need to generate a hash value for each imaged file and save the hash value inside the AFF4 selective image along with the file's content. Thus, the integrity verification requires matching two

hash values for each file (hash value inside the targeted storage and the other inside AFF4 image). The image integrity is ensured only when the integrity of all the relevant files are ensured, in other words, correctly imaged. By the way, it should be noted that this research paper uses the SHA-1 hash function, instead of MD4, which is currently used by the AFF4 open source tool used by this research paper, during implementing the proposed framework. The MD5 is replaced with SHA-1 to ignore the collision found in MD5.

Algorithm 2 illustrates the integrity verification process, which simply has to verify two things: the files' hash values and image's integrity and authenticity. To verify the files' hash values, each file's hash value is read from the targeted storage (Line 5), its SHA-1 hash value is generated, and the new hash value is compared with the file's old hash value stored inside the AFF4 image. If the hash values of all the files are correct, the next step is verifying the integrity and authenticity of the entire AFF4 image. Otherwise, the imaging process must be re-executed.

Algorithm 2: Execution of integrity verification process

Input: ORDMR, Storage Device, Partial AFF4 Image, and AFF4DigSignature

Output: Result (yes or no)

```
1     LET i = 1
2     LET n = number of data items inside the ORDMR
3     LET Ipri = investigator's private key
4     WHILE i is less than or equal to n THEN
5         FirstHash = file[i]'s hash in the Storage Device
6         SecondHash = file[i]'s hash in the Partial AFF4 Image
7         If FirstHash = SecondHash THEN DO increasing i by 1
8         ELSE PRINT "file[i] hash value is incorrect"
9             LET i=n+1
10        End If
11    END WHILE
12    2ndAFF4Hash = Hash (Partial AFF4 Image)
13    2ndAFF4DigSignauter = Sign (AFF4Hash) with Ipri
14    IF AFF4DigSignature = 2ndAFF4DigSignature THEN
15        PRINT "Files' hash values and image signature are matched"
16    ELSE PRINT "Image's signature is incorrect"
17    End IF
18    END
```

To verify the AFF4 image's integrity, the image is hashed as presented in Line 12, and the new hash value is compared with the old hash value. If they match, the image is integrated. The authenticity is ensured by digitally signing the new hash value with the investigator's private key and then comparing it with the signature of the old AFF4 image. This leads to ensuring the non-repudiation security feature as well. Finally, a very important advantage provided by the proposed framework is that the relevant forensic data can be imaged into more than one AFF4 forensic image, for instance, based on their types. This helps in distributing the workload of the forensic analysis (executed at the digital forensics lab) to

more than one investigator and, as a result, decreasing the analysis time with any urgent case or crime. Therefore, this research paper applied the selective imaging concept for image only relevant data to reduce the investigation cost (time and resources). The suspect's storages can be imaged into different AFF4 forensic images and also each storage can be imaged into several AFF4 forensic images.

7 Selective Analysis Module (SAM)

The SAM consists of four subcomponents, as illustrated in Fig. 2, which are analysis interface, search and decryption engine, analysis viewer, and auditing trail. The SAM is designed in such a way that the collected AFF4 images can be more effectively analyzed using existing forensic analysis tools. The AFF4 image can be analyzed using two other well-known digital forensic analysis tools, namely X-Ways and FTK AccessData. These tools can directly analyze any non-encrypted forensic file while the encrypted forensic files still can be analyzed based on their metadata.

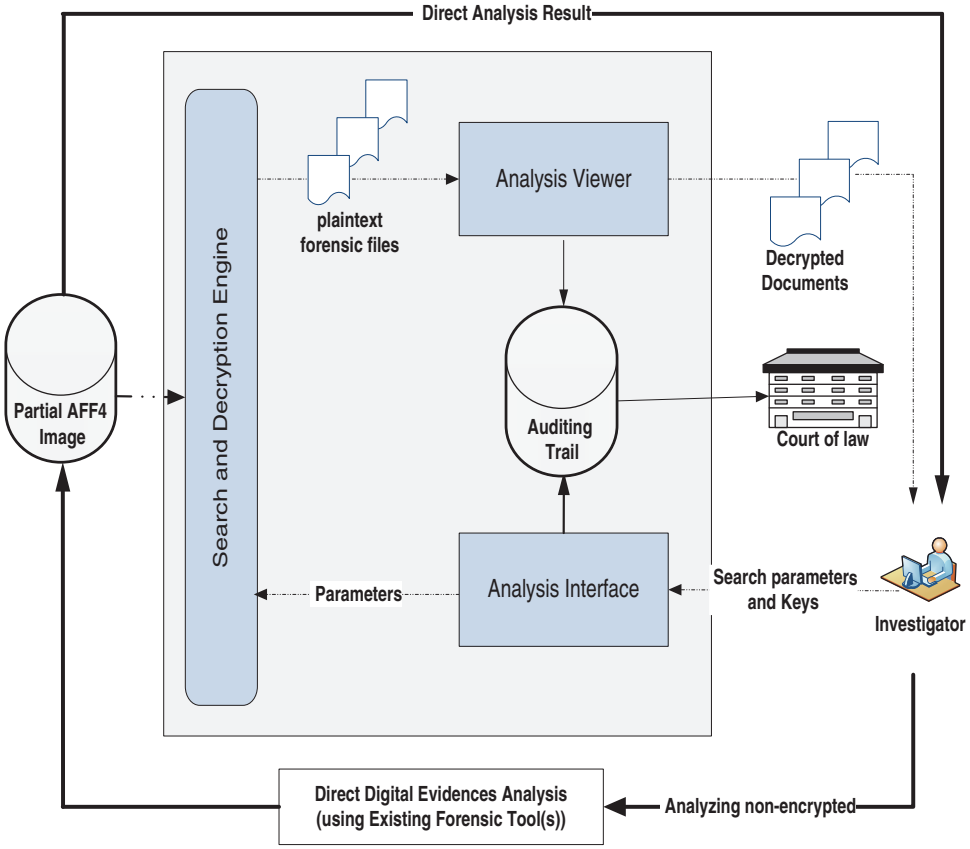

Figure 2: General architecture of the selective analysis module (SAM)

The analysis interface is used for: i) to authenticate the investigator (or analyzer) with his private key; ii) verify the AFF4 image's integrity; iii) receive the investigator's search parameters and pass them to the search and decryption engine; and finally iv) receive the secret key, decrypt it with the investigator's private key, and pass it to the search and decryption engine.

8 Implementation

To ensure that the proposed framework is workable, both the SIM and SAM are implemented and evaluated. However, the designed framework is refined based on its implementation and evaluation results. The NetBeans Java programming IDE (Java SE 8) [60] is used with the assistance of some additional Java application programming inter-faces (APIs) and digital forensics open source tools, as presented below.

The Java Cryptography Extension (JCE) API [61] is used to provide several security mechanisms and features such generating RSA public/private keys, generating a secret AES encryption key, encrypting the private data with the AES cryptosystem, hashing the collected data files and AFF4 image, digitally signing the collected AFF4 image's hash value, authenticating the investigator, and protecting (encrypting and/or signing) the auditing reports and trail. The size used for the generated public and private keys is 1024 bits. The JCE tool is also used to generate an AES secret key, and initializing and using SHA-1. To process the csv files, the Java CSV library [62] is used, which is a free Java package. The csv files include PDMRs, RDMRs, and ORDMR reports. The Advance Forensic File Format (AFF4) Java package [20] is used for creating the forensic image. This open source package is modified in this research paper, by replacing MD5 with SHA-1 and adding other metadata. The forensic data files can be imaged using the following four imaging cases: 1) Encrypted AFF4 imaging without compression; 2) Encrypted AFF4 imaging with compression; 3) Normal AFF4 imaging without compression; and 4) Normal AFF4 imaging with compression. The implementation code of the normal AFF4 imaging covers both compression and without compression cases. The only difference between these cases is the change in the zipping level (which is a number) inside the AFF4 open source [20]. The same thing is true when the encrypted AFF4 imaging type is used.

Regarding implementing the SAM in terms of accessing the AFF4 image, the open source SleuthKit (TSK) digital forensic framework [21] is used for reading the forensic files from the targeted storage. This helps to ensure that these files are read in a forensically sound manner. This means in read-only mode, without altering the contents of the targeted storage. This framework is comprehensive, but only the functions required for accessing the targeted storage is used.

9 Results and Discussion

The proposed framework is evaluated using two main criteria, namely imaging efficiency and forensic search sufficiency. The result, at the end, is compared with other related works [53–56], since they are the only work, we found, related to cryptographic-based, privacy-aware digital forensics solutions. To evaluate the efficiency, 100,000 text files are made and used and each file includes 600 words, as used in [53]. To make a tag for validating the search results during digital evidence analysis, 25 percent of these files contain a "forensics" word, 25 percent contain a "privacy" word, and 25 percent contain a "security" word. This experiment is executed in Windows 10 using a personal laptop with 8 GB of RAM and Intel Core i7-6500U (4 CPUs @ 2.5-GHz) processor.

9.1 The Imaging Time

Here, the required time for data imaging of the proposed framework is measured for four different cases: i) encrypted AFF4 imaging without compression; ii) encrypted AFF4 imaging with compression; iii) normal AFF4 imaging without compression; and iv) nor-mal AFF4 imaging with compression. Fig. 3 show the required imaging time results for the proposed framework with these four imaging cases.

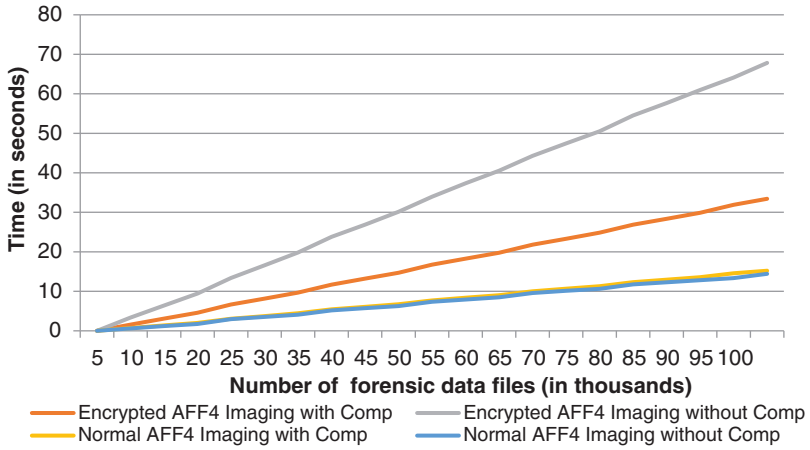

Figure 3: Imaging time of proposed framework

With the encrypted AFF4 imaging, the required times for imaging with and without data compression are 16.7125 and 15.358 s, respectively. Thus, the time is increased by about 1.3545 s (about 8 percent) when the data compression option is used. It is clear that the data compression option increases the execution time. However, data compression is a default option in AFF4 imaging and helps to reduce the size of the forensic image, as will be discussed later. However, the required time for normal AFF4 imaging without compression is 7.215 s. With data compression, the imaging time is 7.60 s. Thus, it can be seen that using the data compression option with the normal AFF4 imaging increases the imaging time by 5.42 percent. So, using AES encryption with AFF4 imaging approximately doubles the imaging time.

Fig. 4 compares the imaging times of the proposed framework with those of other related works, the Law et al. model [53] and Hou et al. model [56]. Here, only encrypted imaging without compression is used because these related works support only this imaging case. The imaging process of the model proposed by Law et al. [53] costs about 15.5 min. This is the required time for building and encrypting the index files and writing them to another external storage (external hard disk). Still, this solution needs more time making a bit-by-bit image from the suspect's storage, at the beginning, and this time is not measured here. In any case, the Law et al. model [53] takes 15.5 min (for building and encrypting the index files and writing them), plus the time required to make the bit-by-bit image.

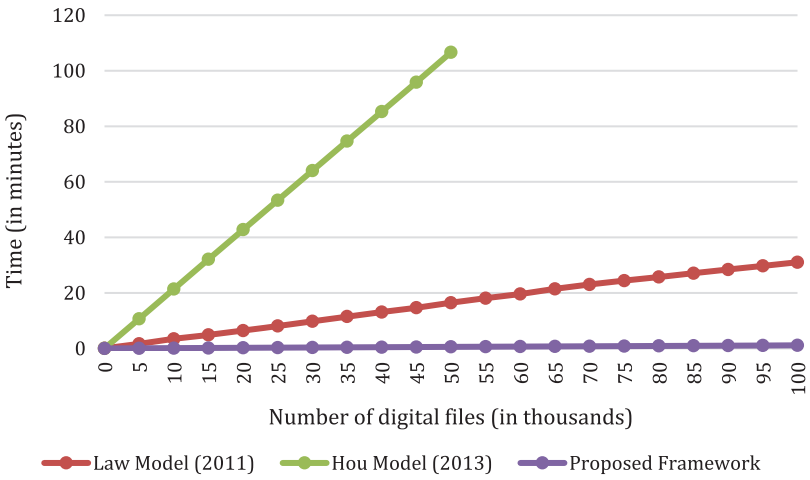

Figure 4: Imaging time of proposed framework and other related works

The model of Hou et al. [56] is the most expensive because of the need to encrypt every word, in each forensic file, using public key cryptography. We implement this model with 1024-bit RSA keys. The system gets overhead and stop responding when the number of forensic files is about 30,000 and, at that point, the time cost is 572.3161 min. With the proposed framework, the total imaging time is only 8.591 min. This is due to the direct encrypting of files with the AES cryptosystem. However, there is no need to tokenize each file and use a searchable encryption method as such process has been proven here to be an inefficient solution.

Finally, it is clear that the imaging time for the models of Law et al. [53] and Hou et al. [56] is about 2 and 222 times, respectively, greater than that of the proposed framework. This is a normal result because these related works use the concept of searchable encryption, which requires a longer time for two reasons. First, the time is greater because it requires that each file be tokenized into words, and then each word is encrypted. Second, it was pro-posed for storing private data in a remote server such as a cloud server, where the server's search efficiency is a concern. The data owner's document encryption and decryption efficiency is not an important issue since the data owner encrypts/decrypts his or her data sequentially (or partially and from time to time). The imaging efficiency of the Hou et al. model [56] is worse because of encrypting of each tokenized word using public key cryptography, which is not efficient compared to secret key cryptography systems such as AES. Based on the above result, the proposed solution increases the efficiency of the imaging process and as targeted by the objective of this research paper.

9.2 The Imaging Size

The size of forensic image can differ from one solution to another. In fact, a huge forensic image increases the imaging cost as it requires more storage space. The size of the dataset used is about 411 MB. Any computer forensics tool seems to produce an image with the same size for the relevant data. However, the case here is different because the data are going to be encrypted and may be compressed while using the AFF4 image to re-duce the image size.

Fig. 5 shows the image size of the proposed framework with the different imaging cases. When encrypted AFF4 imaging without compression is used, the forensic image size is 458 MB (480,830,714 bytes). Thus, compared to the dataset's size, it increased by about 47 MB. This is because of the additional data added to the AFF4 image such as the files' hash values, and metadata (name, size, extension, etc.). Moreover, some additional metadata are added by the used AFF4 programming tool.

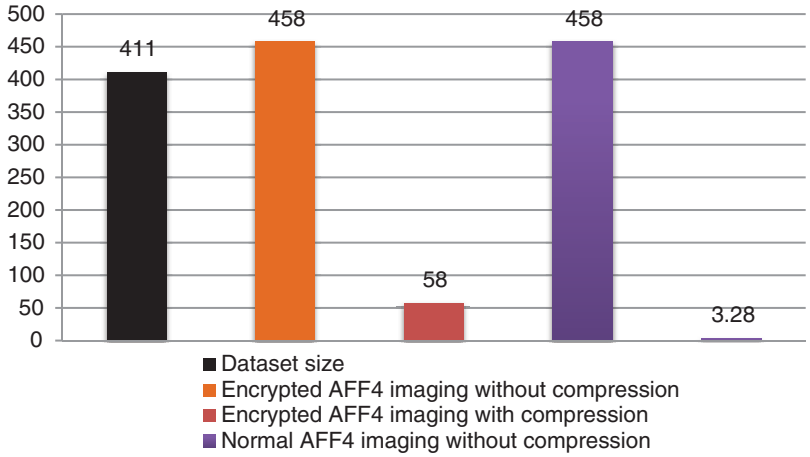

Figure 5: Image size of proposed framework

However, with the normal AFF4 imaging without compression, the image size is 458 MB (480,829,905 bytes). This is smaller than the previous case by 809 bytes. Because of that, the previous case takes more space for encryption padding. When data compression is used, the image size is reduced to 58 MB with the encrypted AFF4 imaging and 3.28 MB with the normal AFF4 imaging. This is a normal result because the imaged data (or dataset) are simple text files. The result would be increased if other data types (such as photos or videos) are used. It is well-known that data compression is more efficient with text and less efficient with the multimedia files.

Fig. 6 compares the image sizes required by the proposed framework and other related works [53–56]. Only two imaging cases are covered here in the proposed framework. These cases are encrypted AFF4 imaging with compression and encrypted AFF4 imaging without compression. The collected image size with the Law et al. model is 2140.99 MB (2,245,000,000 bytes). This is because each encrypted file's size in the Law et al. model is 21.9 KB (22.450 bytes), whereas the size of the used dataset is 411 MB, and the size of each file (inside the dataset) is only 4.21 KB (4,316 bytes). This is the result of tokenizing each file into words and encrypting each word into a fixed length one (using encryption padding with each word). Thus, the size of each imaged file is about five times greater than that of the original file. On average, the size of each encrypted forensic word is 37.41 bytes, even if a tokenized word is just one character or symbol.

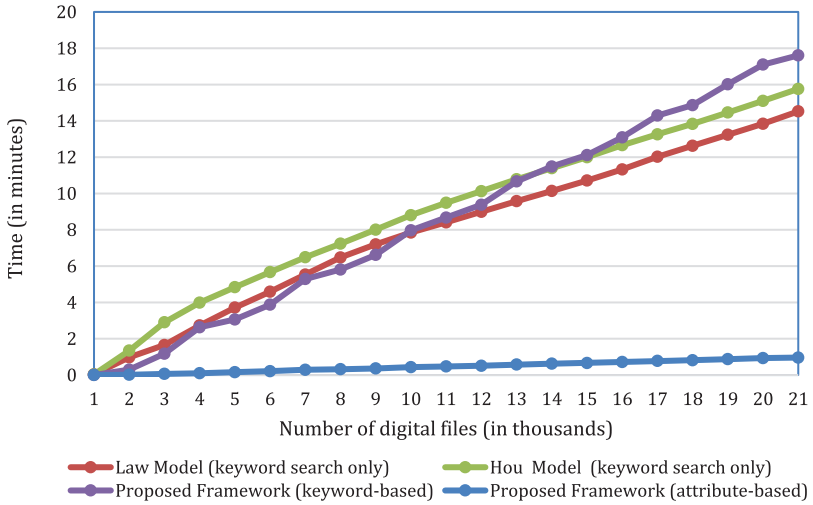

Figure 6: Image sizes in related works and proposed framework

For the model of Hou et al. [56] the image size is even greater. The image size is 7,916.45 MB (830,100,000,000 bytes). In fact, with this model, the size of each imaged file is increased from 4.21 KB to 81.06 KB (830,010 bytes). This means it is doubled about 19.25 times. This is because of tokenizing each file into words and using public key cryptography for encrypting each word.

In the proposed framework, the image size within two imaging cases is 458 MB and 58 MB. The other imaging case results are also shown in Fig. 6. Compared to the models of Law et al. and Hou et al., the forensic image size is reduced by 4.67 and 17 times, respectively. The result is much better if the data compression option is used, which is the default AFF4 option. When using data compression, the image size is reduced by 36.91 and 136.48 times, respectively, compared with the models of Law et al. [53] and Hou et al. [56].

Compared to the other related works, it is clear that the proposed research reduces the collected forensic image size several times. This is because of tokenizing each file into words and encrypting each word

separately by the related works. Therefore, the proposed framework improves the imaging efficiency by reducing the required time and storage. So, the targeted selective imaging module is an efficient one, compared to other works.

9.3 The Analysis Efficiency

The analysis efficiency is evaluated using two different criteria: the search and decryption times. Fig. 7 shows the required keyword-based search time for the proposed framework, along with the other two related works [53,56]. The keyword-based search time for the Law et al. model [53] is 7.26 s, and for the Hou et al. model is 7.48 s. These works require the investigator to encrypt any keyword to start the search process. In the Hou et al. model [56], encrypting a single keyword requires about two seconds since public key encryption is used. With the Law et al. model [53], the time for encrypting a single word is trivial. However, in the proposed framework, the required key-word-search time is reduced to 5.365 s.

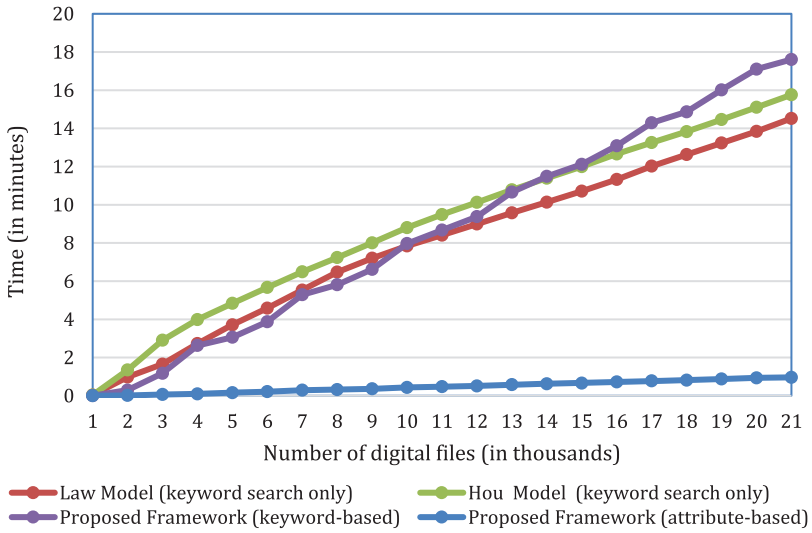

Figure 7: Required search times for related works and proposed framework

For the keyword-based search, the related works require encrypting a keyword and then searching inside the encrypted files. This means searching without decrypting the data files. This is why they come out with acceptable searching engines because they use searchable encryption schemes that are proposed for cloud computing. Using the pro-posed framework, each file needs to be decrypted before searching for it and, even at that, the searching time is still acceptable and more efficient than the related works.

Using the attribute-based search provided only by the proposed framework, the required search time is less than one seconds. This is because of processing the files' attributes (or metadata) inside the AFF4 image, without requiring the decryption or even processing of the files' contents.

Fig. 8 shows the required decryption times for the proposed framework, along with the other related works [53,56]. Using the model of Law et al. [53] decrypting all the collected forensic files requires about 15.5 min. With the model of Hou et al. [56] the system gets overhead when it reaches about 35,807 files. At this stage, the system already spent about 611 min. Thus, the model of Hou et al. cannot scale with more than 35,806 files. This is due to using the public key cryptography for encrypting every tokenized word in every forensic file. The proposed framework reduces the required decryption time to only 8.5 min. It can be observed that the related works have the same decryption efficiency problem found during the encryption process. The problem comes from the same efficiency gap, which is

tokenizing each file into words and decrypting each word separately. The same gap leads to producing a huge forensic image size. In the proposed framework, using AES cryptosystem for encrypting the relevant data files leads to having an efficient decryption and encryption along with producing an acceptable forensic image size. So, the proposed selective analysis module is an efficient solution, compared to other related works, and as targeted by the objective of this research paper.

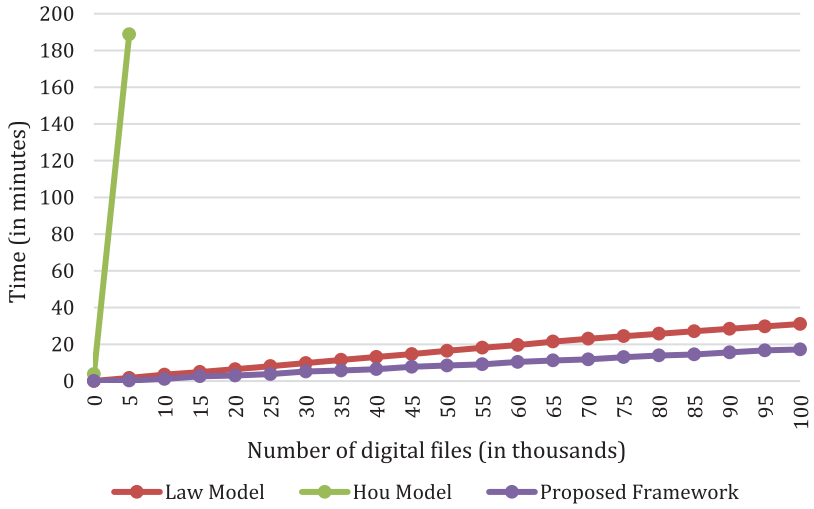

Figure 8: Required decryption times for the related works and proposed framework

9.4 The Analysis Sufficiency

The analysis sufficiency is evaluated using three criteria: i) keyword-based search; ii) attribute-based search; and iii) integration with other computer forensics analysis tools.

In terms of keyword-based search, the ability to use a keyword-based search to find the necessary digital evidence is discussed. The models of both Law et al. [53] and Hou et al. [56] support only a keyword-based search and the keywords must be prepared before collecting the forensic data. The investigator will not be able later to enter any new key-word. Moreover, there is no grantee that the prepared keywords will cover all relevant evidence. In the proposed framework, the investigator uses other computer forensics tools to select the relevant data. Existing tools (such as EnCase or CnWRecover) provide professional search engines with different search types. Then, at the analysis stage, the investigator can sequentially prepare and use several keywords at different times until the needed digital evidence is found.

For the attributes-based search, the ability to use an attribute-based search for analyzing the collected data is evaluated here. In fact, only the proposed framework supports this kind of search. An attribute-based search can be used for both private and non-private data. It relies on forensic files' attributes such as the name, extension, size, and so on.

Regarding integration with computer forensics analysis tools, analyzing the collected encrypted forensic data using existing computer forensics tools (such as EnCase, FTK AccessData) provides several advantages since these tools have advanced analysis and re-porting features that cannot be provided by a single research effort. These tools are also widely used and accepted worldwide by courts of law. The collected encrypted data produced by the models of Law et al. [53] and Hou et al. [56] cannot be analyzed using such tools because the searchable encryption schemes used by these models are not supported at all by these computer forensics tools. In the proposed framework, the two most popular forensics tools (Encase and FTK AccessData) can be used for analyzing the collected AFF4 image. The new versions of these tools

(EnCase 7.1 and above & FTK 3.0 and above) sup-port the AFF4 image. Thus, any non-encrypted data produced by the proposed framework can be directly and completely analyzed. The encrypted (private) data files (inside the AFF4 image) can still be analyzed directly using their metadata for example, using an attribute-based search to list some kind of files based on their type, size, and so on. Also, the encrypted data file can be analyzed by such tools as they support AES encryption. The encrypted files' contents inside the collected AFF4 image by the proposed framework can be analyzed with the latest versions of X-Ways and FTK AccessData. This is because the latest versions of these tools can analyze AFF4 forensic images.

10 Conclusion

In this research paper, a lawful and efficient privacy-preserving computer forensics framework is proposed. It provides an efficient imaging and analysis while providing sufficient analysis methods. The lawful requirements for privacy preservation in digital forensics are investigated. The proposed framework is implemented and evaluated from different perspectives. Potential future work includes developing a data selection tool for privacy-preserving computer forensics, since this issue still presents a research gap. An-other issue is adopting the proposed framework to be applicable to other digital forensics branches, especially network and mobile forensics.

Funding Statement: The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through research group no (RG-1441-531).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present research.

References

[1] P. Stephenson, "The forensic investigation steps," *Computer Fraud & Security*, vol. 2002, no. 10, pp. 17–19, 2002.

[2] P. Stephenson, "Comprehensive approach to digital incident investigation," *Information Security Technical Report*, vol. 8, no. 8, pp. 42–54, 2003.

[3] P. Stahlberg, G. Miklau and B. Levine, "Threats to privacy in the forensic analysis of database systems," in *Proc. of ACM SIGMOD*, New York, NY, USA, pp. 91–102, 2007.

[4] M. Khanafseh, M. Qatawnehand and W. Almobaideen, "A survey of various frameworks and solutions in all branches of digital forensics with a focus on cloud forensics," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 8, pp. 610–629, 2019.

[5] C. Horan and H. Saiedian, "Cyber crime investigation: Landscape, challenges, and future research directions," *Journal of Cybersecurity and Privacy*, vol. 1, no. 4, pp. 580–596, 2021.

[6] M. Burmester, Y. Desmedt, R. Wright and A. Yasinsac, "Security or privacy, must we choose?," in *Proc. of Symp. on Critical Infrastructure Protection and the Law*, Computer Science and Telecommunication Board, Cambridge, UK, 2002.

[7] S. Bui, M. Enyeart and J. Luong, *Issues in Computer Forensics*, USA: Santa Clara University Computer Engineering, 2003. [online]. Avaialble at https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.4586&rep=rep1&type=pdf.

[8] E. Spafford, "Some challenges in digital forensics," in *Proc. of Advances in Digital Forensics II, IFIP Advances in Information and Communication*, Orlando, Florida, pp. 3–9, 2006.

[9] N. Croft and M. S. Olivier, "Sequenced release of privacy accurate call data record information in a GSM forensic investigation," in *Proc. ISSA*, Pretoria, South Africa, pp. 1–14, 2006.

[10] W. Halboob, M. Abulaish and K. S. Alghathbar, "Quaternary privacy-levels preservation in computer forensics investigation process," in *Proc. ICITST 2011*, Abu Dhabi, United Arab Emirates, pp. 777–782, 2011.

[11] F. Armknecht and A. Dewald, "Privacy-preserving email forensics," *Digital Investigation*, vol. 2015, no. 14, pp. S127–S136, 2015.

[12] M. B. Seyyarab and Z. J. M. H. Geradts, "Privacy impact assessment in large-scale digital forensic investigations," *Digital Investigation*, vol. 33, pp. 1–9, 2020.

[13] J. A. Yaacoub, N. N. Hassan, O. Salman and A. Chehab, "Digital forensics vs. anti-digital forensics: Techniques, limitations and recommendations," arXiv, 2021. [online]. Avalaiable at https://arxiv.org/abs/2103.17028v1.

[14] D. Choi, "Digital forensic: Challenges and solution in the protection of corporate crime," *Journal of Industrial Distribution & Business*, vol. 12, no. 6, pp. 47–55, 2021.

[15] S. S. Kazemi and S. Heidari, "Digital forensics and its role in promoting criminal prosecution," *Electronic Journal of Management, Education and Environmental Technology*, vol. 25, no. 5, pp. 1–15, 2021.

[16] R. Muir and S. Walcott, "Unleashing the value of digital forensics," The Police Foundation Report, UK, London. 2021. [online] Available at https://www.police-foundation.org.uk/2017/wp-content/uploads/2010/10/value_of_digital_forensics.pdf.

[17] I. Jayaraman and A. S. Panneerselvam, "A novel privacy preserving digital forensic readiness provable data possession technique for health care data in cloud," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 4911–4924, 2021.

[18] X-Ways. 2021. [online]. Available at https://www.x-ways.net/.

[19] FTK AccessData. 2021. [online]. Available at https://accessdata.com/.

[20] The Advanced Forensics File Format (AFF4). 2021. [online]. Available at https://www.loc.gov/preservation/digital/formats/fdd/fdd000412.shtml.

[21] The SleuthKit Open Source Forensics. 2021. [online]. Available at https://www.sleuthkit.org/.

[22] J. Stüttgen, A. Dewald and F. C. Freiling, "Selective imaging revisited," in *Proc. of the Seventh Int. Conf. on IT Security Incident Management and IT Forensics*, Nuremberg, Germany, pp. 45–58, 2013.

[23] P. Joseph and J. Norman, "Forensic corpus data reduction techniques for faster analysis by eliminating tedious files," *Information Security Journal: A Global Perspective*, vol. 28, no. 4–5, pp. 136–147, 2019.

[24] E. E. Kenneally and C. L. Brown, "Risk sensitive digital evidence collection," *Digital Investigation*, vol. 2, no. 2, pp. 101–119, 2005.

[25] P. Turner, "Unification of digital evidence from disparate sources (digital evidence bags)," *Digital Investigation*, vol. 2, no. 3, pp. 223–228, 2005.

[26] P. Turner, "Selective and intelligent imaging using digital evidence bags," *Digital Investigation*, vol. 3, no. 1, pp. 59–64, 2006.

[27] G. Richard and V. Roussev, "Breaking the performance wall: The case for distributed digital forensics," in *Proc. of DFRWS'04*, Baltimore, Maryland, pp. 1–16, 2004.

[28] M. Cohen and B. Schatz, "Hash based disk imaging using AFF4," *Digital Investigation*, vol. 7, no. 2010, pp. S121–S128, 2010.

[29] J. Stüttgen, "Selective imaging: Creating efficient forensic images by selecting content first," Master Dissertation, Friedrich Alexander Universität, Erlangen, Nürnberg, 2011.

[30] G. Richard and V. Roussev, "File system support for digital evidence bags," in *Proc. of Advances in Digital Forensics II - IFIP Advances in Information and Communication Technology*, Springer, Boston, MA, USA, pp. 29–40, 2006.

[31] P. Turner, "Applying a forensic approach to incident response, network investigation and system administration using digital evidence bags," *Digital Investigation*, vol. 4, no. 1, pp. 30–35, 2007.

[32] J. Griera and G. Richard II, "Rapid forensic imaging of large disks with sifting collectors," *Digital Investigation*, vol. 14, no. 2015, pp. S34–S44, 2015.

[33] S. Garfinkel, D. Malan, K. A. Dubec, C. Stevens and C. Pham, "Advanced forensic format: An open extensible format for disk imaging," in *Proc. of Advances in Digital Forensics II, IFIP Advances in Information and Communication*, Orlando, Florida, pp. 13–27, 2006.

[34] E. Imager. 2021. [online]. Available at https://security.opentext.com/document/product-brief/encase-forensic-imager.

[35] FTK Imager. 2021. [online]. Available at https://accessdata.com/product-download/ftk-imager-version-4-5.

[36] W. Halboob, K. S. Alghathbar, R. Mahmod, N. I. Udzir, M. T. Abdullah *et al.,* "An efficient computer forensics selective imaging model," in *Proc. of LNEE*, Berlin, Heidelberg, pp. 277–284, 2004.

[37] M. Caloyannides, *Privacy Protection and Computer Forensics*, 2nd ed., Boston, London, UK: Artech House Publishers, 2004.

[38] M. Saboohi, *Collecting Digital Evidence of Cyber Crime*, Islamabad, Packistan: International Islamic University, 2000. Avaialable at https://www.academia.edu/1375440/COLLECTING_DIGITAL_EVIDENCE_OF_CYBER_CRIME.

[39] C. W. Adams, "Legal issues pertaining to the development of digital forensic tools," in *Proc. of SADFE '08*, Oakland, California, USA, pp. 123–132, 2008.

[40] M. A. Fahdi, N. L. Clarke and S. M. Furnell, "Challenges to digital forensics: A survey of researchers & practitioners attitudes and opinions," in *Proc. of Information Security for South Africa*, Johannesburg, South Africa, pp. 1–8, 2003.

[41] A. Nieto, R. Rios and J. Lopez, "Privacy-aware digital forensics," in *Proc. of Security and Privacy for Big Data, Cloud Computing and Applications*, NICS Lab, Málaga, Spain, pp. 1–39, 2019.

[42] S. Srinivasan, "Security and privacy in the computer forensics context," in *Proc. of ICCT'6*, Guilin, China, pp. 1–3, 2006.

[43] S. Srinivasan, "Security and privacy vs. computer forensics capabilities," *Information Systems Control Journal*, vol. 4, pp. 1–3, 2007.

[44] W. Halboob, R. Mahmod, N. I. Udzira and M. T. Abdullah, "Privacy levels for computer forensics: Toward a more efficient privacy-preserving investigation," *Procedia Computer Science*, vol. 56, no. 2015, pp. 370–375, 2015.

[45] W. Halboob, R. Mahmod, N. I. Udzir and M. T. Abdullah, "Privacy policies for computer forensics," *Computer Fraud & Security*, vol. 2015, no. 8, pp. 9–13, 2015.

[46] A. Gupta, "Privacy preserving efficient digital forensic investigation framework," in *Proc. of 6IC3*, Noida, India, pp. 387–392, 2013.

[47] S. Saleem, O. Popova and I. Bagilli, "Extended abstract digital forensics model with preservation and protection as umbrella principles," *Procedia Computer Science*, vol. 35, no. 2014, pp. 812–821, 2014.

[48] M. Reith, C. Carr and G. Gunsch, "An examination of digital forensic models," *International Journal of Digital Evidence*, vol. 1, no. 3, pp. 1–12, 2002.

[49] A. Nieto, R. Rios and J. Lopez, "A methodology for privacy-aware IoT-forensics," in *Proc. of IEEE Trustcom/BigDataSE/ICESS*, Sydney, NSW, Australia, pp. 626–633, 2017.

[50] "ISO/IEC 29100:2011 Information technology—Security techniques—Privacy framework," 2011. [online]. Avalaiable at https://www.iso.org/standard/45123.html.

[51] R. I. Ferguson, K. Renaud, S. Wilford and A. Irons, "PRECEPT: A framework for ethical digital forensics investigations," *Journal of Intellectual Capital*, vol. 21, no. 1, pp. 257–290, 2020.

[52] L. Englbrecht and G. Pernul, "A privacy-aware digital forensics investigation in enterprises," in *Proc. of the 15th Int. Conf. on Availability, Reliability and Security*, Virtual Event, Ireland, pp. 1–10, 2020.

[53] F. Y. W. Law, P. P. F. Chan, S. M. Yiu, K. P. Chow, M. Y. K. Kwan *et al.,* "Protecting digital data privacy in computer forensic examination," in *Proc. of 6th IEEE Int. Workshop on Systematic Approaches to Digital Forensic Engineering*, Oakland, CA, USA, pp. 1–6, 2011.

[54] 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. of 7th Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing*, Dalian, China, pp. 378–383, 2011.

[55] 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. of 3rd Int. Conf. on Multimedia Information Networking and Security*, Shanghai, China, pp. 595–599, 2011.

[56] S. Hou, S. M. Yiuy, T. Ueharaz and R. Sasakix, "A privacy-preserving approach for collecting evidence in forensic investigation," *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, vol. 2, no. 1, pp. 70–78, 2013.

[57] OECD Privacy Guidelines. 2021. [online]. Available at https://www.oecd.org/sti/ieconomy/privacy-guidelines.htm.

[58] APEC Privacy Framework. 2021. [online]. Available at https://iapp.org/resources/article/apec-privacy-framework/.

[59] CnWRecovery. 2021. [online]. Avalaibale at https://www.cnwrecovery.com/.

[60] NetBeans IDE, 2021. [online]. Available at https://netbeans.org/features/java/index.html.

[61] Java Cryptography Extension (JCE). 2021. [online]. Available at https://www.oracle.com/java/technologies/javase-jce8-downloads.html.

[62] CSVReader. 2021. [online]. Available at http://opencsv.sourceforge.net/apidocs/com/opencsv/CSVReader.html.