Journal of Education and Science (ISSN 1812-125X), Vol: 32, No: 01, 2023 (109-124)

Digital Forensic Tools: A Literature Review

Karam Muhammed Mahdi Salih[1*], Najla Badi Ibrahim[2]

Department of Computer Network, Ninevah University, Mosul, Iraq
Department of Computer Science, Computer Science and Mathematics College, University of Mosul, Mosul, Iraq

E-mail: [1*]Karam.mahdi@uoninevah.edu.iq, [2]Najlabadie@uomosul.edu.iq

(Received December 20, 2022; Accepted February 22, 2023; Available online March 01, 2023)

DOI: 10.33899/edusj.2023.137420.1304, © 2023, College of Education for Pure Science, University of Mosul.

Abstract

Digital forensics is a process of collection, identification, extraction, and documentation of electronic evidence which is used in a court of law. There are a large number of tools that help us to make this process easy and simple. In this paper, four tools have been chosen to explore and study. The best digital forensic tools have been chosen according to different parameters in each type of digital forensics. The (Stellar) basis tool and (Forensic Tool Kit) have been explored for computer forensic tools while (Network Map) has been chosen for network forensic tools and (OSFmount) has been studied as a live forensic tool. This paper also covers other types of forensic tools like Database forensic tools, O.S. forensic tools, and Mail forensic tools. The role of Artificial intelligence in Digital Forensic tools has been discussed in this paper by using both Decision Stump and Bayes net machine learning techniques. After making an investigation of the IoT device traffic dataset using these two techniques, Decision Stump gives us less accurate results compared with Bayes net.

Keywords: Forensics tool, Digital Evidence, Artificial intelligence.

أدوات الأدلة الجنائية الرقمية، مراجعة عامة

كرم محمد مهدي[1*]، نجلاء بديع ابراهيم[2]

[1*]قسم الشبكات، كلية تكنولوجيا المعلومات ، جامعة نينوى، الموصل، العراق
[2]قسم علوم الحاسوب، كلية علوم الحاسوب والرياضات، جامعة الموصل، الموصل، العراق

الخلاصة

الأدلة الجنائية الرقمية هي عملية جمع وتحديد واستخراج وتوثيق الأدلة الإلكترونية والتي يتم استخدامها في المحاكم. هناك الكثير من الأدوات التي تساعدنا في جعل هذه العملية سهلة وبسيطة. في هذه المقالة تم اختيار أربعة أدوات للاستكشاف والدراسة. تم اختيار أفضل أدوات الأدلة الجنائية الرقمية وفقًا لمعايير مختلفة في كل نوع من أنواع الأدلة الجنائية الرقمية. تم استكشاف أداة أساس Stellar ومجموعة أدوات الأدلة الجنائية الرقمية (FTK) لأدوات التحليل الجنائي للكمبيوتر بينما تم اختيار خريطة الشبكة (Nmap) لأدوات التحليل الجنائي للشبكة. وتم دراسة OSFmount كأداة تحليل مباشر Online للأدلة الجنائية الرقمية. تغطي هذه الورقة أيضا أنواعًا أخرى من أدوات الأدلة الجنائية الرقمية مثل أدوات الأدلة الجنائية الرقمية لقاعدة البيانات، وأدوات الأدلة الجنائية الرقمية لنظام التشغيل، وأدوات الأدلة الجنائية الرقمية للبريد الإلكتروني. وقد تم مناقشة دور الذكاء الاصطناعي في أدوات الأدلة الجنائية الرقمية في هذه الورقة باستخدام تقنيات التعلم الآلي لكل من Decision Stump و Bayes net. بعد اجراء تحقيقات على مجموعة بيانات مرور لجهاز IoT باستخدام هاتين التقنيتين أعطتنا Decision Stump نتائج أقل دقة مقارنة بشبكة Bayes.

الكلمات المفتاحية: الدليل الرقمي، الذكاء الاصطناعي، أداة التحليلات الجنائية.

1. Introduction

Electronic devices like computers, laptops, and portable devices are expanding exponentially. These kinds of devices use a large amount of data, interconnected to the network so the Internet is considered the main cause of cybercrime. Digital forensics (DF) is considered a newborn compared with other forensics sciences. The role of DF science starts after the occurrence of the crime [1]. DF is a process of collection, identification, extraction, and documentation of electronic evidence from different electronic devices and then use in a court of law as legal pieces of evidence. [2]. The investigation process mainly depends on the DF tools and it will give us effective and efficient results. They are different types of data to deal with these tools like the Internet of Things (IoT) devices data, computer devices, mobile devices cloud computing, etc. [3].

Most of these tools' goals are to collect and recover the original files from the devices. DF tools are used for solving problems related to computer crimes like phishing, money laundering, bank Fraud, and child exploitation. The most of shreds of evidence have been found on computers [4]. As shown in figure 1, DF tools are divided into computer forensics network forensics, live forensics, Operating System forensics, database forensics, and Mail forensics.

As a part of Artificial intelligence (AI) machine learning (ML) generally and deep learning especially have an important role in DF. As we know the AI technique can work with big data in a short time with accurate results. So, AI helps the investigators in the DF analysis process. The exuberance of forensic tools will make it hard for users to choose the relevant tool for their requirements [4], [5]. So, we explored the most popular tools and collect information about others to make a comparison between them. However, the investigator can choose the appropriate one for him and for the crime that he will investigate.

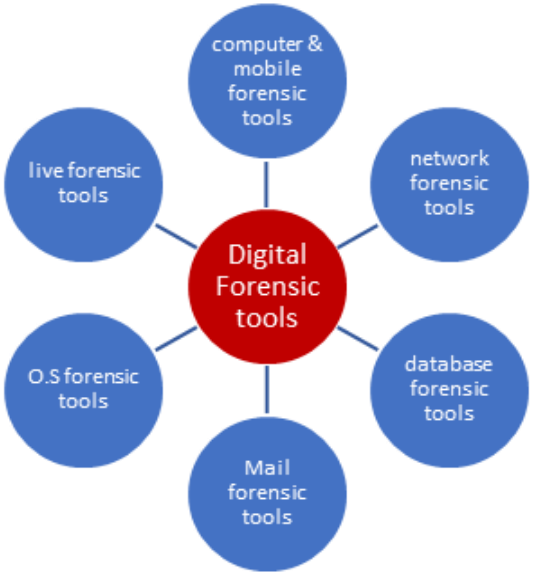

Figure 1 Types of digital forensic tools

In this study, we highlighted the best DF tool in each type: Computer, Network, and Live forensic tools according to several key parameters that consider important parameters in that type. For example, imaging and hashing consider important parameters in computer forensic tools while port scanning and packet analyzing are important key parameters for network forensic tools. Live log analysis and RAM dumping are important key parameters for live forensic tools. The use of Artificial intelligence in DF tools has been discussed in this paper by using both Decision Stump and Bayes network machine learning techniques on IoT device datasets and a comparison between them has been made. Decision Stump gives us less accurate

results compared with Bayes net because it doesn't care that much about the attributes or their relationships. On other hand, we got the best results from Bayes Net because it represented the conditional dependencies of a set of random variables. Each node in the network represents a variable, and each directed edge in the network represents a conditional relationship.

In section 2, literature reviews have been made about digital forensics, digital forensic models, digital forensic tools, and the techniques used in this field, especially with artificial intelligence techniques like machine learning and pattern recognition. In section 3, digital forensic stages have been explored with their process model. In section 4, a study of digital forensic tools have been made. Different types of digital forensic tools have been highlighted like computer forensic tools, network forensic tools, live forensic tools, Operating System (O.S) forensic tools, mail forensic tools, and database forensic tools. Section 5 highlighted the importance of using metrics to validate Digital Forensic tools to help the investigator choose the best one. Section 6 focuses on the role of artificial intelligence in digital forensic tools. A simulation of DecisionStump and Bayes net have been made with a comparison of their results. Section 7 explores the challenges which face digital forensic tools while section 8 provides us the conclusion of this study and future work.

2. Literature reviews

In [4] they analyzed many DF analysis tools. It said the pattern recognition technique is perfect for the analysis stage of the DF. The recognized patterns provide features that are used to develop many DF tools. So, various tools are not only used to preserve and analyze pieces of evidence data but are also important to find the solution to all the conflicts that occur in the execution phase. In [6] described different techniques used for live and dead forensic analysis. It keeps the important commands by different DF tools like WIRESHARK, Autopsy, O.S forensic, TRUECRYPT, Forensic Tool Kit (FTK) Imager, and SANS SIFT and also generates a comprehensible atmosphere to help a detective. Besides, they accumulate information that can be transformed using live analysis, which sidesteps destroying the information due to the stoppage of the target node. [7] said the major process done by criminals is for destroying files by deleting, damaging, or overwriting hard disks, etc. The team only focused on how to recover the destruction data. To recover the damaged data with the help of different tools such as WIRESHARK, Autopsy, TRUECRYPT, FTK Imager, Operating system forensic, X-WAYS, and SANS SIFT.

Researchers in [8] explained the attributes, constrictions, and applications of DF tools and compared them with other tools in assisting investigators or users in employing composite DF tackles for their inspection. [9] employed a machine learning technique and advising a scheme to diagnose abnormal packets and attacks. Naive Bayesian provided the best accuracy against other classifiers. [10] used NLP techniques to analyze DF shreds of evidence. [11] focused on the recent readiness and advances of DF tools in the composite atmosphere. [12] proposed a method to build a new intelligence DF model for storehouse willingness. [13] suggested an effective model for DF cloud Investigation called Cloud Forensics Investigation Model (CFIM) to pattern the crimes happening in the cloud forensically. [14] proposed a DF framework methodology for the social media network community. This system contains operative classifying digital devices, procedures, analyzing and obtaining DF pieces of evidence. [7] showed DF terms in the cyber world and informed a comparative analysis of the current stream state of forensics. [15] proposed building architecture for AI applications in the DF especially in the analysis stage. [16] described an analysis of up-to-date DF artificial intelligent schemes to raise these procedures in forensic correction. [17] said that the compression of data can disturb different DF stages. [18] analyzed the different ML techniques and their usability in recognizing evidence by tracking file systems. The Machine Learning algorithms achieved good outcomes. [19] proposed a classification model for network traffic using Machine Learning techniques. The results revealed that the best outcome had been done by a random

forest classifier. [20] analyzed network traffic to discover windows ransomware by spread on ML and accomplished a Total Form (TF) a percentage of 97.1% with the decision tree method.
Researchers in [21] proposed a process of the text description of Natural Language Processing and spam email discovery. [22] suggested a model for the cataloging of attacks in the cloud atmosphere using ML procedures with a DF method. [23] proposed model of managing intellectual cybersecurity. The model practices AI procedures to make the analysis procedure of cybersecurity more proficient compared with old-style security instruments. Through the speedy development of technologies, it is important to select DF methods and frameworks.
DF methods from 2015 to 2022 are offered in the next lines. [24] examine the environments, cruise anomaly information, and control relation report. [25] Conformist data collection process strategy, provision law of shaping the consistency of the DF pieces of evidence. [26] study the DF on IEC/ISO ethics. [27] employing Digital Forensic Readiness (DFR) mechanism in amenability through the IEC/ISO ethics. [28] proposed a model based on Online Natural Language Processing (NLP) for forensic investigation. The paper compared different DF tools in different groups such as computer Forensic Tools, Network Forensic Tools, O.S Forensic Tools, Live Forensic Tools, Database forensic tools, and Email Forensic Tools. Consequently, the investigators can choose the accurate tool used for their requirements easily. The paper also highlighted the use of AI in DF investigation.

3. Digital Forensic stages

In DF the first prototypical suggested has four stages: Collection, Identification, Assessment, and Admission. then, a different prototypical is suggested to describe the stages of collecting, analyzing, preservation, and reportage of the pieces of evidence produced by many devices. Recently, a growing number of extra complicated prototypical are suggested. The goal of these models is to speed up the whole investigation procedure. The variety of sources and devices of digital shreds of evidence results in a variety of DF procedure models [29]. There is no common procedure model appropriate to use for all forms of the investigation process. [30]. Figure 2 shows the different stages of DFs. The role of each phase is discussed below:

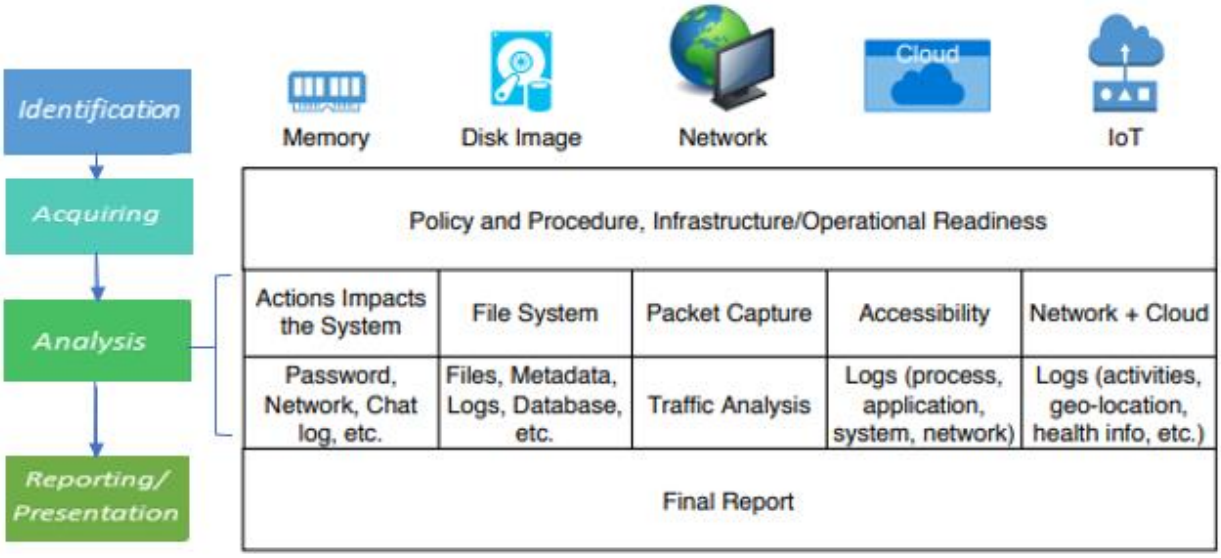

Figure 2 Different Types of digital forensics

Despite DF being a new study zone, already has completed important growth. The growth is done by the enhancement of methodologies and technology, for example, tools for gathering and analyzing DF pieces of evidence. In DF, a method to do an investigation process is called a process model which is a context

with a sum of stages to do an investigation. In DF investigation a standard methodology should define the sequence of actions need in the investigation process. A perfect process model should be wide-ranging, which means it should be applied to a large number of cases. If a framework is very simple and has fewer phases, the result is not provided good guidance to the process of investigation. Otherwise, if a framework has more stages with sub-steps of each stage, the result is more limited to its usage.

A lot of studies are extra emphasis on clarifying the all procedure of DF investigation, important DF frameworks had been discussed in [31]. Further lately, research on the DF framework emphasizes a single step or solving more specific problems like evidence collection, examination, analysis, and preservation. For example, the triage model [4], [32] is efficient for the situation when time is a critical parameter. Through using the DF pledge, detectives get information about the illicit faster as a replacement for waiting for all reports which might gross weeks or months or lengthier.

3.1. Identification stage

This stage is used to define and examine the pieces of evidence and their location and where it is found. Shreds of evidence should be handled properly and carefully. The goal of this stage is to protect the integrity of pieces of evidence, it should be protected along with a log called Chain Of Custody (COC) which is known by way of the paper track, the DF linkage, or the sequential certification of DF evidence. It indicates the collection, sequence of control, transfer, and analysis. Figure 3 shows the usage of a chain of custody at the investigation time.

Your Logo Here		Your Address Here

[Agency Name] Case #:	

Item #	Date/Time Removed	Reason for Removal of Evidence	Signature

Figure 3 A sample of a chain of custody (COC)

3.2. Acquiring stage

For more analysis, this stage helps to save the state of the pieces of evidence. In this stage, hard disk imaging is done as a copy of the data on the hard disk. Three kinds of acquisition are accepted according to law enforcement forensic duplication, mirror image, and live acquisition. A mirror image makes a forensics duplication which saves the backup of the device's hard disk as a bit-for-bit cloning copy.

3.3. Analysis stage

In the analysis stage, three kinds of analysis can be done: limited analysis, partial analysis, or full analysis. The limited analysis works with only specific shreds of evidence. Partial analysis works with the cookies, log documents, e-mail files, etc. while complete investigation helps to discover the original cause of the crime happening. [4] different tools designed for the analysis stage like FTK and Encase which can deal with a large number of scripts to get information from the data that is to be analyzed.

3.4. Reporting/Presentation stage

The reporting stage helps to deduce, in a documented report form based on pieces of evidence. This is done with the help of digital crime laws represent the information for further investigation.

4. Digital Forensic tools

Digital Forensic Tools are the program applications designed for the DF investigation process in digital Criminalities. Different DF tools are accessible in the marketplace. They are moreover commercially licensed or general forms. We will talk in this paragraph about DF tools in different groups and make a proportional study of various tools in each group. DF tools have been selected according to different criteria like technical considerations, general issues, disk imaging, string searching, and legal issues.

Various DF tools have been chosen to explore in this study as shown in figure 4. We will explain them in detail in the next sections.

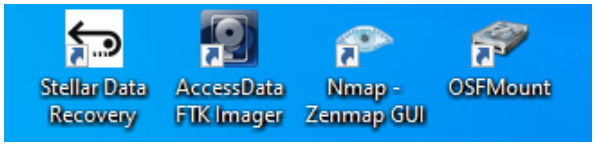

Figure 4 Different Digital Forensic tools Icons

A. Computer forensic tools

Computer DF tools are intended to certify that the pieces of evidence taken out from computers are correct and dependable. There are different types of computer DF tools like Data and Disk seizure DF tools. A comparative investigation of five Computer forensic tools based on feature parameters questions have been made. For example, hashing, imaging, and data recovery. In this paper, the Stellar tool and Forensic Tool Kit (FTK) have been explored in this review for computer forensic analysis.

Stellar: Stellar tool helps the investigator to find all files they want from the computer disk. Stellar is designed to be a comprehensive recovery tool to help its users to deal with all types of data loss scenarios, without needing any expert knowledge. Digital investigators can do normal or deep scanning. Figure 5 show the deep scanning mode. It does a whole signature-based file search which is useful for recovering the files that normal scanning could not found it. Figure 6 show the recovery of deleted files using Stellar tool.

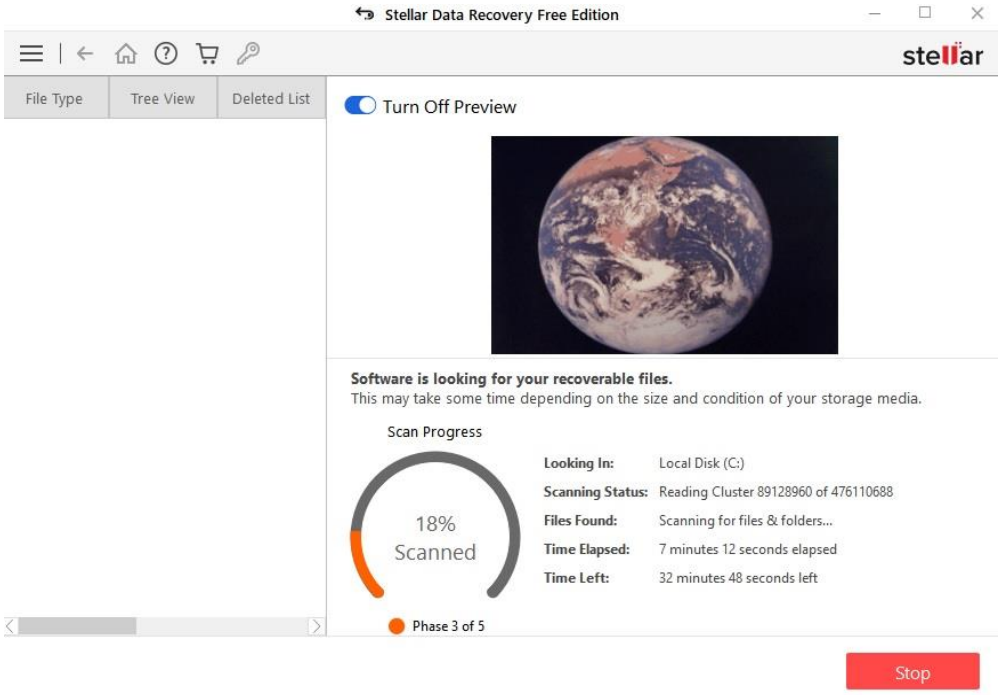

Figure 5 Deep scanning using Stellar

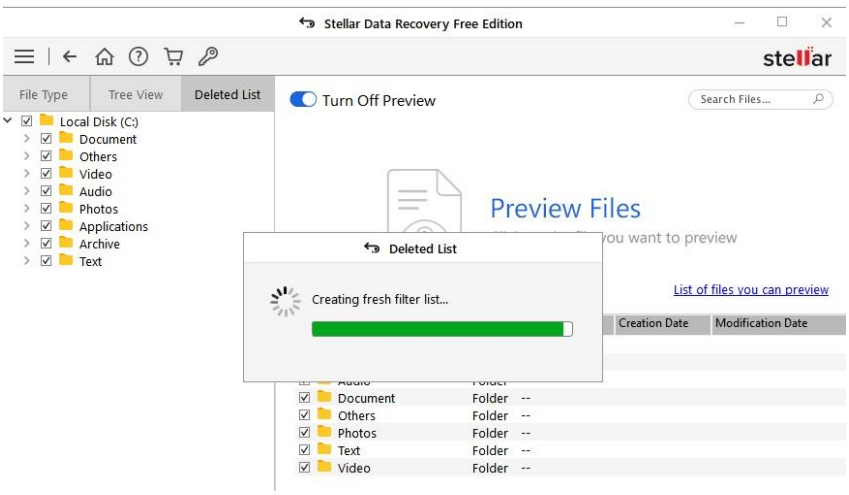

Figure 6 Deleted files recovery using Stellar

Forensic Toolkit, or FTK made by AccessData, is a computer forensics tool that scans a hard drive for finding different information. for example, it can scan a disk or locate deleted emails for text strings to crack encryption by using them as a password dictionary.

FTK is also associated with a disk imaging program named Forensic Tool Kit Imager. This tool takes an image clone of a hard disk and then calculates hash values i.e., message-digest algorithm (MD5) or Secure Hash Algorithms (SHA1), and validates the integrity of it by comparing it with the original one. The forensic data image can be saved and analyzed in different formats, like E01, DD/raw, and AD1 as shown in Figure 7 using FTK tool in the files analysis process.

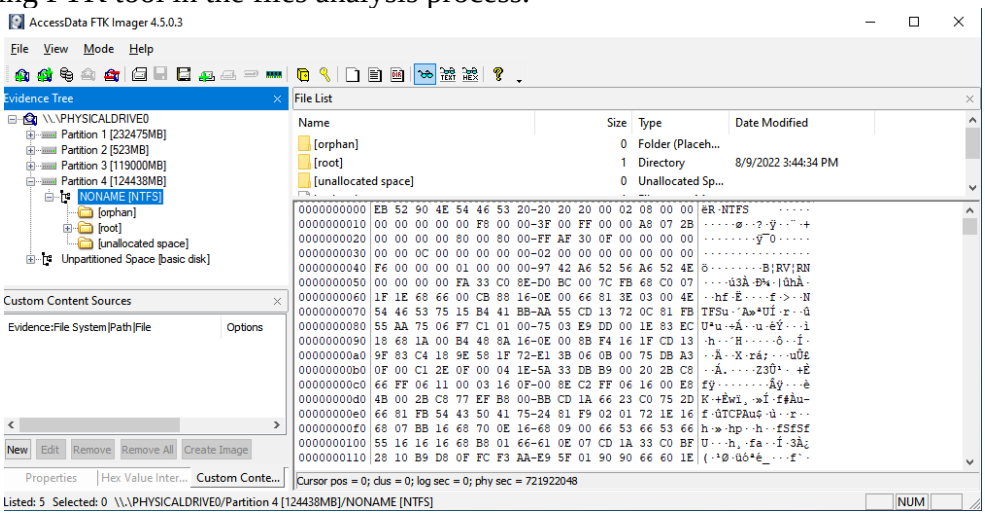

Figure 7 Analysis files using FTK

Table 1 has the comparison of key parameters like imaging which is a technique of copying physical storage for making investigations and gathering shreds of evidence.

The copy does not only include files, but every bit, sector, partition, files, deleted files, folder, and also unallocated spaces. The copy image is identical to all the device or drives architecture and contents. The second key parameter is hashing, the professionals in Digital forensics should use hashing algorithms, like MD5 and SHA1, to produce hash values of the original files which they use in an investigation to ensure that the pieces of evidence are not changed or modified during the investigation, pieces of evidence collection and analysis so they protect their integrity. Another reason for using hash values is that electronic pieces of evidence are shared with various parties during the investigation process like legal

professionals, law enforcement, etc. So, we need to ensure that everybody has the same copies of the pieces of evidence. Stellar forensics that we chose to explore in this study calculates hash values automatically.

Table 1 Comparison of computer forensic tools

Feature	Stellar [33]	FTK [34]	Pro discover [35]	Autopsy [36]	Encase [37]	Cyber check suit [38]
User interface	Simple & easy to use	Simple & easy to use	Simple & easy to use	Five Complicated area	Professional training	Simple & easy to use
Hashing	MD5 & SHA-1	MD5 & SHA-1, Hash Set	MD5	MD5, MD5sum, MD5deep	MD5	MD5 & SHA-1
Apply Imaging	Yes	Yes	No	Yes	No	Yes
Recover data from formatted disks or partitions	Yes	Yes	Yes	Yes	Yes	Yes
Recover data from any disk-based device	Yes	Yes	Yes	Yes	Yes	Yes
Extension & full path	Yes	Yes	Include the file name	Yes	Yes	No
Seizer	Yes	Yes	Yes	Yes	Yes	Yes
Acquire	Yes	Yes	Yes	Yes	NO	Yes
Photo & video repair capabilities	Yes (Premium edition)	NO	NO	NO	NO	NO
Availability	Limited free edition & Commercial edition	commercial	Trial	Trial	Trial	License

B. Network forensic tools

Network forensics works through interpreting and controlling networks to make an intrusion detection and find unknown malicious and abnormal threats through networks and their associated devices. The Nmap DF tool has been discovered in network investigation in this paper as shown in Figure 8.

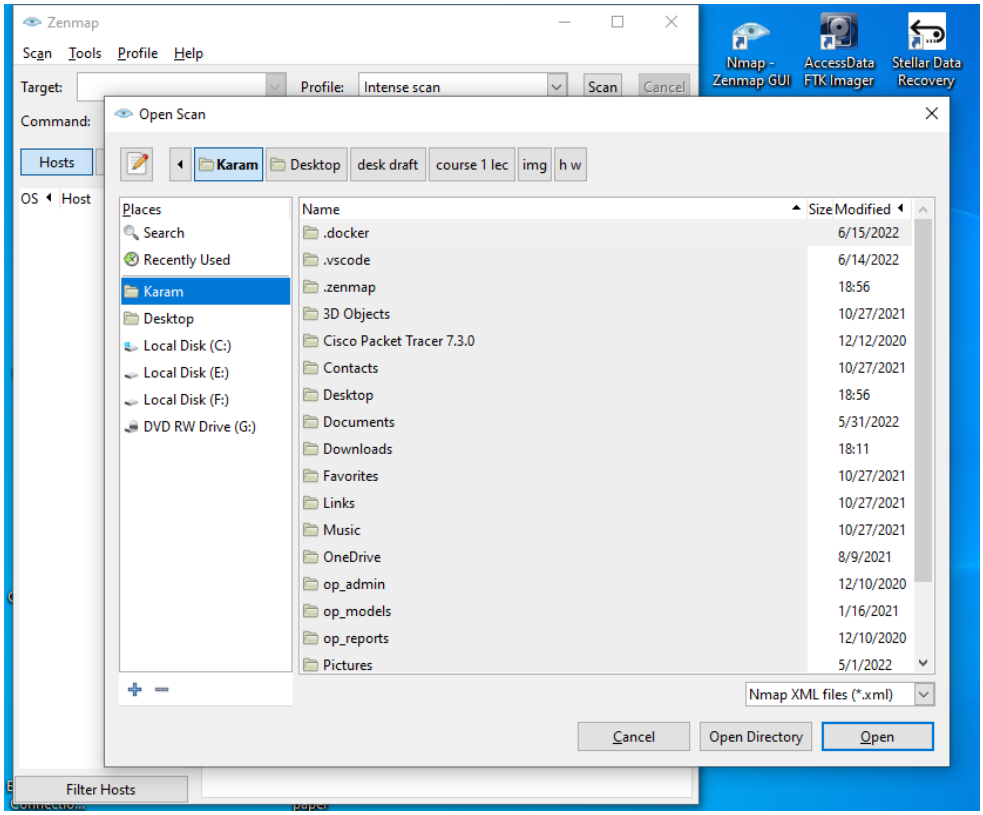

Figure 8 Nmap network analysis tool

Network Map: This tool is used to discover services and hosts on a computer network by analyzing the responses of sending packets. Nmap gives us several services for probing networks, like host discovery and O.S. detection, vulnerability detection, and other advanced services. Nmap can also work in different circumstances of the network like congestion, latency, and heavy traffic during the scan process.

Table 2 shows a comparison between five network forensic tools according to various key parameters like open port scanning which is the process of analyzing the security of all the ports in a network. It detects any vulnerabilities in packet data. Also, it analyses the parameters like topology and protocol detection and detects any spoofing or Packet spoofing or concealing the identity of the sender or impersonating another computing system.

Table 2 Comparison of network forensic tools

Tools	Nmap [39]	Wireshark [40]	Nessus [41]	Snort [42]	Ettercap [41]
Data integrity	Yes	Yes	Yes	Yes	Yes
Port scanning	Yes	No	No	Yes	No
Topology	Yes	No	Yes	No	Yes
Packet analyzing & spoofing	Yes	Yes	Yes	Yes	Yes
Protocol	Yes	Yes	Yes	Yes	Yes
Availability	Yes	Yes	Yes	Yes	Yes

C. Live Forensic Tools

Live forensic deals with active systems. It makes a forensic analysis for it and focuses on RAM attribute extraction. So, live forensics provide consistent and accurate data for investigation and it is considerably better than incomplete data provided by other DF process. In this paper, we explored the OSF mount tool as a live forensic tool.

OSF mount: It holds this name because it mounts image files that were created by disk cloning applications like OSFClone. The image file can be analyzed using OSForensics which is mounted as a virtual drive on Windows. OSFMount Forensic tool also can be used to mount DVD/CD-ROMs as RAM disks. Figure 9 show the OSF mount tool.

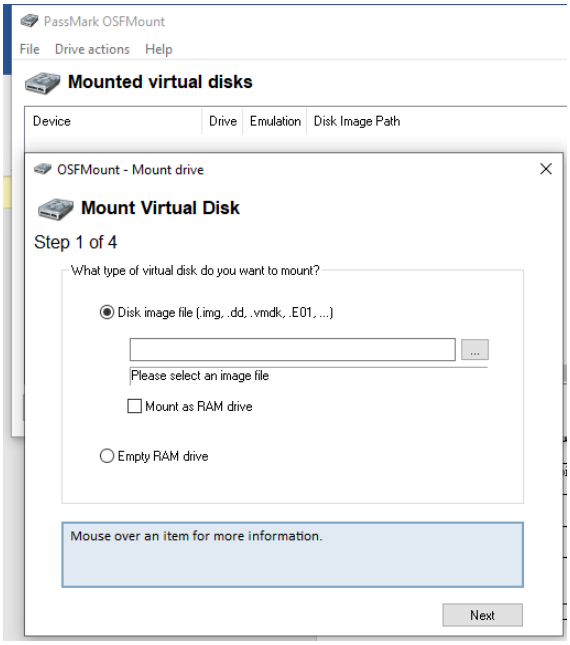

Figure 9 OSF mount live analysis tool

Four live DF tools are chosen grounded on key parameters like dealing with Search, Logs analysis, memory dumping, and Live logs analysis which is the process of taking all the content in RAM and writing it to a storage device. Table 3 shows this comparison of four live DF tools.

Table 3 Comparison of live forensic tools

Tools	OSF mount [43]	Magnet RAM [42]	Belkasoft [41]	Volatility framework [44]
Live analysis	Yes	Yes	Yes	Yes
Live logs	Yes	Yes	Yes	Yes
RAM dumping	Yes	Yes	Yes	Yes
Search	Yes	No	No	No
Logs	Yes	No	No	No
Availability	Trial	Trial	Trial	Trial

D. Other Forensic Tools

There are sub-branches of forensic tools like Database Forensic tools, O.S Forensic tools, and Email Forensic tools. These three types of tools have been explored in this section. Critical data is warehoused in various Database Management System (DBMS) i.e., Oracle as a Relational Database Management System store commercial data, MySQL work with web stores as a back-end packing, while SQLite stores personal data like SMS and browser bookmarks. So, databases need their special set of forensic tools. DF

investigators still need the necessary DF tools to investigate Database Management Systems forensic objects. Also, we require to establish a special standard for artifact storage and its mechanisms to develop advanced analysis tools for Database. [45] Operating System Forensics tools are used for recovering and gathering important information from the Operating System of the device.

The goal is to find practical proof against the criminal. Four methods are used for Operating system forensics: disk-to-disk clone, disk-to-image file, disk-to-data file, and the backup of a file. This tool identifies abnormal files and makes a hash-matching signature. In a live manner, the data has been loaded and exported with all key parameters like module, run count, title, file size, category, last run time, date, time, etc. then, the report is generated and presented to the investigator includes I/O read-write, threads, total CPU, etc. Emails played an important role in communication through the internet like business communications and transmitting information between different devices. Unfortunately, there are a lot of encounters in email DF, for example, spoofing, forged emails, and Unsigned Re-emailing. Investigator has to collect the proof, identify the criminal, and show up the judgments. It can work with various Email formats, for example, .msg, .emlx, .pdf, .mht, .xps, etc. by examining header information, message body content, and other key parameters like time. In addition, it has a filtering option, exporting, saving, and analysis. Based on popularity, four open-source (Database, email, and O.S.) forensic tools have been chosen, as shown in Table 4.

Table 4 Database, O.S and Mail forensic tools

Type	Tools	Availability
Database Forensics [45]	SQL Server	Test
	X-Ways forensic	Test
	Mandiant	Free
	Red Line	Test
O.S Forensic Tool [46], [47]	OS Forensic	Test
	ExifTool	Test
	Autopsy	Test
	Hashmy files	Test
Mail Forensic tool [48], [49]	Add4 Mail	Test
	Mail Xaminer	Test
	eMail Tracker Pro	Test
	Paraben E-mail examiner	Test

5. Digital Forensics Tools Evaluation Metrics

It is important to use different metrics to validate DF tools to help the investigator's community to compare various tools autonomously. Also, developers will recognize which part of a tool is to be enhanced. To provide extreme accuracy and fulfill the maximum requirements, metrics should cover the maximum attributes of DF Tools. Until recently a few methodologies have been proposed in this field with little research on the metrics of DF Tools. [50] proposed a solution by defining metrics to measure the number of pieces of evidence correctly produced from the list of pieces of evidence, which is called the accuracy rate, and the number of files produced by the list of files which is called the precision rate. Researchers in [51] defined seven metrics to validate DF tools as shown in Table 5.

Table 5 Digital Forensic tools metrics

Metric	Description
Absolute speediness	The time essential for the forensic tool to finish a job.
Relative speediness	The average dispensation evidence rate and reading data rate from its source.
Accurateness	The rate of right result.
Completeness	The rate of evidence originates from the collection of pieces of evidence obtainable in the forensics copy.
Reliability	counts the number of tools failing during an investigation.
Auditability	The auditability of the result.
Repeatability	The rate of tests where the procedure was utilized was exactly as specified.

The paper [52] suggested a methodology to evaluate the performance of the tool. If the reproduced pieces of evidence are identical to the original one that means the result is correct. This can be done by using a hashing algorithm like MD5. However, this method also has disadvantages for example, it does not work if one bit is lost or changed throughout the collection stage because the signatures will be changed.
Furthermore, the collection stage might not recover the precise pieces of evidence for the reason that complications like the disk being damaged are not caused by the tool.

6. The use of Artificial Intelligence in Digital Forensic tools

Digital investigators have a difficult time finding pieces of evidence in digital information. It has become difficult to specify an investigation and its source of proof. The various technology, specific procedures, and processes used in the DF investigation are not keeping up with the development of criminals. So, criminals use these weaknesses to do their crimes. Artificial intelligence (AI) is very important in identifying crime in DF investigations. An algorithm based on AI is very effective and highly recommended in detecting and preventing risks and criminal activity. Also, it is important in forecasting illegal activity. Researchers have used the available evidence data in court to condemn a person.
The pattern recognition techniques are the best for the Analysis stage of the DF. Recognition of the Pattern has two procedures. The first is an examination and the other one is recognition. The features are taken out from the patterns to be recognized in the analysis step. Then, applying different methods of pattern recognition to these features are practical for DF investigation. These techniques are projected to improve diverse DF tools to identify and gather pieces of evidence that would be cooperative to deal with explicit kinds of digital criminalities. For example, the Jaro Winkler algorithm [53]and Cosine similarity function [54] are considered advanced pattern recognition algorithms for identity resolution in DF they are typically based on making similarity metrics for more complex strings.
The increasing popularity of IoT devices and their privacy concerns encourage us to choose an IoT device traffic to analyze and make some investigations. We chose an IOT Fridge device traffic dataset from the University of New South Wales (UNSW) Canberra at Australian Defense Force Academy (ADFA). It contains six attributes (Time, date, temperature, condition, label, type) with full training set classifier. We examine this dataset using two different machine-learning techniques. We chose these techniques because they are two separate concepts. The first one is the decision stump tree which is the ML technique of a single-level of decision tree [55]. Decision stumps frequently work with apparatuses named base learners or weak learners in ML [56]. For nominal attributes, it builds a stump that has a sprig for every probable attribute rate or a stump has a double of leaves, the first one matches a specific class, and the second one has matched all the other classes [57]. The second machine learning technique we used is the Bayesian

network, it is ideal for predicting the probability of several possible known causes the occurrence of an event was the contributing factor.

As we see in Table 6, compared with Bayes Net, Decision Stump gives us less accurate results because it does not care that much about the attributes or their relationships. It focuses only on how these attributes affect the target.

Table 6 Experiments results of Decision Stump and Bayes Net

Method	DecisionStump	Bayes net
Correctly Classified Instances	536395 (91.3672 %)	582648 (99.2458 %)
Incorrectly Classified Instances	50681 (8.6328 %)	4428 (0.7542 %)
Kappa statistic	0.6722	0.9716
Mean absolute error	0.0293	0.0042
Root mean squared error	0.1211	0.038
Relative absolute error	38.6323 %	5.4697 %
Root relative squared error	62.1557 %	19.4962 %
Total Number of Instances	587076	587076

On another hand, the best results, we got from Bayes Net were because it represents a conditional dependency of a set of random variables. Each node in the network represents a variable, and each directed edge in the network represents a conditional relationship the Confusion Matrix of Bayes net is shown in Table 7.

Table 7 Confusion Matrix of Bayes net

A	B	C	D	E	F	G	< << classified as
500827	0	0	0	0	0	0	A = normal
0	35568	0	0	0	0	0	B = backdoor
0	0	9052	0	1181	0	0	C = ddos
0	0	0	7079	0	0	0	D = injection
0	0	0	0	28425	0	0	E = password
0	2902	0	0	0	0	0	F = ransomware
0	0	0	0	345	0	1697	G = xss

These results bring to light that a digital investigation and solution are important to protect the privacy of an IoT device owner. Also, it indicates the importance of artificial intelligence techniques in this field, especially machine learning techniques, and that IoT-specific concerns must be considered in the ongoing policy debate around ISP data collection and usage.

7. Challenges

This section highlights the limitations of DF tools. Four challenges in DF have been emphasized in [58]. The first one is the law of enforcement and legal systems challenges, like legal process, Jurisdiction, privacy, bad provision for legal criminal, standards, and the lack of studies on DF Tools. Second, technical challenges like big data, encryption, volatility, bandwidth crush, cloud computing, and emerging technology. Third, lack of unified formal representation, lack of forensic knowledge, standardized process, and qualified professionals in DF. Fourth, the complications of incidence response discovery, trust of audit trails, and the readiness of DF. The researchers in [17] focused on globule in the hard disk with the storage size for computers. Also, the growth of computers, cameras, and portable devices. So, we want to reconsider the DF process. [59]highlights the massive data of DF challenge, especially in the Internet of Things (IoT), and suggested a data modification process in DF by distinguishing the imaging in a massive amount of forensic data. [60] emphasized DF process limitations with cloud atmosphere like volatility,

namely records, data integrity, and creation of the forensic image. [61] presented DF process difficulties with the smartwatches.

8. Conclusion and Future work

When any digital crime or attack is done, an appropriate investigation and incident response procedure are to be used to finish the examination. The different stages of DF examination were mentioned above with a comparison of different DF Tools. These tools are chosen for the investigation according to the type of crime or attack. Artificial Intelligence (AI) acting a momentous part in prediction and analysis. This is done by different Machine Learning methods and validated with different metrics to choose the best one. The paper analyzed various tools like Computers, Networks, Databases, O.S, Live, and Mail DF Tools. In computer forensics, the Stellar tool has been chosen relatively to a comparison with other tools according to some features like imaging, hashing, recovery data, reparation capability, seizure, acquisition, and availability. In network forensic Nmap tool has been chosen according to some features like Port scanning, Packet analyzing & spoofing topology and protocol analyzing, and availability. OSF mount for the live forensic tool has been chosen according to some features in this study according to Live log analysis, RAM dumping, search, and availability. It likewise introduced the tender of AI in the DF framework. Additionally, some challenges are emphasized through supplementary the DF examination procedure. The future road of DF research should focus on the main challenges in this field like IoT forensics, Cloud DF as a service, big data, and new tools of DF. For example, determining specific data in IoT is stimulating the investigator to identify where to locate or straight the examination. Accordingly, the above challenges can consider as a research opportunity to continue in this field. As we mention before, the main problem in DF is the big forensic data, especially in network forensics and IOT forensics. So, dealing with this data in a reliable forensically manner is a big challenge in DF and this is considered a good opportunity for the researchers to innovate new techniques and new tools to deal with this big data. The researchers also can use artificial intelligence techniques with DF for example, using NLP technique for analyzing DF data and using Artificial Neural Networks (ANN) for complicated pattern recognition in various branches of DF. Also, future research should focus on developing modern techniques and tools to analyze more complicated environments, for example, cyberspace-like clouds and networks to give us the best investigation results.

References

[1] K. K. Sindhu and B. B. Meshram, "Digital Forensics and Cyber Crime Datamining," *Journal of Information Security*, vol. 03, no. 03, pp. 196–201, 2012, doi: 10.4236/jis.2012.33024.

[2] J. K. Alhassan, R. T. Oguntoye, S. Misra, A. Adewumi, R. Maskeliūnas, and R. Damaševičius, "Comparative evaluation of mobile forensic tools," in *Advances in Intelligent Systems and Computing*, 2018, vol. 721, pp. 105–114. doi: 10.1007/978-3-319-73450-7_11.

[3] O. Osho, U. L. Mohammed, N. N. Nimzing, A. A. Uduimoh, and S. Misra, "Forensic Analysis of Mobile Banking Apps," in *Computational Science and Its Applications – ICCSA 2019*, 2019, pp. 613–626.

[4] S. Sachdeva, B. L. Raina, and A. Sharma, "Analysis of Digital Forensic Tools," *J Comput Theor Nanosci*, vol. 17, no. 6, pp. 2459–2467, Sep. 2020, doi: 10.1166/jctn.2020.8916.

[5] H. Hibshi, T. Vidas, and L. Cranor, "Usability of forensics tools: A user study," in *Proceedings - 6th International Conference on IT Security Incident Management and IT Forensics, IMF 2011*, 2011, pp. 81–91. doi: 10.1109/IMF.2011.19.

[6] C.-H. Yang and P.-H. Yen, *Fast Deployment of Computer Forensics with USBs*. 2010. doi: 10.1109/BWCCA.2010.106.

[7] D. Joseph and K. Singh, "Review of Digital Forensic Models and A Proposal For Operating System Level Enhancements," *International Journal of Computer Science and Information Security 1947-5500*, vol. volume 14, pp. 797–806, Nov. 2016.

[8] J.-U. Lee and W.-Y. Soh, "Comparative analysis on integrated digital forensic tools for digital forensic investigation," *IOP Conf Ser Mater Sci Eng*, vol. 834, no. 1, p. 012034, 2020, doi: 10.1088/1757-899X/834/1/012034.

[9] A. Abirami and s Palanikumar, "Proactive Network Packet Classification Using Artificial Intelligence," 2021, pp. 169–187. doi: 10.1007/978-3-030-72236-4_7.

[10] F. Amato, G. Cozzolino, V. Moscato, and F. Moscato, "Analyse digital forensic evidences through a semantic-based methodology and NLP techniques," *Future Gener. Comput. Syst.*, vol. 98, pp. 297–307, 2019.

[11] T. Wu, F. Breitinger, and S. O'Shaughnessy, "Digital forensic tools: Recent advances and enhancing the status quo," *Digit. Investig.*, vol. 34, p. 300999, 2020.

[12] J. Cosic, C. Schlehuber, and D. Morog, "Digital Forensic Investigation Process in Railway Environment," in *2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 2021, pp. 1–6. doi: 10.1109/NTMS49979.2021.9432658.

[13] E. E.-D. Hemdan and D. H. Manjaiah, "An Efficient Digital Forensic Model for Cybercrimes Investigation in Cloud Computing," *Multimedia Tools Appl.*, vol. 80, no. 9, pp. 14255–14282, Apr. 2021, doi: 10.1007/s11042-020-10358-x.

[14] Y.-J. Jang and J. Kwak, "Digital forensics investigation methodology applicable for social network services," *Multimed Tools Appl*, vol. 74, no. 14, pp. 5029–5040, 2015, doi: 10.1007/s11042-014-2061-8.

[15] S. Costantini, G. de Gasperis, and R. Olivieri, "Digital forensics and investigations meet artificial intelligence," *Ann Math Artif Intell*, vol. 86, no. 1, pp. 193–229, 2019, doi: 10.1007/s10472-019-09632-y.

[16] A. Krivchenkov, B. Misnevs, and D. Pavlyuk, "Intelligent Methods in Digital Forensics: State of the Art," in *Lecture Notes in Networks and Systems*, 2019, pp. 274–284. doi: 10.1007/978-3-030-12450-2_26.

[17] D. Quick and K.-K. R. Choo, "Impacts of increasing volume of digital forensic data: A survey and future research challenges," *Digit Investig*, vol. 11, no. 4, pp. 273–294, 2014, doi: https://doi.org/10.1016/j.diin.2014.09.002.

[18] R. Mohammad and M. Alq, "A comparison of machine learning techniques for file system forensics analysis," *Journal of Information Security and Applications*, vol. 46, pp. 53–56, Mar. 2019, doi: 10.1016/j.jisa.2019.02.009.

[19] J. Pluskal, O. Lichtner, and O. Rysavy, "Traffic Classification and Application Identification in Network Forensics," in *Advances in Digital Forensics XIV*, 2018, pp. 161–181.

[20] O. M. K. Alhawi, J. Baldwin, and A. Dehghantanha, "Leveraging machine learning techniques for windows ransomware network traffic detection," in *Advances in Information Security*, vol. 70, Springer New York LLC, 2018, pp. 93–106. doi: 10.1007/978-3-319-73951-9_5.

[21] S. Srinivasan, V. Ravi, M. Alazab, S. Ketha, A. Al-Zoubi, and S. Padannayil, "Spam Emails Detection Based on Distributed Word Embedding with Deep Learning," 2020, pp. 161–189. doi: 10.1007/978-3-030-57024-8_7.

[22] S. Sachdeva and A. Ali, "Machine learning with digital forensics for attack classification in cloud network environment," *International Journal of System Assurance Engineering and Management*, vol. 13, no. 1, pp. 156–165, 2022, doi: 10.1007/s13198-021-01323-4.

[23] I. Sarker, "Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects," *Annals of Data Science*, pp. 1–26, Sep. 2022, doi: 10.1007/s40745-022-00444-2.

[24] A. R. Jadhao and A. J. Agrawal, "A Digital Forensics Investigation Model for Social Networking Site," in *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, 2016. doi: 10.1145/2905055.2905346.

[25] R. Montasari, "A standardised data acquisition process model for digital forensic investigations," 2017.

[26] I. Kigwana, V. R. Kebande, and H. S. Venter, "A proposed digital forensic investigation framework for an eGovernment structure for Uganda," in *2017 IST-Africa Week Conference (IST-Africa)*, 2017, pp. 1–8. doi: 10.23919/ISTAFRICA.2017.8102348.

[27] A. Singh, I. Adeyemi, and H. Venter, "Digital Forensic Readiness Framework for Ransomware Investigation: 10th International EAI Conference, ICDF2C 2018, New Orleans, LA, USA, September 10–12, 2018, Proceedings," 2019, pp. 91–105. doi: 10.1007/978-3-030-05487-8_5.

[28] D. Sun, X. Zhang, K.-K. R. Choo, L. Hu, and F. Wang, "NLP-based digital forensic investigation platform for online communications," *Comput Secur*, vol. 104, p. 102210, 2021, doi: https://doi.org/10.1016/j.cose.2021.102210.

[29] X. Du, N.-A. Le-Khac, and M. Scanlon, *Evaluation of Digital Forensic Process Models with Respect to Digital Forensics as a Service*. 2017.

[30] M. Scanlon, "Battling the digital forensic backlog through data deduplication," in *2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, 2016, pp. 10–14. doi: 10.1109/INTECH.2016.7845139.

[31] M. D. Kohn, M. M. Eloff, and J. H. P. Eloff, "Integrated Digital Forensic Process Model," *Comput. Secur.*, vol. 38, pp. 103–115, Oct. 2013, doi: 10.1016/j.cose.2013.05.001.

[32] B. Hitchcock, N.-A. Le-Khac, and M. Scanlon, "Tiered forensic methodology model for Digital Field Triage by non-digital evidence specialists," *Digit Investig*, vol. 16, pp. S75–S85, 2016, doi: https://doi.org/10.1016/j.diin.2016.01.010.

[33] C. Dietzel, T. U. Berlin, / De-Cix, M. Wichtlhuber, G. Smaragdakis, and A. Feldmann, "Stellar: Network Attack Mitigation using Advanced Blackholing."

[34] DHS, "Access Data Forensic Toolkit (FTK) Version 7.0.0.163: Test Results for String Search Tool." [Online]. Available: https://www.dhs.gov/science-and-

[35] K. M. A. Kamal, M. Alfadel, and M. S. Munia, "Memory forensics tools: Comparing processing time and left artifacts on volatile memory," in *2016 International Workshop on Computational Intelligence (IWCI)*, 2016, pp. 84–90. doi: 10.1109/IWCI.2016.7860344.

[36] A. Dizdarević, S. Baraković, and J. Baraković Husić, "Examination of Digital Forensics Software Tools Performance: Open or Not?," in *Advanced Technologies, Systems, and Applications IV -Proceedings of the International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies (IAT 2019)*, 2020, pp. 442–451.
[37] D. Quick and K.-K. R. Choo, "Digital Forensic Data Reduction by Selective Imaging," 2018, pp. 69–92. doi: 10.1007/978-981-10-7763-0_4.
[38] A. Shaaban and N. Abdelbaki, "Comparison study of digital forensics analysis techniques," in *Procedia Computer Science*, 2018, vol. 141, pp. 545–551. doi: 10.1016/j.procs.2018.10.128.
[39] N. Agrawal and S. Tapaswi, "Wireless Rogue Access Point Detection Using Shadow Honeynet," *Wirel Pers Commun*, vol. 83, no. 1, pp. 551–570, 2015, doi: 10.1007/s11277-015-2408-0.
[40] P. S. Kenkre, A. Pai, and L. Colaco, "Real Time Intrusion Detection and Prevention System," in *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*, 2015, pp. 405–411.
[41] N. A. Hassan, "Computer Forensics Lab Requirements," in *Digital Forensics Basics: A Practical Guide Using Windows OS*, N. A. Hassan, Ed. Berkeley, CA: Apress, 2019, pp. 69–91. doi: 10.1007/978-1-4842-3838-7_3.
[42] A. Ghafarian and C. Wood, "Forensics Data Recovery of Skype Communication from Physical Memory," in *Intelligent Computing*, 2019, pp. 995–1009.
[43] N. A. Hassan, *Digital Forensics Basics*. Apress, 2019. doi: 10.1007/978-1-4842-3838-7.
[44] J. Seo, S. Lee, and T. Shon, "A study on memory dump analysis based on digital forensic tools," *Peer Peer Netw Appl*, vol. 8, no. 4, pp. 694–703, 2015, doi: 10.1007/s12083-013-0217-3.
[45] J. Wagner, A. Rasin, K. Heart, R. Jacob, and J. Grier, "DB3F & DF-Toolkit: The Database Forensic File Format and the Database Forensic Toolkit," *Digit Investig*, vol. 29, pp. S42–S50, 2019, doi: https://doi.org/10.1016/j.diin.2019.04.010.
[46] S. Alqahtany, N. Clarke, S. Furnell, and C. Reich, "A Forensic Acquisition and Analysis System for IaaS: Architectural Model and Experiment," in *2016 11th International Conference on Availability, Reliability and Security (ARES)*, 2016, pp. 345–354. doi: 10.1109/ARES.2016.58.
[47] S. C. Clark, J. R. Gill, S. S. Aiken, J. L. Arden, S. Denton, and D. McNally Executive Director, "Forensic Autopsy Performance Standards The National Association of Medical Examiners 2020 Standards Committee President Vice-President Chair of the Board Secretary-Treasurer Executive Offices," Online, 2010.
[48] G. Chhabra, C. Professor, D. Singh, and B. Professor, "Review of E-mail System, Security Protocols and Email Forensics," *International Journal of Computer Science & Communication Networks*, vol. 5, pp. 201–211, Jan. 2015.
[49] V. Singh, "Forensic Investigation of Email ARTEFACTS by using various Tools," vol. 2, pp. 2321–2613, Jan. 2015.
[50] L. Pan and L. M. Batten, "An Effective and Efficient Testing Methodology for Correctness Testing for File Recovery Tools," in *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007)*, 2007, vol. 2, pp. 103–107. doi: 10.1109/IIH-MSP.2007.78.
[51] D. Ayers, "A second generation computer forensic analysis system," *Digit Investig*, vol. 6, pp. S34–S42, 2009, doi: https://doi.org/10.1016/j.diin.2009.06.013.
[52] L. Pan and L. M. Batten, "Robust performance testing for digital forensic tools," *Digit Investig*, vol. 6, no. 1, pp. 71–81, 2009, doi: https://doi.org/10.1016/j.diin.2009.02.003.
[53] M. Bilenko, R. Mooney, W. Cohen, P. Ravikumar, and S. Fienberg, "Adaptive name matching in information integration," *IEEE Intell Syst*, vol. 18, no. 5, pp. 16–23, 2003, doi: 10.1109/MIS.2003.1234765.
[54] W. W. Cohen, P. Ravikumar, and S. E. Fienberg, "A Comparison of String Distance Metrics for Name-Matching Tasks." [Online]. Available: www.aaai.org
[55] L. Reyzin and R. E. Schapire, "How Boosting the Margin Can Also Boost Classifier Complexity," in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 753–760. doi: 10.1145/1143844.1143939.
[56] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int J Comput Vis*, vol. 57, no. 2, pp. 137–154, 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.
[57] J. J. Oliver and D. Hand, "Averaging over decision stumps," in *Machine Learning: ECML-94*, 1994, pp. 231–241.
[58] N. Karie and H. s Venter, "Taxonomy of Challenges for Digital Forensics," *J Forensic Sci*, vol. 60, Jul. 2015, doi: 10.1111/1556-4029.12809.
[59] D. Quick and K.-K. R. Choo, "Big forensic data reduction: digital forensic images and electronic evidence," *Cluster Comput*, vol. 19, no. 2, pp. 723–740, 2016, doi: 10.1007/s10586-016-0553-1.
[60] S. Hraiz, "Challenges of digital forensic investigation in cloud computing," in *2017 8th International Conference on Information Technology (ICIT)*, 2017, pp. 568–571. doi: 10.1109/ICITECH.2017.8080060.
[61] S. Alabdulsalam, K. Schaefer, T. Kechadi, and N.-A. Le-Khac, "Internet of Things Forensics – Challenges and a Case Study," in *Advances in Digital Forensics XIV*, 2018, pp. 35–48.