Digital evidence extraction and analysis of Disk Image

Sai Pavan Rangu
School of Computer and Cybersecurity Engineering
Amrita Vishwa Vidyapeetham
Chennai, India
Email: rangusaipavan@gmail.com

Tata seetha rama Sai Vamshi krishna
School of Computer and Cybersecurity Engineering
Amrita Vishwa Vidyapeetham
Chennai, India
Email: tatasaivamsi1@gmail.com

Abhiram Patel U
School of Computer and Cybersecurity Engineering
Amrita Vishwa Vidyapeetham
Chennai, India
Email:uppulaabhirampatel2601@gmail.com

***Abstract*—In today's investigations concerning cybercrime, virus detection, and legal actions, the forensic analysis of digital evidence—especially from disk images—is essential. One essential tool in this field is the Disk Image Analyser Tool (DIAT), which gives investigators the capacity to examine disk images in great detail, find buried information, recognize malicious code, and provide evidence that is useful in court. The usefulness, approaches, and importance of DIAT in forensic analysis are all thoroughly examined in this work. It talks about the key components of DIAT, such as how to parse disk pictures, and partitions, pull out metadata, find anomalies, Database Tables and make investigative procedures easier. This work also emphasizes the useful uses of DIAT in forensic contexts, including malware signature identification, file system artifact analysis, and digital timeline reconstruction. The effectiveness and drawbacks of DIAT are evaluated via a study of case studies and experimental assessments, providing insight into how it might improve forensic investigations and progress digital forensic techniques. Additionally, some improvements and future approaches for DIAT are suggested to enhance its capacity to handle the changing demands of digital forensic investigation.**

Keywords: Disk image analyse, forensics, disk image, data hiding, Digital Evidence, evidence extraction, extraction tools, file sorting, Keyword search.

I. Introduction

Specializing in the methodical extraction, examination, and exposition of digital evidence for legal processes, digital forensics is an indispensable field in contemporary investigations. Identifying illegal activity, evaluating its effect on system integrity, gathering proof for legal purposes, and preventing future security breaches are just a few of the goals that fall under this subject. Digital data gathering, preservation, and analysis are the three main steps that this procedure usually entails. This study emphasizes methods applicable to a variety of file systems frequently seen in forensic investigations, with an emphasis on the collection and examination of digital evidence linked to breaches within computer systems. Although file systems are widely used in operating systems, there is a significant lack of information in the literature on the vulnerabilities that are present in file systems and the constraints of current digital forensic techniques. To clarify disk image structure and vulnerabilities, analyse popular digital forensic techniques and suggest improvements for static analysis methodologies used with disk images, this study seeks to close this knowledge gap. Researchers want to improve knowledge of file system vulnerabilities and improve digital forensic methods through this study to better prevent cybercrimes and safeguard digital ecosystems.

II. Digital Forensic Proof

A. Types of Digital Proof

Volatile and non-volatile digital evidence are the two types that are needed for digital forensics. erratic evidence is stored in digital devices' Random Access Memory (RAM) and can only be gathered while the devices are in use. Non-volatile evidence may be obtained even while the devices were off the power supply. Non-volatile evidence sources include data hidden in files, files used for swapping, index-type data files (index.dat files), clusters that are not allocated, unused partitions, hidden partitions, saved record settings, and event-related logs.

B. The value of electronic proof

Large amounts of information are kept automatically. Electronic devices log everything that a user does, including file transfers, access histories, browsing habits, user inputs, and passwords. This degree of tracking makes the related data extremely valuable.

C. Extracting evidence

Tracing digital evidence for both sorts of crimes and vandalism begins with evidence extraction. A predetermined series of procedures is followed during the evidence extraction process to capture activity traces for additional examination and accusations. It varies depending on the evidence's purpose and might be either logical or physical. Similar to secondary storage in a computer system, physical extraction is locating and obtaining evidence from a digital source. This is typically accomplished utilizing header and footer-based searches, keyword-based comparison, and space that hasn't yet been assigned. Gathering proof from the point of storage concerning the file system of the underlying operating system is a necessary step in the logical extraction process. Based on

the connected device, file management system, active apps, and underlying platform, a logical extraction is carried out. Obtaining information from current and deleted files, file systems, unused space, compressed, protected, and genuine data are also included.

D. Elements impacting the extraction of evidence

The fast evolution and proliferation of digital devices, budgetary concerns, inadequate training opportunities, a large volume of unprocessed digital evidence, equipment scarcity, and the potential for examiners to leave the team and achieve unachievable results are some of the factors that impact or influence evidence extraction. One of the main factors influencing evidence extraction is also the lack of consensus regarding the necessity of digital evidence acquisition.

E. Digital evidence's admissibility in court

Extracts in some of the Evidence Acts from various jurisdictions expressly provide for the application of such provisions in the proof of the contents of electronic records. According to the provisions, any information contained in an electronic record—whether it be a document or correspondence printed, stored, recorded, or transferred in any medium by a computer—is considered documentary evidence and may be used as evidence in lieu of other proof of the actuals' production, provided the requirements outlined in them are met.

III. Forensic Analysis Process

A. Disk Image Acquisition

The first was to acquire the disk image. Through a user-friendly interface built with PyQt5, the user selects the disk image file. A read and prepare the disk image for analysis follows. At this stage, an exact copy of the disk image is available for further processing to support data integrity.

B. Case Management

The case directory is the second stage, which manages and maintains all data regarding forensic investigations. This keeps every investigation in line and maintains the integrity of the data. The directory is the central storage of files, findings, and reports relating to that case.

C. Extracting Information About the Disk Image

At this stage, information from the disk image is extracted, which includes, but is not limited to, information about the image size and information about partitions. This information helps in understanding the structure of the disk image; it will later help in pinpointing areas that need closer scrutiny.

D. Evidence Searching

The fourth stage involves searching for evidence of malicious activities on the duplicate image. It is done by scanning for signatures, time stamps, or other hidden data. The evidence may be searched for by using many customized developed methods on the duplicated image to ensure that the same is completely probed for evidence. This would lead to the detection of possible security breaches, unauthorized access, and other such malicious activities.

E. Database Integration

In the entire process, it is integrated with an SQLite database for storing and managing data. The database design has been done to make any information stored and retrieved effectively. This integration would ensure that all the data is safely stored and can be efficiently questioned while analysis is performed. It keeps track of records of the file information, findings, and other relevant data in the database for structured data management

F. Generation of reports

In the end, it generates a complete report of the findings after the tool has made an analysis. The report details the case, the information image of the disk, and the result of searching for evidence. This report is very important in the case documentation of the results of the analysis and for forensic research. It produces the investigation clearly and concisely, thus helping to present the findings to relevant parties

IV. Literature survey

Evidence extraction from hard drives, mobile phones, and social media renders enormous help in establishing the act performed with malicious/criminal intent upon cyberspace resources. Kailash Kumar et al in their work titled "Hard drive identification and analysis in digital forensics" highlight the process of acquiring the necessary evidence from a hard disk drive to support a lawsuit[1]. The work implements the evidence-gathering process from hard drives after their identification through BIOS information, through the steps of

(I) acquiring data from the hard disk through the forensic data acquisition and analysis tool EnCase Forensic Edition Version 4 a Windows-based one, without losing integrity upon the files prospectively created and stored by the individual who committed the crime.

(ii) authenticating collected evidence through the creation of MD5 hash value, used similarly to fingerprint identification during imaging of a hard drive.

(iii) Evidence analysis through the creation of MD5 hash value for the extracted file to check mismatch of content in the future. This step involves powering down the system under suspicion, and forensically Imaging the Drive with an EnCase DOS Boot Disk for further investigation.

(iv) According to Det. Cindy Murphy, the following procedures could be used to retrieve data from mobile devices[10] They are:

(I) the Evidence Intake phase, which covers the ownership chain, the proof of custody, the kind of incident to which the device was connected, and a broad description of the kinds of data that should have been obtained from the phone.

(ii) The identification phase addresses the following topics: the phone's make, model, and important details; removable and secondary data storage; legal authority to work on the device; and other parts of prospective evidence, if any

(iii) This includes researching the phone to be tested, preparing the devices to be used in the examination, and making sure

the examination machine is ready with all hardware, software, cables, and drivers.

(iv) Isolation phase: This is a phase where the phone is isolated from any possible sources of communication before it is examined. It avoids the appending of new data to the phone via calls and text messages received and erasure of data via a signal or an overwrite of available data by virtue of calls and text messages. Isolation is what prevents the examiner from instinctive access to voicemail, e-mail, browsing history, or whatever is possibly kept with the service provider's connectivity.

(v) Either processing stage to obtain data residing on this primary storage, by removing installed data storage/memory cards for processing separately using established computer forensics methods to ensure that date and time information of files maintained on the data storage/memory card is not changed during the investigation.

(vi) Verification phase to establish accuracy of information gathered through comparison of gathered data to the mobile, usage of tools and comparing the results, and applying hash values for matching.

(vii) Documentation and reporting phase, this involves the start date and time of when the examination was started, the description of the physical status of images of phone and components for example SIM card and memory expansion card. The label with key information of on and off status of phone during receipt, Make model and tools used as part of examination data documented while the examination is conducted.

(viii) Presentation phase is concerned with the way the information is extracted, recorded, and provided to another investigator, lawyer, and court. The beneficiaries should also receive the extracted and consumed data in hard copy and electronic formats. This data is used to arrange call history and other related information, or it can be fed into other software to facilitate future work. Data extracted from images or alternative data formats could be provided for better understanding.

(ix) It becomes very important that the cellular phone preserves and documents the extracted data in a format that will be useful in support of the extraction process for the court proceeding, future reference purposes, and record-keeping needs.

The procedure of extracting evidence from WhatsApp that guarantees the confidentiality of messages sent and received by altering the encryption strategy of the SQLite database stored on the installed device's memory has been the focus of Nagendar Rao Koppolu [4]. Because so many people use the social networking app, it is regarded as one of the most appropriate and important venues for gathering evidence.

V. System design of Disc Forensics

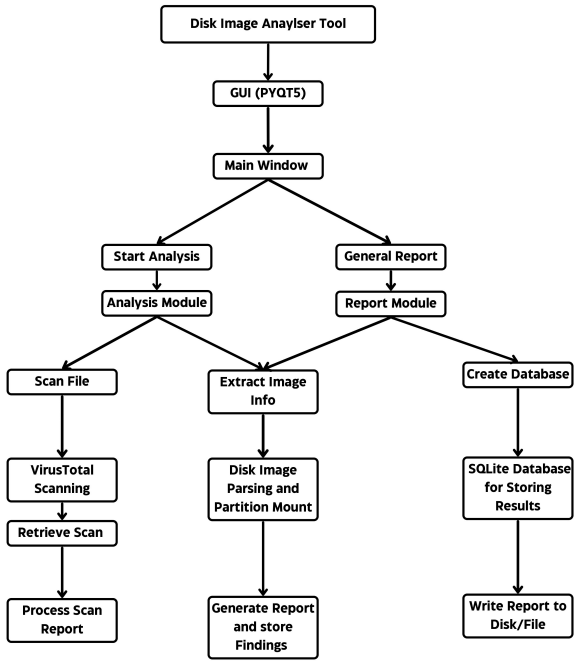

Fig. 1. Design of the project

User Interface:
Main Window: This will be the basic interface from which all users can initiate and manage the analysis.
Start Analysis: Button or any kind of feature that will start disk image analysis.
Report Module: This will be the module acting as a frontend for report generation/viewing.
General Report: This will be a sub-section under the user interface that will present general findings and results.
Scan File: It scans files inside the Disk Image.
Extract Image Info: It extracts thorough details from the Disk Image.
Create Database: This starts an SQLite database and creates it to keep results from the analysis, meaning it holds data from Virus and malware findings.
Virustotal scanning: This helps to scan files against known malware signatures appended at Virustotal.
Disk Image Parsing and Partition Mounting: Disk image is parsed to identify and mount partitions to enable further detailed analysis.
Database Manager (SQLite):
Store Results: Store results from the scans and analysis.
Retrieve Results of Scans: It retrieves the stored results of the scans into the program for further processing in report generation.
Report Generator:
Process Scan Report: Processes the retrieved data of the scans.
Generate Report: This module compiles the gathered information into a detailed report.

VI. Architecture of Forensic Disk Image Analysis

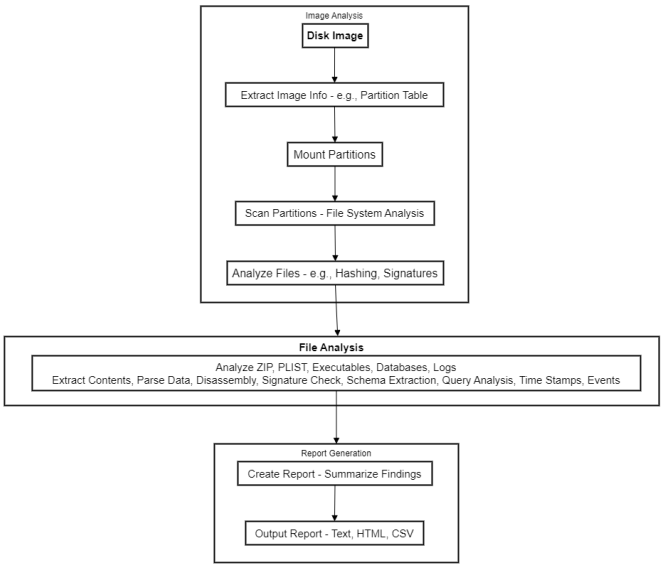

Fig. 2. Architecture Diagram

A. Image Acquisition Layer

Bit-for-bit copy of the storage device is acquired in the first layer. This layer ensures the integrity of the original evidence is maintained by creating an exact duplicate for analysis. The acquired disk image should be a replica of the original, inclusive of all sectors, partitions, and unallocated spaces, in a manner that no data is forgotten in the analysis done later.

B. Image Metadata Extraction Layer

This is the layer responsible for the examination of the metadata in the case disk image. In most cases, important information such as the partition table, among other structural information about the disk, is extracted here and is always paramount in giving insight into the structure and organization of the data. This step gives the foundational context necessary for partition mounting and file system analysis with a high degree of accuracy, ensuring that all data structures are correctly interpreted.

C. Partition Mounting Layer

In this layer, the identified partitions mount as virtual drives. This is essential for making the underlying file systems accessible for more detailed inspection and analysis. Mounting partitions enables forensic tools to interact with file systems, appearing active and live, thereby availing comprehensive access to stored data, file structures, and system configurations.

D. File System Analysis Layer

The file system analysis layer allows the detailed examination of mounted partitions to be perused. The perusal in this layer then screens the file systems for files and directories that bear relevant metadata, thus being able to overview data contained in the partitions. It makes sure that all files, including hidden and system files, are accounted for and that all their attributes and properties are documented for future analysis.

E. File Analysis Layer

This layer checks every file located in the partitions. Multiple methods, such as hashing and signature matching, add to different facets of data carving to ensure file integrity, identify their type, and attempt to recover deleted or fragmented files. Such a detailed investigation thus forms a structure in which files can be identified and categorized, anomalies detected, and potentially important information extracted that may not be visible in any other way.

F. Specific File Type Analysis Layer

During this level, specific types of files that generally pertain to forensic investigations are carefully dissected. Of these are the contents of compressed files—such as ZIP—PLIST, executable files, databases, and log files—all of these are subjected to their own specific type of scrutiny in terms of extracting information relative to the case, interpreting the meaning, and locating encrypted or encoded information which may be key to the case at hand.

G. Reporting and Documentation Layer

This layer summarizes the findings of the analysis to structured reports. It contains timelines of events, summaries of evidence, and detailed descriptions of techniques used during the analysis with their corresponding results. Such a layer will document all the steps that were undertaken while laying the analysis; for this reason, it can easily be reproducible and verifiable by other experts in the forensic field.

H. Report Dissemination Layer

The last layer ensures the results of the analysis are effectively communicated to the stakeholders. Reports can be derived in various forms like text documents, HTML reports, and CSVs, ensuring easy distribution, collaboration, and the continuation of analysis. The process of dissemination would help in reporting the findings in a very lucid and understandable format so that lawyers, investigators, and all other concerned people find it easier to analyze and make use of the same.

VII. Methods and Procedures

A. Project Planning and Requirements Gathering

Define objectives for the Autopsy forensic tool, like analysing disk images, identifying malicious files, and generating comprehensive reports.

Requirements Analysis: Identification and Documenting—All functional and non-functional requirements are identified and documented. Documentation includes user interface design, database schema, core functionalities of image analysis, mounting partitions, scanning for evidence, etc.

B. Design and Architecture

System Architecture: Detail the overall architecture of the application, specifying important components, as well as the main parts to be implemented: user interface, database, core forensic analysis modules. User Interface Design. Propose wireframes and prototype the user interface in Figma or Adobe XD. Ensure that the design is as user-friendly and intuitive as possible. Database Design: Propose a database schema in SQLite for storing information related to files and their findings. Implementation Setup Development Environment: Set up the development environment with all the necessary libraries and frameworks; for example, PyQt5 for the GUI and SQLite for the database. User Interface Development: Next, implement all GUI elements using PyQt5. This will create the main window, input fields for image path and case name, and buttons for actions. Core Functionality Development: Implement the core functionalities. The following are some examples: Image Path Selection: As shown in, provide a facility to the user for selecting the path of the disk image.

C. Case Management

Provide the user the ability to create cases and be able to manage them. Image Analysis: Develop methods to extract information from the disk image and analyse such information. Database Integration: Integrate SQLite for storing and retrieving data. Report Generation: Add functionality for creating, viewing, and exporting reports

D. Deployment

Packaging—The product is packaged for distribution. This may be in the form of an installer or just a distribution of the application as an executable that users can execute on their own. 5.0 Deployment—The application will be deployed to the target environment, or environments, and access provided to users.

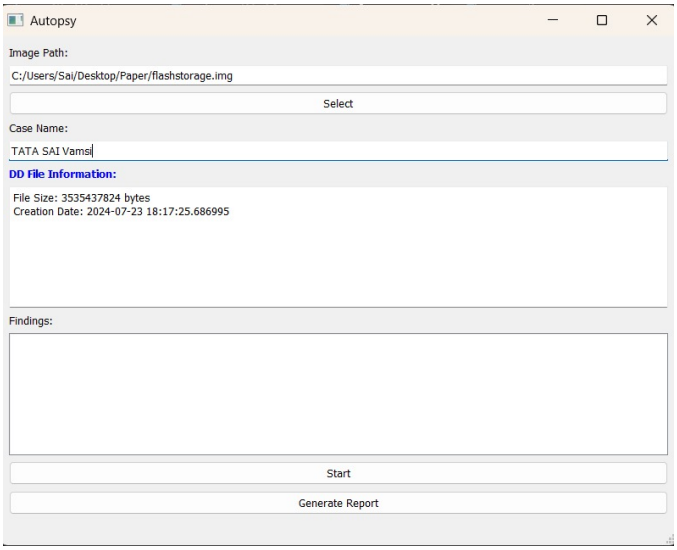

Fig. 3. Initial interface

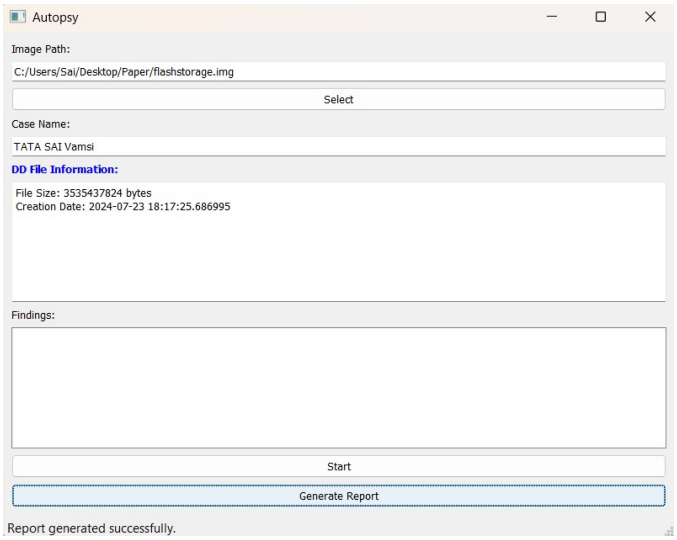

Fig. 4. Image Details

E. Maintenance and Support

Bug Fixes: Address any bugs or issues reported by users Updates: Apply updates and improvements based on feedback and changing requirements from the user. Support: Help the users with any hiccups they may encounter in using the application

VIII. Implementation Details

A. Key Classes and Functions

AutopsyMainWindow: This class runs the main GUI window.
selectimagepath: This function picks the image path.
extractimageinfo: This function pulls data from the disk image.
createdatabase: This function sets up the SQLite database.
generatereport: This function crafts the autopsy report.

B. Libraries and Tools

Python: This language powers the whole thing.
PyQt5: This builds the fancy user interface.
SQLite: This handles all the database stuff.
Requests: This grabs data from the web (if needed).
Zipfile, Hashlib, Datetime, and friends: These tackle file squishing, number crunching, and time tracking.

C. Analysis and Graph

1) Description of the Table: Case: This column lists the different cases that were analysed.
Malicious Files Found: This column shows the number of malicious files detected in each case.

Case	Malicious Files Found
Case 1	5
Case 2	2
Case 3	7

TABLE I
MALICIOUS FILES FOUND IN DIFFERENT CASES

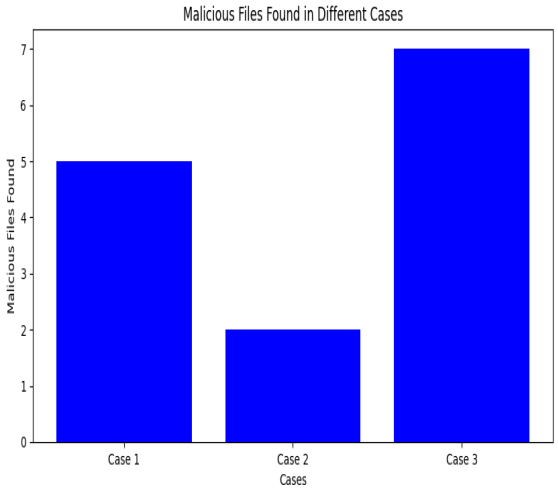

Fig. 5. Malicious found in the different cases

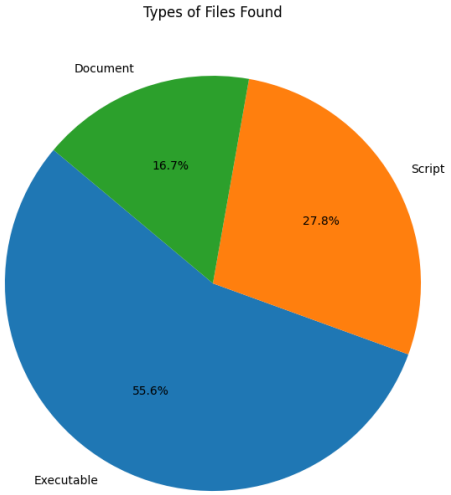

Fig. 6. Types of files that are mostly to have vulnerabilities and malicious code with them and that are found in the tests.

IX. CONCLUSION

We had to work systematically in developing the Autopsy forensic tool to get a robust and user-friendly application. We have been in a position to come up with a complete tool that really helped in the digital forensic investigation through meticulously planning and executing each phase of the process. It is a very important and useful tool that allows forensic analysts to acquire disk images, manage cases, obtain detailed information about a disk image, find evidence, integrate with a database, and finally generate exhaustive reports. This project assessed the need for a systematic approach toward forensic tool development to ensure its accuracy, efficiency, and reliability. Details of the process are included here for reference in similar projects in order to elaborate on methodologies and best practices.

X. FUTURE WORK

Though effective, there are avenues in the Autopsy forensic tool that could be improved, even open to further research. The following is investigated: Higher level of automation: The forensic analysis process is time-consuming and error-prone; increasing the level of automation in this process could be time-saving. It can involve automated scanning for more types of evidence or integration with machine learning algorithms for identifying patterns indicative of malicious activity.

Additional File System Support: Expanding the tool's compatibility to other file systems beyond NTFS, such as EXT4, HFS+, APFS, would increase the applicability and usefulness of the tool in different forensic scenarios.

Better User Interface: There's always space for improvement, even if the existing interface is quite user-friendly. GUI could be improved so that it has more intuitive navigation and better visualization of data for better user experience and efficiency.

Cloud Integration: Given the increasing trend of using the cloud, it is desirable to integrate capabilities for the forensic analysis of cloud-based data. It would consist of developing methods for securely accessing and analysing data stored in a variety of cloud environments.

Forensic analysis in real time: Forensic analysis in real time capabilities would be very vital for incident response. It would help forensic analysts track and analyse data as it is generated, immediate insights into active investigations, and decision-making.

These areas, when worked on in the future, will further the progression of Autopsy as a forensic tool and keep it moving with the ever-changing nature of digital forensic investigations. These developments not only enhance the tool's functionality but also ensure its applicability and efficiency for the fast-moving high-tech world.

REFERENCES

[1] Kailash Kumar Hard drive identification and analysis in digital forensics
[2] Muhammad Iqbal; Benfano Soewito, "Digital Forensics on Solid State Drive (SSD) with TRIM Feature Enabled and Deep Freeze Configuration Using Static Forensic Methods and ACPO Framework", International Journal of Computer Science and Information Security (IJCSIS), Vol. 18, No. 11, November 2020, pp. 44-56.
[3] Aya Fukami, Radina Stoykova, Zeno Geradts, "A new model for forensic data extraction from encrypted mobile devices", Forensic Science International: Digital Investigation, Vol. 38, September 2021, Article 301169.

[4] Nagendar Rao Koppolu, “A Deep-dive Analysis on WhatsApp Artifacts and their Relevance in Crime Investigation”, International Research Journal of Engineering and Technology (IRJET), Vol. 08, Issue 06, June 2021.

[5] Yash Gorasiya, “Types and Sources of Digital Evidence”, DFI Part - 3, Cyversity, June 21, 2021. [Online]. Available: https://medium.com/cyversity/types-and-sources-of-digital-evidence-b8fb1f64060f

[6] Tushar Panhalkar, “Understanding Digital Evidence and Its Types”. [Online]. Available: https://info-savvy.com/understanding-digital-evidence-and-its-types/

[7] “Why electronic evidence is important and how to prove its authenticity”, LifeHash. [Online]. Available: https://www.lifehash.com/post/why-electronic-evidence-is-important-and-how-to-prove-its-authenticity

[8] Ajay Bhargava, Aseem Chaturvedi, Karan Gupta, Shivank Diddi, “Use Of Electronic Evidence In Judicial Proceedings”, Mondaq, June 01, 2020. [Online]. Available: https://www.mondaq.com/india/trials-appeals-compensation/944810/use-of-electronic-evidence-in-judicial-proceedings

[9] Tushar Panhalkar, “List of Mobile Forensics Tools”. [Online]. Available: https://info-savvy.com/list-of-mobile-forensic-tools/

[10] Det. Cindy Murphy, “Cell Phone Evidence Extraction Process Development”, Digital Forensics Magazine. [Online]. Available: https://digitalforensicsmagazine.com/blogs/wp-content/uploads/2010/07/Cell-Phone-Evidence-Extraction-Process-Development-1.8.pdf

[11] Oxygen Forensics, “WhatsApp Challenges: Finding Evidence With Oxygen Forensic Detective”, Forensic Focus, January 30, 2019. [Online]. Available: https://www.forensicfocus.com/news/whatsapp-challenges-finding-evidence-with-oxygen-forensic-detective/

[12] NowSecure, “Android Forensics”. [Online]. Available: https://github.com/nowsecure/android-forensics

[13] Exterro, “FTK® Imager”. [Online]. Available: https://www.exterro.com/ftk-imager

[14] KD8BNY, “LiME Linux Memory Extractor”, GitHub, March 21, 2021. [Online]. Available: https://github.com/504ensicsLabs/LiME

[15] HancomGMD, “MD-NEXT, Mobile forensic software for data extraction”. [Online]. Available: http://www.hancomgmd.com/ir/press/md-next-mobile-forensic-software-for-data-extraction/

[16] “Overview of MOBILedit Forensic Express product features and functions”. [Online]. Available: https://www.mobiledit.com/forensic-express

[17] B. Carrier, “File system forensic analysis”, Addison-Wesley Professional, USA, 2008.

[18] S. Ardisson, “Producing a Forensic Image of Your Client’s Hard Drive? What You Need to Know”, Qubit, vol. 1, pp. 1-2, 2007.

[19] M. Andrew, “Defining a Process Model for Forensic Analysis of Digital Devices and Storage Media”, Proceedings of SADFE2007, Second International Workshop on Systematic Approaches to Digital Forensic Engineering, pp. 16-30, 2007.

[20] E Investigation, “Electronic Crime Scene Investigation: A Guide for First Responders”, US Department of Justice, NCJ, 2001.

[21] M. Reith, C. Carr, G. Gunsch, “An examination of digital forensic models”, International Journal of Digital Evidence, vol. 1, pp. 1-12, 2002.

[22] K. Bejtlich C. Rose, ”Real digital forensics: computer security and incident response”, Addison-Wesley Professional, USA, 2008.

[23] H. Carvey, ”Windows Forensic Analysis DVD Toolkit”, Syngress Press, USA, 2007.

[24] L. Naiqi, W. Yujie H. QinKe, ”Computer Forensics Research and Implementation Based on NTFS File System”, CCCM’08, ISECS International Colloquium on Computing, Communication, Control, and Management, 2008.

[25] J. Aquilina, E. Casey C. Malin, ”Malware Forensics Investigating and Analyzing Malicious Code”, Syngress Publishing, USA, 2008.

[26] E. Huebner, D. Bem C. Wee, ”Data hiding in the NTFS file system”, Digital Investigation, Elsevier, 2006, vol. 3, pp. 211-226.

[27] S. Hart, J. Ashcroft D. Daniels, ”Forensic examination of digital evidence: a guide for law enforcement”, National Institute of Justice NIJ-US, Washington DC, USA, Tech. Rep. NCJ, 2004.