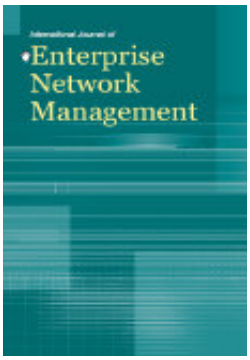


International Journal of Enterprise Network Management

ISSN online: 1748-1260 - ISSN print: 1748-1252
https://www.inderscience.com/ijenm

Effectiveness of digital forensic investigation through excavation methods of various Linux based tools

T.M. Bhraguram, P.S. Rajakumar, Arshia Arjumand Banu

DOI: 10.1504/IJENM.2024.10053188

Article History:

Received:	17 April 2022
Accepted:	20 October 2022
Published online:	19 March 2024

Effectiveness of digital forensic investigation through excavation methods of various Linux based tools

T.M. Bhraguram* and P.S. Rajakumar

Computer Science and Engineering Department,
DR. MGR Educational and Research Institute,
Chennai, India
Email: bhruguram@gmail.com
Email: rajakumar.subramanian@drmgrdu.ac.in
*Corresponding author

Arshia Arjumand Banu

College of Computer Science and Information Technology,
Jazan University,
Jazan, Kingdom of Saudi Arabia
Email: abanu@jazanu.edu.sa

Abstract: Digital forensic is a process of pre-processing, identification, modelling, extraction, and documentation of computer evidence. The forensic investigations in today's human life are more important due to the high-level cyber crime activities and other proof-less investigations happening under various public and private domains. The computer world updates various methods to do the investigation activities and most of the methods are working based on the existing activity monitoring and proof-based content available for the processing. Various computer platforms give many procedures to continue the investigation process, but the effectiveness and accuracy is completely depending on the tools and data proof used while processing the data. Linux is one of the most eligible and rich tools providing platform with various proofreading mechanisms. We are trying to furnish the most effective methods used for digital forensic investigations in Linux platform, which were proven to be with high level of accuracy and integrity. This article can provide various mechanism used in the tools and its effectiveness through an excavation method.

Keywords: digital forensic; cyber crime; excavation method; Linux platform; platform based.

Reference to this paper should be made as follows: Bhraguram, T.M., Rajakumar, P.S. and Banu, A.A. (2024) 'Effectiveness of digital forensic investigation through excavation methods of various Linux based tools', *Int. J. Enterprise Network Management*, Vol. 15, No. 1, pp.70–92.

Biographical notes: T.M. Bhraguram received his MSc in Computer Science and Engineering (CSE) from Bharathidasan University, India and ME in Computer Science and Engineering (CSE) from Vinayaka Missions University, India. His distinguished career spans 15 years of academic and one year of corporate experience. He has published more than 16+ articles, which include Scopus and Web of Science (WoS). He holds multifarious memberships from prominent professional bodies specifically from IACSIT, CSTA, IAENGG,

IAHFP and IARCP. His research interest and areas are data mining, cloud-based services, digital forensics, metaverse and blockchain. He has won many awards and accolades during his career and presently pursuing as a research scholar in the Department of CSE at Dr M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India.

P.S. Rajakumar received his MTech in Computer Science and Engineering (CSE) from Dr M.G.R. Educational and Research Institute, India and PhD in CSE from Jawaharlal Nehru Technological University, Hyderabad, India. His distinguished career spans 24 years of academic and two years of corporate experience. He has published more than 20+ articles, which include Scopus and Web of Science (WoS). He has also published two books with ISBN. He holds multifarious memberships from prominent professional bodies specifically from CSI, ISTE, IEEE, MIEI and MTSI. His research interest and areas are data mining, SVM classifier, IoT, cloud computing, DL and ML. He has won many awards and accolades during his career and presently serves as a Professor in the Department of CSE at Dr M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India.

Arshia Arjumand Banu received her Bachelor of Science (BSc) in Maths, Statistics and Computer Science from R.B.V.R.R Women's College, Osmania University, Telangana, India in 1998. She pursued her Master of Computer in Computer Science from Osmania University, Telangana, India in 2001. Currently, she is working as a Lecturer in the Department of Computer Science, College of Computer Science and Information Technology, Jazan University. Her research interests are artificial intelligence, machine learning, image processing and IoTs.

1 Introduction

This investigation analysis gives an efficient method for digital forensics based on the tools providing in Linux platform. The methods such as Hierarchy pattern matching, Hex code generations, Database forensics, Email spam prevention, disk and data capture, memory forensic are some of the effective analysis procedures for advanced research works. The tool performance analysed here consists of various categories of procedures as mentioned above. Some times various these categories are packed together and provide multiple functionalities under a single umbrella. The available procedures here are the exhaustive list (Majore et al., 2014) and we must focus more on the accuracy level produced by these tools in various situations. Some tools are packed up with more effective algorithm but due to the analyst skill set, the results get blurred during the analysis time. The noise and disturbance might be highly affecting core characters by the various conditions like lab conditions, equipment availability, existing laws of the region, and some contact obligations. Example a tablet or node considered as computer forensic instead of mobile forensic, once the SIM card is not available in the device. But most of the tools are providing a rich content category and possibilities while collecting and processing information. The device update and new tool introduction is important while collecting data due to the dynamic characteristics of the cyber activities and digital evidences.

Digital forensics have negatively suffered the impact of the boom of cloud computing due to its dynamic nature. The tools and procedures that were successfully proved and

used in digital investigations are now becoming irrelevant, making it an urging necessity to develop new forensics capabilities for conducting an investigation in this new environment. The major challenge faced by the digital forensic experts are the requirement of highly skilled labour force and involving a sizeable time investment. In reality these challenges might take high level impact during the investigation process. This may lead to the bad results in digital evidence collection and its proof-reading process.

Cloud forensic is a subclass of digital forensic and it assesses the need for digital investigation in a cloud environment. Due to the dynamic nature of cloud environment, it's highly sophisticated to handle the real-time evidence collection process of digital investigation. Many tools and processes are developed to handle the situation, but most of the analysis processes are failed due to the insufficiency of dynamic factor reconsideration. Problem of multiple operating systems running on mobile devices

- Type and nature of the data storage and device architecture.
- Forensics tools should be engineered to support heterogeneous investigations, preserve privacy, and offer scalability.

The framework presented here can handle the dynamic factors like time to cover the investigation, synchronous data transfer, fast and dynamic database update, evidence file updating, result analysis from various sites if applicable, file matching process, etc. information is distributed across different datacentres to enhance performance and allow load balancing, scalability, and redundancy features. There are different scenarios in cloud forensic where it can be reconstructed by investigators.

Currently the system supports various Linux based devices and its enhancements. The forensic analysis tools are used to analyse the data and to recover the deleted or hidden data from the digital devices. Normally multiple platforms are running under various mobile devices. These devices do not support a homogeneous data storage mechanisms or device architecture. Such heterogeneous diversity might become complicated while implementing the mobile device support. Forensic tools have the capability of preserving privacy and high scalability. This heterogeneous investigation environment must be supported by a good forensic investigation tool.

Digital forensics are not only useful for computer related crime, but rather in all fields of criminology. More data exchange takes place in online environment even for large and small business organisations leaving sensitive data open for criminals. So, tools are needed in real-time included smartwatch forensics, finger printing and DNA analysis

1.1 Digital forensic investigation umbrella

This is the broad category of digital forensic considerations where most of the incidents and digital proofs are verified. Most advanced research areas in this era deviating more other categories according to the new digital investigation devices and situations. When these things coming under a detailed proof reading, then digital world can occupy various processed digital evidences (Lee and Shon, 2014; Tobin and Gladyshev, 2015; Kessler and Carlton, 2014; Voorst et al., 2015) and incident reports. Today's available resources and tools are costly according to the environment that we consider. This might be the case once the digital evidences are distributed in a wide range of servers or physical locations. Consider the following digital forensic tools (Lee and Shon, 2014; Kessler and

Carlton, 2014; Vacca, 2005; Adams et al., 2013) categories in spite of evidence processing environment. The Table 1 depicts the various Digital forensic retrieval models and its processing time with risk factors involved. The processing data volume depends on the system configuration also. The risk factors involved contains various 3rd party application processing capacity also. The algorithm execution time depends on the system supporting levels.

Table 1 Digital forensics umbrella

Digital forensic tools models	*Estimated fraction level of processing*	*Risk factors involved*
Disk analysis	120 GB per hour	Damaged disk, disk I/O error, non availability of cluster
E mail investigation	1 GB per hour	Bypass the address and incubation content, spam readiness
Multimedia content evaluation	105 GB per hour	Compressed model evaluation, format conversion, unsupported devices
Browsing history and VPN	5 GB per hour	Private content, VPN bypassing, content trespassing
Network analysis	1 GB per minute	Packet tracing, network speed, load of the network
File clearance	50 GB per hour	Format conversion, Hex code generation, file permission
Computer based models	500 MB per minute	Application performance, tools functionalities, system processing capabilities
Memory content check	500 MB per hour	Cluster and segment check, fragmented content, impossible defragmentation
Database transaction evaluation	1 GB per day	Successful and roll back transaction history including database scalability
Digital image excavation	1 GB per day	Cryptographic evaluation of images, Embedding code under digital images

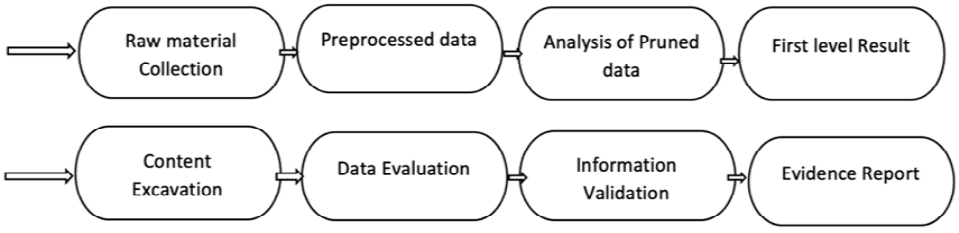
According to Warren G. Kruse II and Jay G. Heiser, authors of computer forensics: incident response essentials, computer forensics is 'the preservation, identification, extraction, documentation, and interpretation of computer media for evidentiary and/or root cause analysis.' Every digital forensic investigation contains a serious of processes/activities where the investigator must ensure the integrity and the law enforcement of the evidences collected and the government sector permissions. The tools used for the investigations must choose based on the various factors affected during the evidence collection (Majore et al., 2014; Jo et al., 2018). The criteria/factors affected including Skill level of the tools based on the competent factor, Output report based on the raw data, cost of the tool, area/focus of the tools and the additional accessories demanded by the tool.

1.2 Digital forensic scope in the current technological environment

Digital forensic investigation (Ryu et al., 2019; Tobin and Gladyshev, 2015; Carrier, 2005) is a process with various stages/phases. Each phase has its own collection method

and the data scrutiny technique which can pass the necessary information or facts to the next level. Most of the digital forensic techniques are digging the device or disk level information as the pre-process level. The assurance of this excavation highly depends on the quantity of raw data processed and the quality of the processed data. Various digital forensic investigation methods are depicted below.

Figure 1 Stages of digital forensic investigation methods



Various Digital evidence categories

a The evidences collected from any electronic devices

b The digital or electronic transactions or the evidences based on the electronic document under the Govt. laws of the respected country

c The audio or video content, email or SMS based, erased file, log files, user credentials (name and password) image and lost file encrypted file, steganographic content call logs and basic document or text files.

1.2.1 How to manage the extracted evidences?

The system must properly organise and pre-process all the high-quality digital proofs based on the parameters applicable to the process cycle. The evidence integrity or authenticity must be supervised under the government laws of the region. The investigator and other digital content organisers must maintain the evidence custodian properly and the collected evidences must match with the existing case file. The evidence must be acceptable to the current case file, its must be authentic and sufficient enough to complete at least one process cycle and also reliable.

1.2.2 Most effective evidence collection methods

According to the existing algorithmic methods, the current researches emerged many kinds of tools used for digital evidence collection. All these tools are responsible to extract, generate and analyse the evidence content. The tools like crowd strike, volatility (Tobin and Gladyshev, 2015; US Department of Justice, 2004), USB historian, sleuth kit, autopsy tool, plainsight, Linux dd, EXif tool, Neo hex editor, bulk extractor, Pladin suite, DEFT, last activity view, UDB write blocker, FTK manager, SANS SIFT, fire eye redline, HxD, Helix, ENCASE, etc. are providing high quality pre-process activities and thus generate reports with most recent and categorised output based on the regression or classification model (Kessler and Carlton, 2014).

Tools are needed and its selection If any command is executed directly, it may delete some files or loaded into memory there is a possibility to replace evidence file. Recently

used files may get updated. Also, when investigator executes some commands, it may affect the output of other file as and will intervene into system files. Different file formats and large size hard disk store more complex file hashes to database and invigilator can map suspected files. Technological changes need updated forensic tools and its documentation describing the steps as skill level, output, cost, focus area.

The Linux kernel provides a detailed process integration and high-speed synchronisation in most of the batch processing environment. This helps to implement a cloud based and high-speed computing module which requires dynamic response capability. The Linux based tools for digital forensic investigation matters here becomes more reliable compare to other heterogeneous computing environments due the following key features:

1 Fast computing environment for cloud-based systems.

2 Fats and reliable synchronous communication.

3 More reliable security features provided by Linux platform.

4 High speed computing capability of network-based tools.

5 Firewall and antivirus reliability for efficient kernel level processing.

6 Rich set of tools and platforms which can be diversely changed its nature due to the open-source atmosphere.

2 Methods of evidence extraction

2.1 Capturing data packets from high-speed networks

The network information capturing tools (Tobin and Gladyshev, 2015) are available in Caine and other open-source categories. The IP addresses, mac address, log file and other login information, cookies used for specific websites can be retrieved by the said tools and can process a detailed used activities log based on the time frame. Some tools are expert to retrieve the contents like metadata and raw data where the user's cookie information might be available through the browser or network accessing port settings. Wireshark tool can provide such raw data information about the tag sources of a particular user while his surfing history is recorded. The connected user's description can be obtained through Caine or Abel tools (Voorst et al., 2015).

2.2 Re excavate the erased contents

Normally the crime committing entities may erase the data associated with the criminal offense. Therefore, the recovery of this deleted content is very much important. The recovery process is highly useful to regenerate the deleted files and folders from the file system depends on the platform mode. The files systems like FAT and NTFS can be more generic and may support most of the data recovery tools. These file systems can incorporate the cluster management with the data recovery tools. But the file systems line Ext4, Exfat (Jo et al., 2018; Mandia and Prosise, 2001; Technical Working Group for Electronic Crime Scene Investigation, 2010) and other mac based journalised file systems needs more attention to recover the cluster content. Many of the recovery tools are failed

due to the reference segment of the deleted content. So some times it failed to produce the data which were deleted before a particular time period or from a particular disk drive. The folder and file recovery tools are making some assumptions which are normally inappropriate. Some times the file recovery process gets an assumption that the initial cluster and content size of the file or folder is part of the sequential file content. When the initial cluster identification fails then it will refer another file or folder which were deleted during the same period. Most of the tools are failed to distinguish the file entry point which was removed or overwritten. The tools like Foremost, Scalpel, PhotoRec (Voorst et al., 2015) and Datalifter (Vacca, 2005; Adams et al., 2013) can overcome these issues while recovering the data.

2.3 Social media or data science based engineering

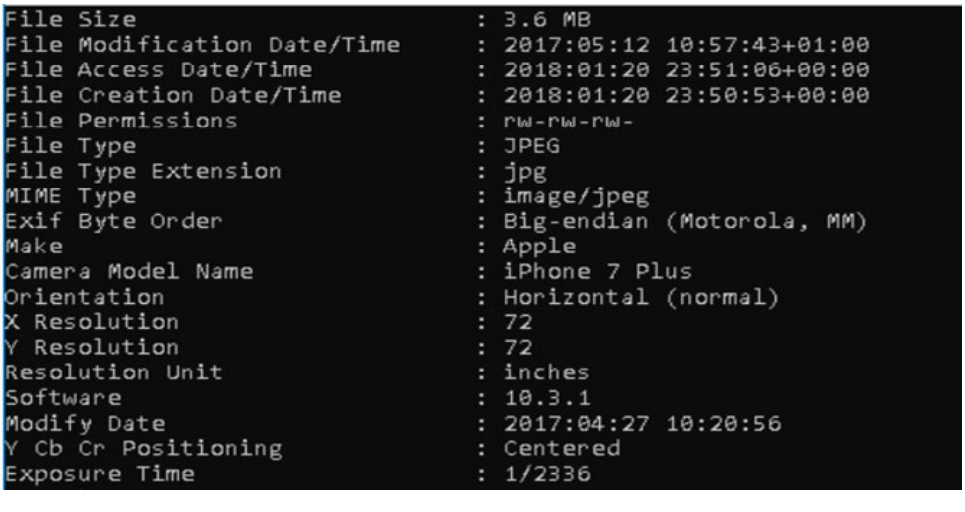
Most of the high-level spam or offenses include high political pressure and other law based loopholes made by the culprits in the case. The investigators cannot continue with the existing proof or evidences in such cases so they need more engineering related activities to do so. This type of blind situation can be overcome with special tool tips and high-performance computing methods to obtain the evidences.

2.4 Excavate embedded metadata from various sources

The tools like exchangeable image file (Exif) (Majore et al., 2014; Jo et al., 2018; Lee and Shon, 2014) are very much useful and with high computational power while extracting the metadata content of a particular file or the collected evidence. This might be useful to authenticate the genuineness and integrity of the data.

ExifTool is a console-based application used modify the metadata information of the files. It is fast, powerful and supports a large range of file formats. This can be used for analysing the suspicious files with its static properties in a host-based forensic investigation. The meta data content retrieval through Exif tools is depicted below.

Figure 2 Sample metadata content from Exif tool



The following challenges faced during the excavation tool development and its demonstration.

Easy availability of hacking tools, life span limitation of digital media, cloud service requirement, security lack, device type, data format, the increase of PC's and extensive use of internet access, lack of physical evidence makes prosecution difficult, the large amount of storage space into terabytes that makes this investigation job difficult and any technological changes require an upgrade or changes to solutions.

2.5 Pre-process the steganographic and encrypted contents

The software's used to remove password protection for particular files are more dependable to save protected and encrypted file. The tools like WinHex (Lee and Shon, 2014; Ryu et al., 2019; Tobin and Gladyshev, 2015) can remove the password protection effectively and also some tools like PRTK and distributed network attack (DNA) can bypass or recover the password used for protection. AccessData (Ryu et al., 2019; Tobin and Gladyshev, 2015; Kessler and Carlton, 2014) is an environment where we can save a number of protected file data with various encryption standards. But most of these tools are hardware dependent and some deficiency shows while processing the password recovery procedures. DNA tools can perform Brute force 40-bit encryption (Voorst et al., 2015; Jo et al., 2018) method and also can combine multiple computers for faster execution. The Adobe Acrobat and Microsoft office-based file formats can be supported by this. The investigators must use large capacity files which will not be in normal level must use some steganographic applications and password recovery method by some encryption solutions.

With the popularity of social media, many people explicit their information and criminals can use these social networking sites to commit crimes. Including data retrieval procedures, algorithmic structures and the damaged hardware components are the big risk factors included in this method. The detailed description about all these factors is depicted according to the variation of the risk factors. The transformation tools based on the datamining pre-processing has been furnished in this article relate to extract transform and load (ETL) based and also Z score normalisation.

Summarised, our contributions are:

- ∞↓ High quality analysis of various Linux based tools used for digital evidence retrieval processing.
- ∈↓ A framework for digital evidence retrieval using harvesting method.
- ∋↓ The integrity and report quality checking of various Linux based digital forensic tools.

3 Related works

In this section we give a short overview of relevant related work on Evidence management. Since our novel frame work is based on evidence processing, we also give a short overview of this research area.

3.1 Digital evidence collection using two step injection (Lee and Shon, 2014)

This proposes a new method by Nana Rachmana Syambas and Noufal from Bandung Institute of Technology Indonesia. This paper reveals some good ideas for Digital evidence collection method by using a special model called Injection model. They have developed an evidence retrieval method with two steps of digital evidence injection. This method trying to prevent the loss of digital evidences with a strong two step verification strategy. The system suggests to work the application based on the collaboration model and maintain secrecy and accuracy. But this system shows the accuracy level not more than 65% as it is tested in various damaged disk based and switched off computer systems. The system fails to produce a remote data collection strategy and lossy compression performed while the evidence collection is from a damaged system.

3.2 Revisiting risk sensitive digital evidence collection (Adams et al., 2013)

This work contains a detailed study made by Erin E Kenneally and Christopher L T Brown regarding the risk factors included while collecting and processing the digital evidences. The work modelled and framework for debate and change drivers for risk sensitive approach and cost benefit analysis of current methods. Also, the system gives a methodology component for legal and risk assessment. The article suggests a detailed data connection, integrity, data representation and file data unit implementation. It gives the sector level representation for sector data unit, mandatory supporting artefacts, the identification of access control artefacts and templates.

3.3 Review of evidence collection and protection phases in digital forensics process (Kessler and Carlton, 2014)

This article published in international journal of Information and security science by Asaf Varol and Yesim Ulgen Sonmez based on the digital evidence collection and its protection. The direct crime scene investigation and also the collection and protection based on some existing level literature content represented. Under the information technology law, how to collect the evidence from electronic devices and also how the hardware and software utilisation is maintaining while modelling the evidence collection. A crime sense investigation activity flow represented in detailed manner. This might be useful for a new investigator to manage the situation-based proof collection. The hardware devices introduced for cloning the drive or disk and also for block image capturing. Separate graphical interface for automated image restore introduced with low compression model. But the infrastructure demanded for this article is high and most of the time evidence clarity is affected due to the blurred restore activity. You may find the digital forensic cycle model is depicted below and we have a serious of steps involved here.

3.4 New Approaches to digital evidence acquisition and analysis (Vacca, 2005)

This approach implementation found as a book chapter by Martin Novak, Jonathan Grier and Daniel Gonzales under National Institute of Justice. This chapter gives a detailed description of DFORCE2 system combines the autopsy tool and Amazon EFS. The system contains various module like spark streaming job, spark cluster and DESH cluster.

Each cluster work independently and produces certain outputs that can be passed to the next cluster module. But the system shows the limitations of current prototype complexity. The standalone version demands expertise level of implementation and stand-alone server need more attention. Different set of complex activities needs to set up the system to migrate it into commercial cloud. Secure communication needs to pass the evidence structures. The disk visualisation technique explained here is useful while developing a framework for direct disk retrieval.

Figure 3 Digital forensics cycle model (see online version for colours)

3.5 Collecting evidences from a running computer (Jo et al., 2018)

This book chapter report found on The National Consortium for Justice Information and Statistics and prepared Todd G. Shipley, CFE, CFCE, Director of Systems Security and High-Tech Crime Training for SEARCH, The National Consortium for Justice Information and Statistics, and Henry R. 'Dick' Reeve, General Counsel and Deputy District Attorney, Denver, Colorado. This paper was written under the direction of the legal committee of the working group of the internet crimes against children task forces.

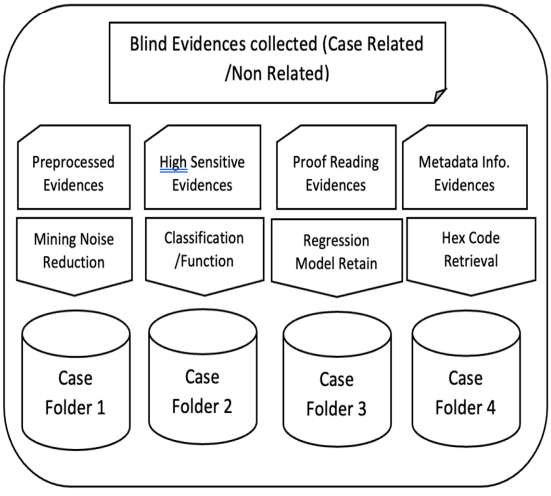
This document describes a methodology for the law enforcement collection of volatile data. The collection of this data can be of substantial use in the investigation of various criminal activities. Volatile data is evidence collected from the crime or suspected scene. A serious of steps had been described with the importance of training the investigators and collecting the data. This chapter mainly aims to produce the data collection from a running computer. Screen photograph, maintain log details, Identify the platform/operating system, note date and time, Dump the RAM, encryption status of the disk, hardware and its configuration level, etc. are some of the steps to perform the recovery process. To carry out SQL injection attacks, a malicious query is created to reveal extremely private data. The SQL injection attack interference shall be well executed through the public interface because it is the current source that an application delivers when the host-level entry point and the network are sufficiently secure (Durai and Baskaran, 2019). According to some of the literature on authentication systems, the

suggested technique conceals the original biometric template such that it cannot be reconstructed even if an attacker has access to the system's recorded data. We assess the system using a relativity matrix between security concerns, strategies, and processes (Vijayalakshmi and Karpagam, 2018).

4 Digital evidence retrieval by using harvest data method

This method uses a four-stage data/cluster module and evidence retrieval (Majore et al., 2014; Jo et al., 2018) by using disk cloning strategy. This contribution can overcome the existing data retrieval performance loss. This method can initiate flash disk pushing strategy where the first phase can collect the evidences without considering any clarity content or the validity or integrity. This blind evidence collection contains a lot of noises and disrupted data values which can be removed by performing a deep pre-processing method.

Figure 4 Harvest cycle storage container



Every evidence collection phase will iterate three times to review the collected contents and output. The second phase also tries to iterate three times to perform the initial evaluation of the evidences. In this phase the evidences are divided into four case folders. The first folder contains all the pre-processed evidences. Second case folder carries all the highly sensitive evidences related to the current investigation. Third folder contains the proof reading evidences which can be used for training purposes. The last case folder is responsible to store the Metadata Information (Tobin and Gladyshev, 2015; Kessler and Carlton, 2014; Voorst et al., 2015) of the highly sensitive evidences contained in the second folder. The storage module available as depicted below. This contains various levels of storage and each database container contains various case history data.

In first stage action, the step emphasis to prevent loss of material collection due to accidental material removal or any other situation where the data loss is the major concern. This stage initiates an application module to record all the possible evidences once the situation is analysed.

The second step reinvestigate and excavate most matching evidences from the step one and applying the excavate retrieval function and dump the specified content in the corresponding storage module. The excavate retrieval function is a classification method where the most eligible data mining algorithm may be used. The KNN classification (Lee and Shon, 2014; Ryu et al., 2019; Technical Working Group for Electronic Crime Scene Investigation, 2010) and decision tree methods are used to find the good quality classifiers and to produce the high accurate classified data.

The third stage is responsible to initiate a hex code retrieval method where the corresponding embedded meta data information (Majore et al., 2014; Jo et al., 2018) is stored and processed.

The extraction of various evidences files are through 4 stage excavation method and the requirements consists of flash drive initiation, Linux based tool activation, and existing evidences database pool. In case of a hard drive failure or other hardware malfunction, you can access your files on the cloud, which acts as a backup solution for your local storage on physical drives.

Excavation method benefits according to the evidence collection parameters gives high priority of evidence process compared to all other models. The benefits of excavation method given below based on the standard characteristics.

Table 2 Characteristics of 4 stage harvest method

1	It preserves the IT Law of the corresponding region
2	All the evidence collection rules are satisfied
3	Dynamic proof reading mechanism introduced
4	Cloning operation can be performed with or without specified contents
5	Processed Encryption standards effectively
6	Remote investigation is possible
7	Installation process needs for key loggers
8	Automated method to run the test

4.1 Investigation actors/operations representation and description

The four stage harvest method can be used by the specific investigative actors who are taking the evidence collection responsibility in various stages. The actors are: investigation head, technical rolling action, informer deviate, legal policy manager, final source node, decision making rollout. The system consists of various actions or operations which are directly operated on the target system from the initial pre-process state. Make ready with storage containers, installing the pre-process module, excavate information, proof reading action, Hex code retrieval and decision making by classified data, report generation are the actions of the system which are capable of maintaining a good evidence collection strategy.

The description and actions below are framed under the categorical level and the entire application is working with the emphasis principle of preventing the evidence damage.

The Harvest method is built on the Linux BlackArch distribution (Tobin and Gladyshev, 2015) platform and it energise the process of evidence collection during a forensic investigation. After making a successful MoU with the investigation law of the

corresponding region, the system trying to prepare the investigations steps. The database containers are created and configured for storing each level extracted evidences. The application modules can be installed into the node which is responsible to coordinate the system and a flash drive installation has been started. In case all the source modules are ready, immediately the blind evidences are collected and passed to the next level. The evidences produced by the cloning operations are pre-processed by the basic data mining pre-processing modules and the classification strategy done through Knn and decision tree operations and the classified data has been verified by the top-level module once the 4-stage operation has been concluded. The pre-processing operations (Ryu et al., 2019; Vacca, 2005) are done on the collected blind evidences to remove the noises occurred in the evidences while collecting it. The system trying to apply the excavation method with proper proof reading action. The content must be analysed the can be checked by then content relation before moving to the next level. Once the content relation is satisfied the analysed data might be moved to the decision making (Kessler and Carlton, 2014); Vacca, 2005) module which is responsible to make the decision on the data to be moved to the corresponding database container. If the content relation is not satisfied according to the investigator level, the process restarts from the exaction method again.

Table 3 User/actor descriptions of 3 stage harvest method

Acting user	*Description*
Investigation head (Ur-HAR-01)	This system user is responsible to handle the entire investigation operations and initiate direct instructions to the technical users
Technical rolling action (Ur-HAR-02)	This system user is responsible to perform the application operations with successful evidence collection
Informer deviate (Ur-HAR-03)	This collaborative user is the person giving specific investigation information to the head actor
Final source node (Ur-HAR-04)	The source of evidences to be caught by this user includes browsing history, chat based log, documents, passwords, images and other history logs
Decision making rollout (Ur-HAR-05)	The evidences to be analysed and categorised based on the relation and importance by applying the programming module
Legal policy manager (Ur-HAR-06)	This actor is responsible to make the legal rights obtained according to the law of the specific region

The second level evidence retrieval performs after the successful decision making and once the operation is done, the system can go with the accuracy level checking initiation of the evidence data. If then accuracy is low and not up to the satisfied level, the proof reading action can be made again to analyse the classified data. The accuracy level is satisfied; the first level results can be moved submitted to the investigators. The system moves to the next level of Hex code retrieval (Kessler and Carlton, 2014), which can be applied to the meta data information of the analysed content. The meta data content checked thoroughly and looking for a modified content if any. Once then post hex code analysis is ready, the harvest method can integrate all the related evidences into a single evidence file. Otherwise the hex code retrieval operation must be performed again.

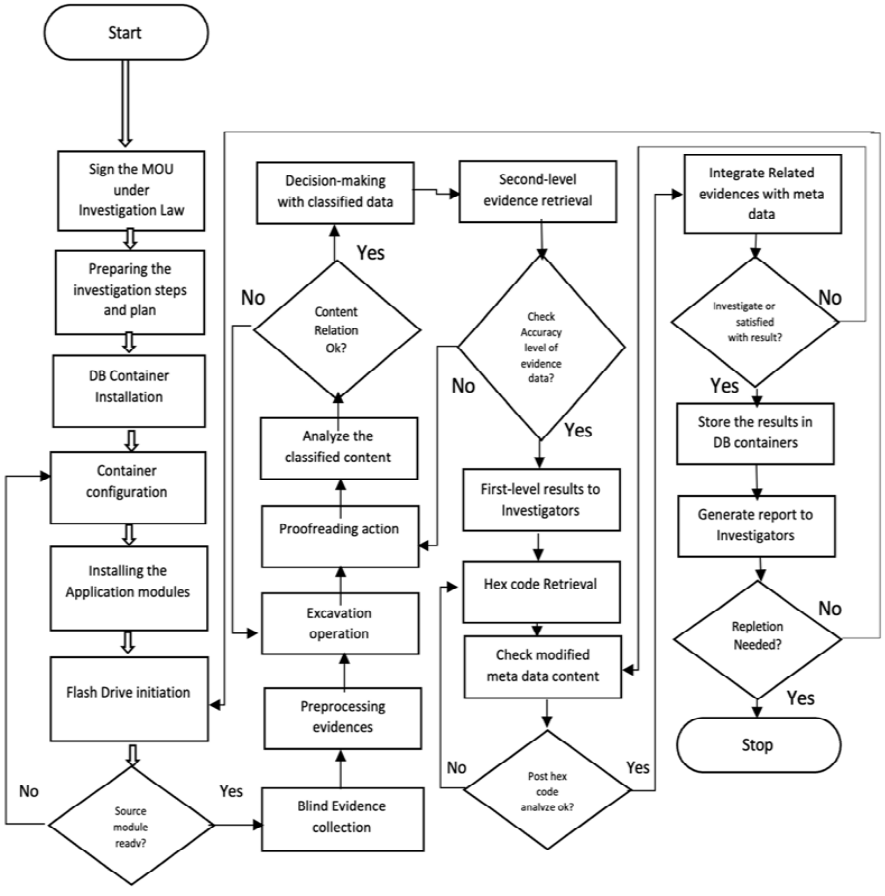
Table 4 Operational modules of the harvest method

Action/operation module	*Description*
Storage container installation (OPR-HAR-01)	The database installation for various evidence collection is coming under this operation
Installing the pre-processing module (OPR-HAR-02)	The noise reduction and pre-processing algorithms applied by this action and the said data to be moved to the corresponding containers
Excavation operation (OPR-HAR-03)	This module will install client application console with key logger module and the installation will proceed with the Linux based distribution. Here in our case it will be with BLACKARCH Linux distribution. The flash disk used to get the document layout
Proof reading action (OPR-HAR-04)	The classification model applies to the retrieved detailed data and categorised to the specific containers. Flash disk applied in this case also
Hex code retrieval (OPR -HAR-05)	The meta data information of each document is retrieved and the document layout is stored in the specified container. This step is highly important to get the misused information as well as to retrieve damaged document content
Decision making on classified data (OPR-HAR-06)	The decision making algorithm applied in this level and the core information content is moved to the report module.
Report generation (OPR -HAR-07)	This module is responsible to generate the report according to the investigator's requirement

The proof reading methods in the second stage depicted in this article shows the methods of new evidence injection technique which can be initiated during the copy process. The evidence collected can be reiterated with the existing database modules if there is any similarity found. In case the recovered data doesn't match the existing contents, the system will try to create a new evidence file module and thrown to the database storage. The stored contents may push to the reporting section with a marked procedure and indicate to the investigators about the new evidence content found and need attention to the verification manually. The integrated evidence file submitted to the investigators for their satisfied level. If the investigator is ready with the current evidence file, then the file can be moved to the corresponding database container and can generate a report. Otherwise system must move to the metadata modified (McMillian, 2000; Farmer and Venema, 2000) content checking and restart the process. The system gives an option to repeat the whole process if needed.

The operations are in four stage levels and the standard operation procedure depicted in Figure 5.

Figure 5 Standard operation procedure of four stage evidence collection harvest method



5 Experimental results and analysis

The Harvest system tested with five different people who are portrayed as the target and the evidences collected based on that. The storage media used here as the external disk with the storage capacity of 500 GB USB 3.0 from Seagate GoFlex (Majore et al., 2014; Jo et al., 2018). The configuration of the target system as follows.

5.1 Level 1 (Stage one) experimental result

The stage one represents the blind evidence collection mode and it directly acts on various inf and bat files of windows systems, ext partition files in Linux kernel (bin, bz2, conf, deb, dsc, ebuild, log, pid, ps, rpm, etc.) and Journalised file system in Mac (app, cnf, dmg, icnf, etc.). (Tobin and Gladyshev, 2015; Voorst et al., 2015). The considered file types and extensions are directly applied to the various file systems. The autorun file failed to execute due to the antivirus block and it opens a door to make a manual

execution (Khobragade and Malik, 2014b) of the specified file type. This operation can be initiated by the harvest method automatically.

Table 5 Hardware configuration of the target system used in harvest method

Target Node 1 (social engineering network based)	*Target Node 2 (low speed computer network oriented)*	*Target Node 3 (high speed computer network oriented)*	*Target Node 4 (high configured Antivirus with password protection)*	*Target Node 5 (high configured antivirus with password, firewall, disk encryption and VPN based protection)*
Linux Mint modified kernel	Mac Os High Sierra kernel	Windows 10 with modified features	MX Linux modified kernel	Mac Os Catalina kernel
Intel i5 core 2.3 GHz	Intel i7 core 2.9 GHz	Intel i7 core 2.9 GHz	Intel i5 core 2.3 GHz	Intel i5 core 2.3 GHz
Turbo boost with 3.2 GHz	Turbo boost with 3.2 GHz	Turbo boost with 3.2 GHz	Turbo boost with 3.2 GHz	Turbo boost with 3.2 GHz
Nvidia GeForce GT 520 4 GB RAM	AMD Radeon HD 6250 4 GB RAM	AMD Radeon HD 6250 12 GB RAM	Nvidia GeForce GT 520 8 GB RAM	Nvidia GeForce GT 520 6 GB RAM
HD 1 TB 7500rpm	HD 500 GB 7500 rpm	HD 500 GB 7500 rpm	HD 1 TB 7500 rpm	HD 2 TB 7500 rpm

Table 6 First level evidence collection through harvest method

File types/parameters to be executed	*Execution status*	*Description*
Social engineering file types	Succeed	System tries to copy the content from social media data repository
Autorun file types (Adams et al., 2013)	Failed	The antivirus blocked the execution of auto running file types
Batch file execution under windows and instr file from Linux and dmg zipped content from Mac	Failed	The blocked file types are executed manually and tries to open the configuration files for the supported platform
Instruction file extraction	Failed	The content of the specified file type to be revealed and its visible to the next level process.
Key logger module contents and short cut installation (Kessler and Carlton, 2014; Vacca, 2005)	Succeed	Key logger file can be renamed to bypass the suspicious domination and can create an exe file in windows or APK file in Mac. The root folder can host the renamed file
Registry file copying (Carrier, 2005)	Failed	Users class files and log files for each session should be copied
The browsing history files in xml model	Succeed	The chat history and all the internet browsing activities from all the installed browsers and must be exported in xml form
Cloning of root/system file types	Succeed	Cloning of the home, system and root folder in Linux, Home folder from Mac and my documents, my pictures and my downloads folder in Windows

5.2 Level 2 (Stage two) experimental result

The second level tries to clone the root folder and or the entire disk content including the file system mode into a target external drive which can hold the entire data content. The log files and registry contents are most sensitive part and it has the system level protection and also by some antivirus or system maintenance tools in some cases. The closing process can be disabled by most of the operating system kernel and it tries to terminate the process once the file indexing operation (da Cruz Nassif and Hruschka, 2013; Khobragade and Malik, 2014a) is finished. The Harvest method tries to diminish the content cloning prevention and it forwarding the cloning bypass process by using special harvest commands such as Clonedrive, Clonefdr, Clonelog, Clonereg, Clonefl and Clondsk. The said commands are used to clone system drives, folders, log content registry modules, files and disk respectively.

The syntax of ten harvest commands:

%Clonedrive% " C.*" "%drive% <"target disk" : "%drive%">"

where target disk drive might be any external disk drives which has the same file system type

%Clonefdr% "%drive " C.*<Folder names>"<"target disk" : "%drive%">" <"target folder name">%"

%Clonelog% "%drive " C.*<all.log>"<"target disk" : "%drive%">" <"target folder name">%".<session log.log>%"

Where all.log contains the entire session details of the specified time

%Clonereg% "%drive " C.*<system.reg>"<"target disk": "%drive%">" <"target folder name">%".<session log.log>%"

%Clonefl% "%drive " C.*<file name>"<"target disk": "%drive%">" <"target folder name">%".<file name>%"

%Clonedsk% "%disk" C.*<source disk>"<"target disk">

Linux can provide an empirical evidence if the Linux-embedded machine is recovered from a crime scene

/etc [%SystemRoot%/System32/config]

This contains system configurations directory that holds separate configuration files for each application

/var/log

This directory contains application logs and security logs. They are kept for 4–5 weeks

/home/$USER

This directory holds user data and configuration information

/etc/passwd

This directory has user account information.

5.3 Level 3 (Stage three) experimental result

The failed cloning operations are initiated again to increase the success rate and the results produced as below.

Table 7 Second level evidence collection through harvest method

File types/parameters to be executed	*Execution status*	*Description*
Autorun file types (Adams et al., 2013)	Succeed	The antivirus blocked the execution of autorunning file types
Batch file execution under windows and instr file from Linux and dmg zipped content from Mac	Succeed	The blocked file types are executed manually and tries to open the configuration files for the supported platform
Instruction file extraction	Failed	The content of the specified file type to be revealed and it is visible to the next level process
Registry file copying (Carrier, 2005)	Succeed	Users class files and log files for each session should be copied

5.4 Level 4 (Stage four) experimental result

The level 4 initiates the cloning operation again and try to solve all the activities before the report generation. The accuracy level inception implied once the stage 4 results had been generated. The accuracy measurement is a multiple iteration of the results and also trying to compare with the sample data stored in the database. Once all the batch level (Sadiku et al., 2017) experimental results are generated, the system can move with the reiteration or process repeat if needed.

Table 8 Third level evidence collection through harvest method

File types/parameters to be executed	*Execution status*	*Description*
Instruction file extraction	Succeed	The content of the specified file type to be revealed and it Is visible to the next level process

Figure 6 Harvest method iteration status (see online version for colours)

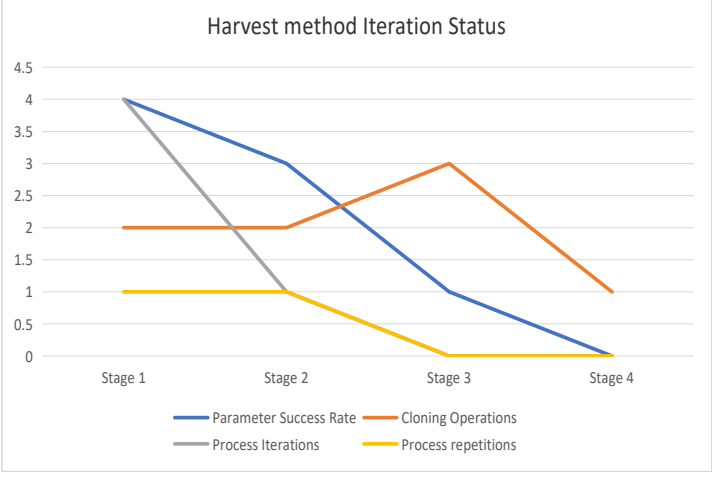
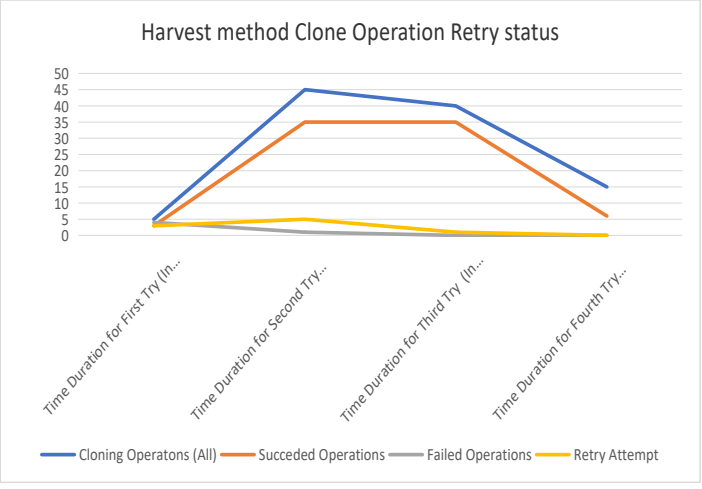


Figure 7 Harvest method clone operation retry status (see online version for colours)



Four measurements considered in Figure 6 shows the harvest method Iteration status. This depicts various parameter success rate, cloning operations considered, number of process iterations and the number of times processes repeated. Parameters success rate touched 100% after the successful fourth stage moving. The cloning operations are succeeded with three operations in stage number three and it need more iteration parts. The repetitions of the processes stopped at stage 2 and none of the stage process repetitions needed after it. Figure 7 gives a detailed picture of how many retries made for successful cloning operations. According to the time duration the retry attempt varied from four to zero and it depicts a high success rate of the cloning operations. Failed operations reached zero after the specified time duration. The system does not show any failure after its successful four stage compilation.

The 4 stage exaction claims that the first stage failure for any cloning operations are successfully reiterated to the next level. The hardest operations like multimedia data content retrieval need multiple level of iterations and the same can be done through the last stage. A detailed Iteration status measurement obtained is referenced below and the system failed to produce more accuracy level only during the cloning operation. Session log copy status is partially succeeded in first iteration in most of the systems.

The method is implemented on the top layer of Linux based distribution and the kernel thread executions activated every time whenever a new process is generated. The response time for every iteration of the evidence collection phase is measured based on the time to execute the method, mean response time and the number of retries performed by the system. Multimedia and any bad sector data retrieval may fail due to the accidental eraser or the poor quality evidence atmosphere. Threshold level for any retries of the operation set to maximum four and the system successfully passed all the various data format retrieval upon this number. So at any circumstances the system can achieve maximum accuracy rate if it performs in a bad environment also. The retrieved data can be pushed up to the classified data accuracy verification procedure once the copy and other cloning operations are completed. The accuracy measurement procedures depicted in the article in detailed manner.

Table 9 Harvest method first iteration accuracy measured

Harvest method first iteration accuracy measured			
Measurement applied	*First iteration status*	*Accuracy produced (%)*	*Suggested for second iteration?*
File category mode	Succeed	100	No
Folder category mode	Succeed	100	No
Disk drive mode	Succeed	100	No
Full disk mode (Cloning status)	Failed	10.5	Yes
Registry log mode	Failed	23.25	Yes
Session log file mode	Partial succeed	90.2	Yes
Browsing history	Succeed	100	No
Multimedia content retrieval	Failed	65.4	Yes

The accuracy and high performance in connection with data retrieval and evidence collection assured by the 4 stage excavation method. The evidences recorded by this method gives zero failure copy and cloning operations and the failure procedures can be reiterated to make it with high accuracy percentage. The existing methods for evidence collections doesn't perform any reiteration so cannot ensure the data collection from failed modules. All the existing methods are performed single stage collection methods and also does not focus on the pre-processing or classification methods.

There are no active exaction tools available to compare the proposed system and the accuracy measures checked with the proposed system criteria. The method experimented with various processing environments and with processing capability issues. The Linux kernel environment gives a high accurate batch processing capability and the system utilises the security measures of Linux distributions.

6 Harvest method supporting of Linux distributions

The harvest method is the latest and advanced digital evidence collection method which can be used in various Linux distributions. The digital forensic distribution in Linux platform is more efficient than any other operating system platforms due to its consistency, portability and security. Here various Linux platform based (Tobin and Gladyshev, 2015) tools are used for evidence collection and processing and the proposed Harvest evidence collection mechanism can be used with many of the Linux distributions and existing tools. The adaptability of the system is high and can make it as a plug-in compatibility. This proposal drawing an attention of the below tool descriptions and supportability of the Harvest method. The adaptability and plug in capability are also added.

The harvest method implemented in cloud-based infrastructure and the frame work is scalable and highly portable. The system can support offline system investigation and local evidence collection in remote places. Once the local investigation done, the system can be integrated with the cloud modules and the evidence file synchronisation carried out.

Table 10 Supporting status of Linux based distributions/platforms of harvest method

Tool name	*Harvest method adaptability*	*Plug in capability*	*Time need to integrate the result*	*Database model/container specification*
Open computer forensics architecture (OCFA)	Supported	Yes	< 5 mins	Postgre SQL
Caine	Supported	Yes	< 5 mins	Postgre SQL, MongoDB
X-ways forensics	Not supported	Yes	< 5 mins	LiteSQL
Digital forensic framework	Supported	Yes	> 10 mins	Oracle
EnCase	Supported	Yes	< 5 mins	Postgre SQL, IBM DB2
Registry recon	Not supported	Yes	> 15 mins	Oracle
Sleuth kit	Not supported	Yes	< 2 mins	MongoDB
Volatality	Supported	Yes	< 5 mins	Postgre SQL, MongoDB
Autopsy	Supported	Yes	>40 mins	Postgre SQL
Llibforensics	Not supported	Yes	< 5 mins	Postgre SQL
Oxygen forensic suite	Supported	Yes	> 45 mins	MongoDB
Coroners tool kit	Not supported	Yes	< 5 mins	Postgre SQL, MongoDB
Bulk extracter	Not supported	Yes	>30 mins	Postgre SQL
Xplico	Not supported	Yes	< 5 mins	Postgre SQL, Oracle
Mandiant Red Line	Supported	Yes	< 5 mins	Postgre SQL
P2 Explorer	Not supported	Yes	> 1 5 mins	MySQL, Oracle
Helix 3	Not supported	Yes	< 5 mins	Postgre SQL
XRY	Not supported	Yes	> 25 mins	Postgre SQL
Cellebrite UFED	Supported	Yes	< 5 mins	Postgre SQL, Oracle
SANS SIFT	Supported	Yes	> 15 mins	Postgre SQL, Orcale
Pro Discover Forensic	Not supported	Yes	> 20 mins	Postgre SQL
Black Arch	Supported	Yes	< 5 mins	Postgre SQL, Mongo DB

7 Conclusions

The Digital forensic evidence collection is the most important process when a new investigation process is carried out. This paper giving a detailed Linux based forensic environment tools and its adaptability with the proposed system one which is drawn out here. The harvest method proposed here is highly useful for the investigation gateway

which must be carried out by the investigators and the supporting technical persons. This system can retain and maintain the meta data information after modifying for the deep investigation. The system helps to collect various digital cyber forensic investigation evidences with its high accuracy and relation with the crime or offense activity. The system is adapted with most of the Linux digital investigation distributions and also with the forensic tools proposed and developed from various third party organisations. The portability is appreciable and its security is depending on the Linux platform firewalls. The system is capable of preserving various evidence category files with more accuracy and with less iterations steps. It can maintain the detailed log and information retrieval from damaged devices or from any other digital media. The four stage process/operation can take a detailed verification of the collected evidences and the separate DB containers used here is capable of classifying then evidences with respected categories. The configuration of the Harvest method is little bit complex but the configuration supports most of the Linux distributions. Further development can be made on the portability with other operating systems and application platforms including mobile devices. The communication channel is weak depends on the device configuration used as target and also it affects the performance of the source devices also. The harvest method is still the most beneficial system approach which can be used in various domains. Integrating excavation method and machine learning algorithms can be an intelligent tool in investigation of suspected machines.

References

Adams, R., Hobbs, V. and Mann, G. (2013) 'The advanced data acquisition model (ADAM): a process model for digital forensic practice', *Journal of Digital Forensics, Security and Law, JDFSL*, Vol. 8, No. 4, pp.25–48.

Carrier, B. (2005) *File System Forensic Analysis*, Addison-Wesley Professional, Boston, MA [online] http://www.awprofessional.com/bookstore/product.asp?isbn=0321268172&rl=1 (accessed 20 August 2020).

da Cruz Nassif, L.F. and Hruschka, E.R. (2013) 'Document clustering for forensic analysis: an approach for imp roving computer inspection', *IEEE Transactions on Information Forensics and Security*, January, Vol. 8, No. 1.

Durai, K.N. and Baskaran, K. (2019) 'Decision tree classification - N tier solution for preventing SQL injection attack on websites', *IJENM*, Vol. 10, Nos. 3–4, p.253, DOI: 10.1504/IJENM.2019.103155.

Farmer, D. and Venema, W. (2000) 'Forensic computer analysis: an introduction', *Dr. Dobb's Journal*, September [online] http://www.ddj.com/documents/s=881/ddj0009f/0009f.htm (accessed 1 November 2001).

Jo, W., Chang, H. and Shon, T. (2018) 'Digital forensic science approach by file recovery research', *J. Supercomput.*, Vol. 74, pp.3704–3725, https://doi.org/10.1007/s11227-016-1909-2.

Kessler, G. and Carlton, G. (2014) 'A study of forensic imaging in the absence of write-blockers', *Journal of Digital Forensics, Security and Law, JDFSL*, Vol. 9, No. 3, pp.51–58.

Khobragade, P.K. and Malik, L.G. (2014a) 'A review on data generation for digital forensic investigation using data mining', *International Journal of Computing and Technology*, April, Vol. 1, No. 3, pp.146–105, https://doi.org/12.1007/s12227-011-117796.

Khobragade, P.K. and Malik, L.G. (2014b) 'Data generation and analysis for digital forensic application using data mining', in *Communication Systems and Network Technologies (CSNT), 2014 Fourth IEEE International Conference on*, 7–9 April, pp.458–462.

Lee, S. and Shon, T. (2014) 'Improved deleted file recovery technique for Ext2/3 filesystem', *J. Supercomput*., Vol. 70, pp.20–30, https://doi.org/10.1007/s11227-014-1282-y.

Majore, S.A., Yoo, H. and Shon, T. (2014) 'Secure and reliable electronic record management system using digital forensic technologies', *J. Supercomput.,* Vol. 70, pp.149–165, https://doi.org/10.1007/s11227-014-1137-6.

Mandia, K. and Prosise, C. (2001) *Incident Response: Investigating Computer Crime*, Osborne/McGraw Hill, Berkeley, CA [online] http://books.mcgraw-hill.com/getbook.php?isbn=0072194510&template=osborne (accessed 20 August 2020).

McMillian, J. (2000) *Importance of a Standard Methodology in Computer Forensics*, May [online] http://www.sans.org/infosecFAQ/incident/methodology.htm (accessed 20 August 2020).

Ryu, J.H., Sharma, P.K., Jo, J.H. et al. (2019) 'A blockchain-based decentralized efficient investigation framework for IoT digital forensics', *J. Supercomput.*, Vol. 75, pp.4372–4387, https://doi.org/10.1007/s11227-019-02779-9.

Sadiku, M., Tembely, M. and Musa, S. (2017) 'Digital forensics', *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 7, pp.274–276, DOI: 10.23956/ijarcsse/V7I4/01404.

Technical Working Group for Electronic Crime Scene Investigation (2010) *Electronic Crime Scene Investigation: A Guide for First Responders*, 2nd ed., Diane Publishing, Darby, PA [online] http://www.ncjrs.org/pdffiles1/nij/199408.pdf (accessed 20 August 2020).

Tobin, L. and Gladyshev, P. (2015) 'Open forensic devices', *Journal of Digital Forensics, Security and Law, JDFSL*, Vol. 10, No. 4, pp.97–104.

US Department of Justice (2004) *Office of Justice Programs, National Institute of Justice, Forensic Examination of Digital Evidence: A Guide for Law Enforcement*, NIJ Special Report Series, NCJ 199408, Washington, DC, April [online] http://www.ojp.usdoj.gov/nij/pubs-sum/199408.htm.

Vacca, R.J. (2005) *Computer Forensics*, 2nd ed., Charles River Media, IBSN: 1–58450–389–0.

Vijayalakshmi, S. and Karpagam, G.R. (2018) 'Authentication as a service in cloud from a fuzzy perspective', *IJENM*, Vol. 9, Nos. 3–4, p.352, Doi: 10.1504/IJENM.2018.094674.

Voorst, R.V., Kechadi, M-T. and Le-Khac, N-A. (2015) 'Forensic Acquisition of IMVU: a case study', *Journal of Digital Forensics, Security and Law, JDFSL*, Vol. 10, No. 4, pp.69–78.